The molecular epidemiology of *Trypanosoma cruzi* infection in wild and domestic

transmission cycles, with special emphasis on multilocus microsatellite analysis.

Martin S. Llewellyn

2008

Abstract

Trypanosoma cruzi is a zoonotic vector-borne unicellular parasite, with a highly complex silvatic ecology, and native to the Americas. Substantial genetic diversity has been identified in *T. cruzi* populations, with six phylogenetic groups or Discrete Typing Groups (DTUs) commonly recognised: TCI, TCIIa, TCIIb, TCIIc, TCIId, and TCIIe. The silvatic affinities of these groups are poorly defined, although broad associations between some lineages and distinct ecological niches are recognised. Additionally, a number of studies have demonstrated a degree of within-DTU diversity, and the current classification may be a poor reflection of the total diversity present.

In this PhD thesis the genetic diversity of silvatic *T. cruzi* is examined, in conjunction with a limited number of domestic strains, to investigate the underlying ecological and epidemiological phenomena that dictate the population genetic structure of this parasite.

>200 new *T. cruzi* and *Trypanosoma rangeli* isolates, including those from silvatic mammals, domestic and peridomestic triatomine bugs, were collected during fieldwork in Venezuela and Bolivia. Where possible, these isolates were genotyped to a DTU level, and the epidemiological significance of these data discussed. Original silvatic genotype data from this study were then compiled with >1000 historical records (1981-2007) for both mammals and triatomines. This dataset was subjected to basic statistical analysis, and strong support found for an association between parasite genotype, silvatic niche, and triatomine vector.

Within-DTU genetic diversity was established for ~200 isolates from two widespread silvatic genotypes, TCI and TCIIc, using a genome-wide panel of 49 microsatellite markers, in tandem with sequence analysis. Substantial genetic diversity was identified in both lineages, coincident with weak spatial structuring. The possibility a population bottleneck was investigated within TCI derived from Andean rodent populations. Moreover, the possibility of a bottleneck was also examined in geographically dispersed human TCI isolates taken from lowland Venezuela. Associated epidemiological implications are discussed.

Genetic diversity in TCI was additionally examined at a within-host level. A total of 211 clones were taken from eight mammals, and analysed using a subset of microsatellite markers. Again substantial genetic diversity was evident, with stable infection of the same mammal by a number of different stains. Limited evidence of genetic exchange was also observed, but could not be confirmed, and the implications of this are also discussed.

Contents

List of Abbreviations

(study site abbreviations are not included)

AMOVA – Analysis of molecular varance

ASI – Amazonian surveillance initiative

API – Andean pact initiative

bp – base pairs

CAI – Central American pact initiative

CSM – Cell surface membrane

DDT - Dichloro-diphenyl-trichloroethane

DNA – Deoxyribonucleic acid

FACS - Fluorescence activated cell sorting

GAI – Great American interchange

GIS - Geographical information system

Gpi – Glucose-6-phosphate gene

GDP – Gross domestic product

HW – Hardy-Weinberg

HWE – Hardy-Weinberg equilibrium

IAM – Infinite alleles model

IgG - Immunoglobulin G

kDNA – Kinetoplast DNA

LD – Linkage disequilibrium

LSHTM – London School of Hygiene and Tropical Medicine

MASP – Mucin-associated surface protein.

ML – Maximum likelihood

MLEE – Multilocus enzyme electrophoresis

MLEW – Michael Lewis (LSHTM)

MLG – Multilocus genotype

MLMT – Multilocus microsatellite typing

MLST – Multilocus sequence typing

MNA – Mean number of alleles per locus

MWG – Dr. Michael W. Gaunt (LSHTM)

MYA – Million years ago

PCR- Polymerase chain reaction

PCR-FLP – Targeted PCR-fragment length polymorphism

RAPD – Randomly amplified polymorphic DNA

RFLP – Restriction fragment length polymorphism

RNA – Ribonucleic acid

rRNA – Ribosomal RNA

SCI - Southern Cone initiative

SLST – Single locus sequence typing

SMM – Stepwise mutational model

SNP – Single nucleotide polymorphism

SSU – Small subunit

T. – *Trypanosoma*

Ta. – *Tassuya*

Tam. – *Tamandua*

TCI, IIa-IIe – *T. cruzi* I-,IIa-IIe

To. – *Toxoplasma*

Tr. Triatoma

UCV – Universidad Central de Venezuela

w/v – weight/volume

v/v – volume/volume

VSG – Variant surface glycoprotein

Glossary of selected words

Allopatric – geographically separate populations or individuals

Arboreal – inhabiting trees

Anthropophily – liking or preferring humans

Autochthonous – of indigenous origin

Congenital – disease transmission from mother to child *in utero*

Ecotope – an ecological type

Epizootic – an outbreak of disease in animals linked to human activity

Fossorial - terrestrial / ground-dwelling

In silico – using a computer

Interpopulation – between major *T. cruzi* genotypes

Intrapopulation – within major *T. cruzi* genotypes

Infrapopulation – the population of parasites within an individual host

Isthmus – a narrow strip of land connecting two landmasses

Monoxenous – requiring a single host to complete the lifecycle

Myrmecophagous – exclusively feeding off ants/termites

Silvatic – from the natural environment

Sympatric – occupying the same geographical area

Syntenous – location of genetic loci on the same chromosome

Topology – tree branching order (phylogenetics), physical environmental features (geography)

Trap-night – a single mammal trap set over the duration of a one night

Xenodiagnosis – sampling of parasites from animals using laboratory-reared vectors

Xerophilous – showing adaptation to limited water supply

Zoonotic – an infectious disease that can be transmitted from wild/domestic animals to humans

Zymodeme – an isoenzyme electrophoretic pattern

List of Figures

List of Tables

18

List of Equations

21

1 Introduction.

1.1 General aspects of *Trypanosoma cruzi.*

1.1.1 *T. cruzi*: biology, genome and life cycle.

T. cruzi, the aetiological agent of Chagas disease, is a flagellate trypanosomatid protozoan parasite of the order *Kinetoplastida. T. cruzi* shares many aspects of its cellular architecture with other eukaryotic cells. Distinctive to *T. cruzi* and other members of the order Kinetoplastida, however, are a number of highly specialised organelles and suborganellar structures (1).

Like all kinetoplastids, *T. cruzi* possesses a single mitochondrion positioned within the cell body at the base of the flagellum. Mitochondrial DNA is contained within a bar shaped structure called the kinetoplast, consisting of a concatenated network of 20-50 circular 22-28kb molecules called maxicircles and tens of thousands of ~1.4kbp minicircles (1-3). Transcription of mitochondrial DNA occurs through a unique mechanism of RNA editing, whereby minicircles encode guide RNAs that are responsible for post-transcriptional modification of maxicircle-derived mRNA. Other characteristic features of the Kinetoplastida include vesicle-bound glycosomes, within which glycolysis is compartmentalised; acidocalcisomes, in which minerals accumulate; and reservomes, which act as membrane stores.

In common with two other important trypanosomatids, *Trypanosoma brucei* and *Leishmania major, T. cruzi* is thought to be a largely diploid organism (4-8). The precise chromosomal number and organisation is poorly defined. *T. cruzi* does not possess true centrioles and undergoes 'closed' mitosis, with a microtubular spindle arising from poorly defined structures in the nuclear membrane (9). Simple cytogenetic observation of chromosomes is not possible as they do not condense at any point during the cell cycle. In the absence of cytogenetic confirmation, a combination of southern blot analysis and pulse field gel electrophoresis identified at least 25 homologous chromosomes in *T. cruzi* (5). Later analyses demonstrated the presence of a greater number (between 55 and 57), but also revealed that the number differed significantly between strains, with considerable size differences (up to 173%) between homologous chromosomes from the same clone (10). The high degree of structural and compositional plasticity in the *T. cruzi* genome is, to a degree, reflected in that of *T. brucei. T. brucei* possesses 11 diploid megabase chromosomes, while the composition

24

of the rest of the genome – comprised of aneuploid intermediate and mini-chromosomes – is more highly variable (11). The *T. cruzi* genome sequence, although poorly mapped, suggests that approximately 50% of the entire 90-150 Mb genome is made up of multiple repeat sequences (6). Many repeats (~15%) comprise large surface protein gene families such as mucins, trans-silidases, and mucin-associated surface proteins (MASPs) which, like the variant surface glycoprotein (VSG) genes and VSG gene expression sites encoded on *T. brucei* mini and intermediate chromosomes respectively, are thought to play a role in the parasite's interaction with the host immune system. Thus the significant degree of surface epitope variation required to survive in the host environment may be, at least in part, responsible for the high level of chromosomal rearrangement and genome size plasticity observed in *T. cruzi*.

American trypanosomiasis is a vector-borne zoonosis requiring both a mammalian and invertebrate host to complete its life cycle. Passage through the triatomine bug vector and subsequent transmission to the mammalian host means the parasite must successfully colonise and exploit two drastically distinct environments. As such the parasite undergoes a number of morphological and physiological metamorphoses which are characteristic of different stages of the lifecycle (9).
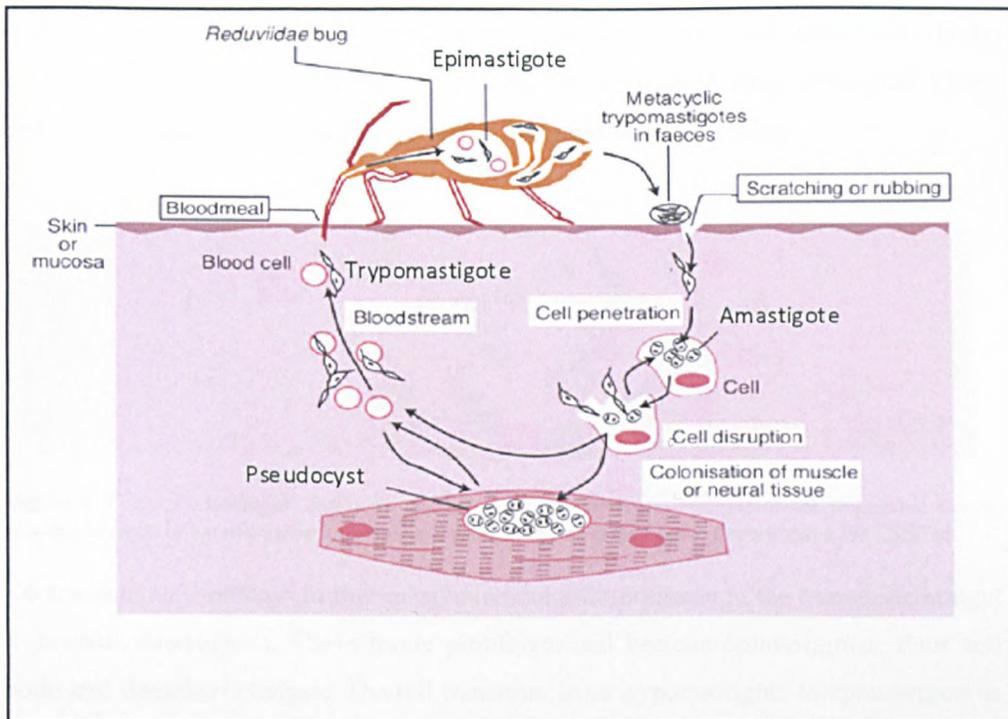
Figure 1 *T. cruzi* **life cycle.** Adapted from Macedo *et al.*, 2002 (12)

Metacyclic trypanosomes ejected from the triatomine hindgut are an elongate, highly motile, non-proliferative stage in the *T. cruzi* life cycle whose role is to establish infection in the mammalian host. The parasite gains entry through the proboscis penetration site, scratches, abraded skin or via a variety of mucous membranes including the conjunctiva and the wall of the upper alimementary tract. Metacyclic trypanosomes can penetrate most nucleated cells and are thought to do so via one of two distinct mechanisms (9, 13). Most commonly the parasite is able to directly recruit lysosomes to the point of entry by manipulation of the host cell's microtubular cytoskeleton. Lysosomes fuse with the cell surface membrane (CSM) to create a vacuolar compartment in which the parasite enters the cell. Alternatively, the host's actin cytoskeleton is activated and the parasite engulfed in a parasitophorous vacuole. Once inside the cell, lysosomes continue to fuse with the parasitophorous vacuole. The increasing acidification inside this structure act as a trigger for the differentiation of the parasite to the intracellular amastigote form and its concomitant escape into the cytoplasm. Once in the cytoplasm, the amastigote, spherical in shape and typified by a shortened flagellum and cell body, undergoes proliferation by binary fission to form an intracellular nest of amastigotes, known collectively as a pseudocyst. When intracellular

amastigotes reach sufficient density, they give rise to small, c-shaped or elongate trypomastigote forms which escape into the blood stream and interstitial fluids. Trypomastigotes are then capable of invading further cells or enter circulation where they may be taken up by a haematophagous triatomine during feeding.
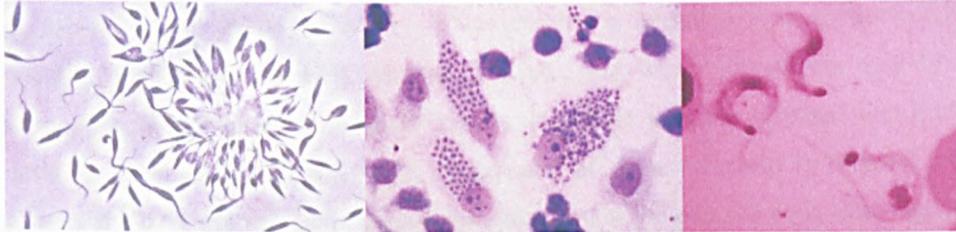
Figure 2 *T. cruzi* **life stages**. Right to left: epimastigotes from the mid-gut of the triatomine vector; amastigote nests in muscle tissue; circulating trypomastigotes in the blood. Source: MLEW, LSHTM.

Trypomastigotes undergo further morphological differentiation in the triatomine midgut to become amastigotes. These forms proliferate and become epimastigotes, their cell body and flagellum elongate. Overall transition from trypomastigote to epimastigote is thought to be a response to movement from a glucose-rich environment in the host to a relatively glucose-poor environment in the vector. Progression to the infective metacyclic form, again due to changes is acidity and nutrient availability, occurs in the hindgut and rectum. Epimastigotes attached to the waxy cuticle of the hind gut wall elongate, the single mitochondrion becomes more active and the parasite, now metacyclic, detaches from the gut wall ready for passage to the host.

Variation in the structural composition of cell surface epitopes occurs alongside *T. cruzi* metamorphosis. Notably, the switch from insect vector to vertebrate host is accompanied by differential expression of two mucin gene families (14). The *T. cruzi* small mucin-like gene family (TcMUG), a relatively small and conserved group of surface molecules which occur on the CSM of insect stage epimastigotes, is replaced by mucins derived from two much larger and more diverse gene families TcMUCI and TcMUCII in vertebrate stage trypomastigotes and amastigotes. The increased immunological complexity of the mammalian host environment is likely to be one of the driving forces behind this change.

1.1.2 Clinical aspects of Chagas disease.

T. cruzi infection, once established in the mammalian host, is life-long and has a variety of clinical outcomes linked to both host response and parasite (15, 16). The clinical course of the disease is traditionally divided into one distinct and two overlapping phases: acute, intermediate and chronic, respectively. The acute phase occurs immediately after initial infection - in humans living in endemic areas first contact is usually in childhood or early adulthood. Blood and tissue parasitemia may be high throughout the body as the parasite evades the innate immune response and infects a wide range of nucleated cell types. Symptoms of acute infection in humans are not always apparent but can include local inflammation at the parasite entry site, high fever, vomiting, and lymphadenitis. In most cases this phase subsides over the course of 4-6 weeks following the mounting of an effective immune response. Mouse models suggest that both cellular and antibody-mediated immunity are necessary to successfully clear initial infection (17, 18). Mortality during acute-phase infection is rare and most frequent in infants and individuals with primary or secondary immunodeficiency (19). Post-acute phase, the patient develops persistent, asymptomatic infection. The level of circulating parasite is low and undetectable by direct microscopy. Nonetheless, *T. cruzi* is highly immunogenic and significant levels of IgG antibody remain in circulation, facilitating the identification of infected individuals. The intermediate condition may persist throughout the lifetime of a patient without ever leading to the symptoms associated with chronic disease. Importantly, however, the level of circulating trypomatigotes is sufficiently high to allow continuing transmission to the triatomine vector species.

Chagas disease refers to chronic, symptomatic, *T. cruzi* infection, and typically manifests several years, sometimes decades, after the acute phase. The rate of conversion from asymptomatic infection to symptomatic disease has been estimated at ~1% per year, and 30-40% of those infected will go on to develop chronic symptoms (15, 20). Clinical Chagas disease is characterised by damage to tissues in the heart and alimentary tract. The physiological mechanism of *T. cruzi* pathology remains poorly understood. It is clear, however, that both direct parasite-mediated damage and host autoimmune responses have a role to play (15, 21, 22). Heart problems arising from Chagas disease include cardiomyopathy, hypertrophy, arrhythmia, disruption of the autonomic nervous system and associated systolic dysfunction (22, 23). In conjunction

with positive serology, early detection of symptomatic disease can be achieved via electrocardiography and a variety of other cardio-imaging techniques.

Complications of the upper and lower alimentary tract, megaoesophagus and megacolon respectively, are rarer but still important clinical manifestations of chronic Chagas disease. Pathogenesis occurs via the loss of enervation to smooth muscle tissue in the gut, associated muscular hypertrophy and luminal enlargement (24). Heart and gastrointestinal complications in chronic Chagas disease are not mutually exclusive and in some cases patients can develop both. Again, the mechanism behind a patient's predisposition to one or both complications is not fully clear. Limited evidence suggests that gastrointestinal Chagas disease is more common in the Southern Cone region of South America. To date, however, studies have yet to define whether this association is the result of parasite genetics, host genetics or factors unknown (25, 26). A more detailed review of this debate in the light of extant parasite genetic diversity is included in Section 1.8.3.

1.1.3   *T. cruzi*: epidemiology and geographical distribution.

T. cruzi infection is endemic to the Americas. Its known silvatic distribution extends from the southern states of the USA (30°17'N) to southern Argentina and Chile (30°45'S) below ~3000 m (27-29). Active disease transmission is limited to the range of viable triatomine vector species, but due to human migration, sporadic, non-autochthonous cases are occasionally found outside these areas (30, 31). Transmission cycles have been loosely categorised as being either domestic or silvatic in their ecology (32). Domestic transmission occurs when triatomine vector species colonise human dwellings, feeding exclusively on the occupants. Peri-domestic foci – chicken sheds, pig pens etc – act as sites from which traitomine reinvasion can occur after the elimination of intra-domiciliary vectors (33, 34). Domestic transmission does not occur throughout the silvatic range or *T. cruzi*. The socioeconomic status, vector composition and vector behaviour of different regions in the Americas all interact to create hotspots of hyperendemic disease transmission (35).
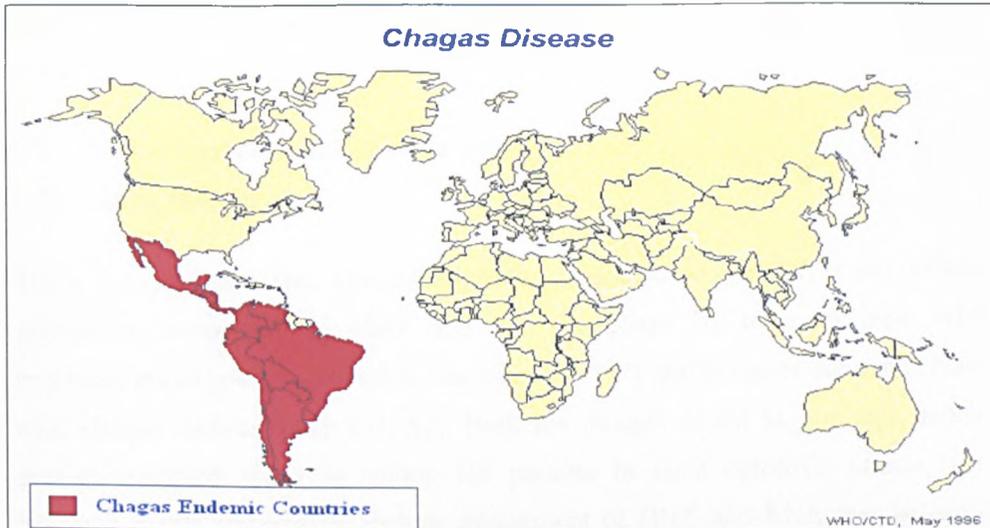
Figure 3 Chagas disease distribution map.
(Source:http://pathmicro.med.sc.edu/parasitology/chagmap1.jpg)

Infection restricted to wild vectors and reservoir hosts is referred to as the silvatic transmission cycle. Sporadic contact may, nonetheless, be made with human populations. Infection can occur orally via the consumption of wild mammals or triatomine-contaminated fruit products of wild origin as well as (more frequently) entry by winged adult silvatic vectors into human habitations (36-39).

T. cruzi transmission cannot always be so easily categorised in the aforementioned manner. As human populations expand beyond their historic boundaries, silvatic vector species are, over time, capable of behavioural and physical adaptation to exploit human habitation. There is evidence to suggest that a number of triatomine species are in the process of doing this (40, 41). Similarly, in areas where domestic and silvatic vector species match, transmission cycles are likely to interlock, providing a considerable challenge for control strategies (42). Finally, recent discoveries in Bolivia suggest that even the most established and widespread domestic vector species (e.g *Triatoma infestans*) have substantial silvatic populations (43, 44).

Although the vast majority of *T. cruzi* infection is vector-borne, transmission can occur congenitally both inside and outside of endemic regions (45-47). The rate of vertical transmission is generally low, but represents a serious public health problem in areas where vector-borne disease has been eradicated. *T. cruzi* can also be spread via blood

30

transfusion and organ transplant where inadequate donor screening systems are in place (48-50).

1.2 Control and intervention strategies.

1.2.1 Drug therapy.

There are currently two chemotherapeutics available to treat *T. cruzi* infection: nifurtimox, introduced in 1965 and now in phase III trials to treat African trypanosomiasis, and benznidazole, introduced in 1971 and the more common choice to treat Chagas disease today (51, 52). Both are thought to act as prodrugs, requiring enzyme-mediated activation within the parasite to elicit cytotoxic effects, which, although poorly understood, include impairment of DNA and RNA metabolism (53, 54). Like many drugs targeted at eukaryotic parasites of human beings, nifurtimox and benznidazole can both have severe side-affects. These include anorexia, central nervous system alterations and allergic dermopathy. Although neither drug is thought to clear systemic parasites in all cases, overall parasite load and antibody titres have been shown to drop considerably over the course of treatment (typically 60 days) and long term studies in mice and humans have demonstrated a reduced likelihood of progression to severe symptoms (55, 56). The considerable time lag between infection and symptomatic disease, as well as the relative recency of drug availability, means that there are still relatively few long-term studies that examine drug efficacy (51). Much work is thus still required to examine the full long term implications of chemotherapy, including the long term effects of drug toxicity to the host and the evolution of drug resistance in the parasite. Additionally, although drug costs are low by comparison to treatments for other trypanosomiases, the duration and toxicity of therapy requires monitoring of the patient during and post-treatment (57). Unfortunately, Chagas disease primarily affects the rural poor, the section of society with the least sustained access to medical professionals.

1.2.2 Vector control.

In the light of the limited success and availability of drugs to treat Chagas disease, control strategies have traditionally focused on blocking transmission by eliminating domestic vector species (58). The importance of this approach was recognised in the early years of the twentieth century, shortly after the description of the disease by Carlos Chagas in Brazil. Before the advent of synthetic insecticides in the 1940s, rural

housing improvement aimed at eliminating intra-domiciliary colonisation sites (cracked adobe walls, palm frond roofs) was the only viable, although costly, solution. Ironically this remains the most effective long-term approach. Post-war, DDT and a variety of organochlorine insecticides were used with considerable success, although effective dosages were high and frequent reapplication was necessary to prevent re-infestation (58). The most commonly used class of insecticides to treat domestic triatomine infestation today, the pyrethroids, came in to use in the 1950s. Effective for longer and at lower doses than DDT and organochlorines, pyrethroids are the mainstay of current vector control programmes (59-62). Residual spraying of domestic and peri-domestic structures to prevent domiciliary infestation as well as bed-net impregnation to block sporadic invasion of silvatic vector species have both been shown to be effective in reducing disease transmission (63, 64).

Despite the recent success of vector control strategies, peri-domestic and silvatic triatomine reinvasion remains a real and present threat (34, 65). Comprehensive spraying strategies and community education can eliminate the risk posed by peri-domestic foci, but no real solution exists to combat that posed by silvatic foci. Local deforestation has been shown to result in a dramatic decline in silvatic triatomine abundance and reservoir mammal infection prevalence, but this has never been a realistic or ethically attractive approach (66).

1.2.3 Transfusional control.

After vector-borne transmission, blood transfusion is probably the second most important route of *T. cruzi* contamination (67). The first stage in blocking transmission involves the screening of potential blood donors, especially from high risk groups (individuals originating from endemic areas), as well as the screening of existing reserves of blood and blood derivatives. Although laws now exist in almost all Latin American countries to ensure mandatory screening, their implementation is often extremely lax (67). Additionally, even if screening is undertaken, multiple procedures are often required to completely avoid the possibility of generating false negatives (68). Once an infected sample has been identified, elimination of parasites by microfiltration (for blood products), exposure to visible light and/or gentian violet is possible (69, 70). The success of such approaches is debatable, however, and early identification and elimination of infected donors is always preferable.

1.3 Chagas disease control initiatives.

Large scale vector control campaigns were in existence, notably in Venezuela and Brazil, as early as the 1970s (58). The first truly national campaign, aimed at total eradication of the domestic vector species, *Triatoma infestans,* was implemented in Brazil in 1983 (71). The Brazilian approach involved widespread mapping of infested localities, comprehensive domestic spraying campaigns and subsequent community-based surveillance to monitor re-infestation. It was soon realised that, in the light of the widespread distribution of many domestically important vector species, an international approach was required. The Brazilian 1983 campaign formed a blueprint for the first international control programme, dubbed the Southern Cone Initiative.

1.3.1 The Southern Cone Initiative (SCI).

In 1991 the governments of Bolivia, Brazil, Argentina, Paraguay, Chile, Peru and Uruguay signed a joint agreement aimed at the total eradication of the principal domestic vector *Triatoma infestans*, the suppression of other domestic vector species and the blocking of transfusion-transmitted disease across the southern cone of South America (72). Each signatory was responsible for financing national control efforts. Annual meetings were held to discuss the overall operating strategy, share methods and outline achievements. Over the next 20 years, with the notable exception of Bolivia (Section 1.5), the scheme achieved considerable success (61). House vector infestation levels and seroprevalence in younger age groups (a key indicator of active transmission) all dropped considerably (Figure 4). In Brazil six and in Argentina ten formally endemic states were declared free of active transmission, while Uruguay and Chile were declared completely free of transmission in 1997 and 1999 respectively.
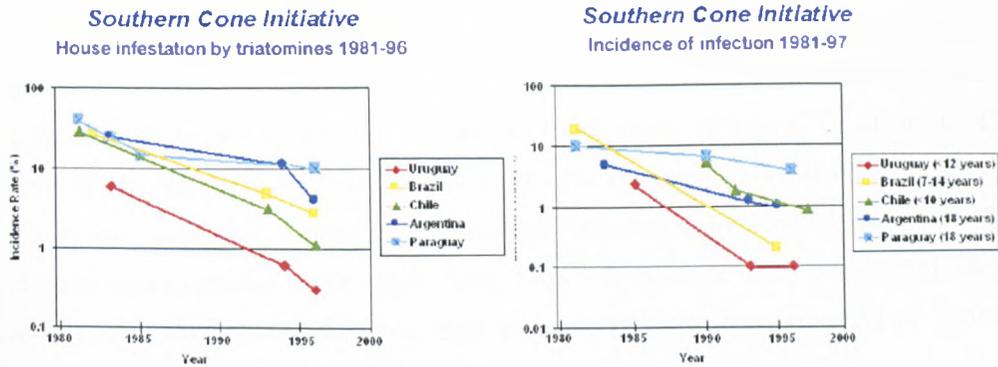
Figure 4 Line plots showing reduction of key Chagas disease transmission indicators over the course of the Southern Cone Initiative. Left: Domestic triatomine infestation. Right: Infection prevalence in younger age group (≤18 years old). Both indicators are associated with active disease transmission. Adapted from http://www.paho.org/english/ad/dpc/cd/incosur.htm.

The financial benefits of the SCI have been considerable (58). Cost benefit analysis undertaken in Brazil suggests that, in terms of reduced disease morbidity and medical costs, US$7.16 has been saved for every US$1 invested.

1.3.2 The Andean Pact Initiative (API) and Central American Initiative (CAI).

Encouraged by the of the success of the SCI, the API and CAI were established in 1997 by the governments of Venezuela, Colombia, Ecuador and Peru, and those of Belize, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua and Panama respectively. Both aim to interrupt domestic disease transmission by 2010 (60). Throughout this region an estimated 5-6 million are infected with *T. cruzi* and a further 25 million at risk (73). Vectorial control in these countries arguably constitutes a greater challenge than that encountered in the implementation of the SCI. Not only is a far more eclectic range of domestic vector species important in disease transmission, but many are also present in silvatic foci throughout the region (62, 65, 74). As such, vectorial control strategies must be modified to identify and counter the high risk of domestic reinfestation in areas where silvatic and domestic triatomine species co-occur (74, 75).

1.3.3 The Amazonian surveillance initiative (ASI).

Established in 2004 by Bolivia, Brazil, Colombia, Ecuador, France, Guyana, Peru, Surinam, and Venezuela, the ASI specifically targets Chagas disease within the Amazon basin (76). While the Amazon basin is not currently endemic for domestic *T. cruzi* transmission, the ASI was established in recognition that, in the light of high

34

immigration and changing environmental conditions (deforestation, land use change), a cooperative surveillance and intervention strategy in the region was vital. Intense silvatic *T. cruzi* transmission (>60% vector infection prevalence) is maintained across the Amazon Basin by a variety of silvatic host and vector species (77). Invasion of houses by adventitious vectors and possible domestic adaptation of local vector species thus carries a high risk of infection for the local population. Interestingly, 111 out of 205 acute cases reported in the region from 1968 – 2000 were orally transmitted, the majority from fruit juices pressed at night and susceptible to contamination by light-attracted silvatic triatomines (38, 78).

1.3.4   The future of Chagas disease control

The success of Chagas disease control strategies over the last 40 years is undeniable. Global disease prevalence estimates of 16-18 million in 1990 have declined to 9.8 million today (79, 80). Such success comes at a cost, however, in the form of reduced political interest in the disease and lower operational budgets allocated to disease control (81). Additionally, the progress made towards disease control is not evenly distributed across all countries. While Brazil, for example, claims to have completely eradicated domestic transmission, Mexico, Ecuador and Panama are still at the earliest stages of disease control (58).

The endpoint of most Chagas disease control strategies is the complete elimination of domestic vector populations. The complex zoonotic ecology of Chagas disease, however, necessitates constant surveillance if this goal is to be achieved in the long term. While the basic infrastructure underlying widespread insecticide spraying campaigns can be downsized once domestic vector eradication has been achieved, national and community-based surveillance systems must remain in place. The efficiency of such an approach relies on effective data collection and collation as well as a remaining capacity for selective interventions, as and when they are required (81).

1.4   Chagas disease in Venezuela

At the time of its description in Venezuela in 1919, active Chagas disease transmission occurred across 750,000 km² (~80%) of the country (82). Vector control strategies over the following 80 years have succeeded in over halving this figure. Currently ~800,000 are thought to be infected with *T. cruzi* and a further 6 million at risk (74, 83). The vast

majority of transmission is now confined to the foothills (500m – 1500m above sea level) of the Northern Andean Cordillera, which includes the states Barinas, Portuguesa, Lara, Cojedes, Yaracuy, Falcon, Carabobo, Guárico, Miranda, the Federal District, Anzoátegui and Monagas (Figure 5) (82). Erratic disease foci also occur across the low-lying 'Llanos' plains (74, 82). Domestic transmission in Amazonas, Bolivar and Delta Amacuru states has not been documented, although silvatic vector species are present (84). The communities worst affected in endemic areas are those of peasant farmers who grow coffee, plantain, manioc and other crops in forested uplands bordering the 'Llanos'.
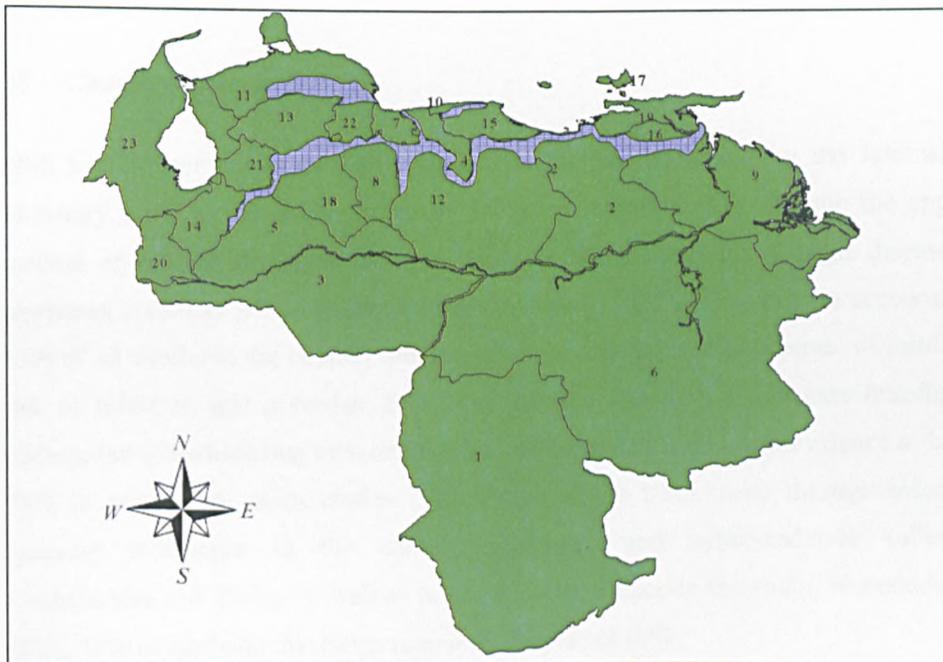
Figure 5 State map of Venezuela - 1: Amazonas, 2: Anzoátegui, 3: Apure, 4: Aragua, 5: Barinas, 6: Bolivar, 7: Carabobo, 8: Cojedes, 9: Delta Amacuro, 10: Distrito federal, 11: Falcon, 12: Guarico, 13: Lara, 14: Merida, 15: Miranda, 16: Monagas, 17: Nueva Esparta, 18: Portuguesa, 19: Sucre, 20: Tachira, 21: Trujillo, 22: Yaracuy, 23: Zulia. Shaded areas represent currently Chagas endemic regions. Adapted from Ache & Matos, 2001(82)

Active domestic transmission coincides with the natural ecotopes of the principal vector species *Rhodnius prolixus*. Two secondary vector species, *Triatoma maculata*, and *Panstrongylus geniculatus* also play a role (41, 74, 85). All three species are present at silvatic foci throughout endemic regions and are known to co-occur domestically and peri-domestically (74, 85). Vector control in Venezuela is hampered by frequent reinvasion of houses from silvatic foci (65). Proximity to palm trees (frequently infested by *R. prolixus* (Section 1.6.1.2)) was identified in a recent study in Barinas state as an

important risk factor for infection (75). The authors also demonstrated that domestic triatomine infestation rates can be extremely low, despite the relative abundance of seropositive individuals. As in the Amazon region, silvatic transmission in endemic areas of Venezuela is often intense, thus adventitious vectors are highly likely to be infected, and could be maintaining human transmission in the absence of significant domestic populations (65, 75, 86). Frequent domestic and peridomestic re-spraying with residual insecticides remains a viable approach to combating Chagas disease in these areas, although further steps must also be taken. These include placing wire gauze across windows and doors as well as the use of insecticide-impregnated bed nets and curtains (75).

1.5   Chagas disease in Bolivia

With a GDP per capita of 3,018 USD, Bolivia and is ranked by the International Monetary Fund as the poorest country in South America (87). Despite the apparent success of control strategies in most Southern Cone countries, Chagas disease still represents a serious public health problem in Bolivia. In 2004 the disease accounted for 13% of all deaths in the country annually with nearly half the population (4 million) at risk of infection and a further 2 million infected (88, 89). Inadequate housing and widespread triatomine bug infestation have corresponded with seroprevalence as high as 74% in some rural communities (90). Progress has been made through insecticide spraying campaigns in the densely populated and hyper-endemic valleys of Cochabamba and Tarija, as well as in the Bolivian Chaco to the south. Nonetheless, by 2001, 70% of Bolivian dwellings remained unsprayed (89).
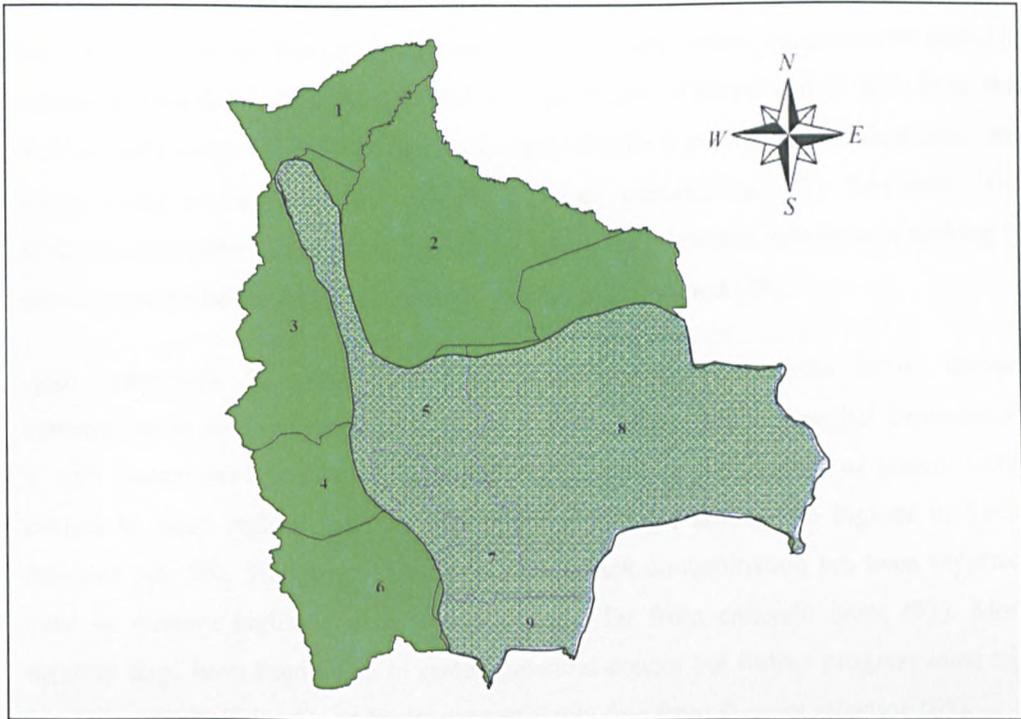
Figure 6 State map of Bolivia — 1: Pando, 2: El Beni, 3: La Paz, 4: Oruro, 5: Cochabamba, 6: Potosi, 7: Chuquisaca, 8: Santa Cruz, 9: Tarija. Shaded areas represent currently Chagas endemic regions. Original figure.

Domestic Chagas disease is endemic to ~60% of the Bolivian landmass (Figure 6) (91). Transmission is most intense in the communities that border the eastern foothills of the Andes (90, 92, 93). This north-south band includes the Yungas region in eastern La Paz state, the valleys of Cochabamba, western Santa Cruz, lower parts of Potosi department, Chuquisaca, and the wine-growing Tarija region. These fertile highland valleys (1200-3000m) are the site of some of the earliest settled agrarian societies in South America (94). They remain the most densely populated and intensively cultivated regions in Bolivia today. *Tr. infestans* is the principal domestic vector, and the presence of local silvatic populations, in conjunction with population genetic data, identify the eastern Andean foothills as one possible origin of this key species (Section 1.6.1.2) (43, 95).

Tr. infestans is undoubtedly the most abundant Chagas disease vector in Bolivia, but a number of other secondary vectors have also been implicated in disease transmission. *Rhodnius stali* was recently identified as an important emergent vector in Alto Beni, capable of colonizing houses and local silvatic palm species, while in Santa Cruz (which includes the southern Chaco region) *Triatoma sordida* has been shown to play a role in disease transmission (40, 96, 97).

38

The sparsely populated western highlands (>3500m) of Bolivia can be expected to be free of vector-borne disease transmission – low mean annual temperatures prohibit triatomine infestation. The lowland eastern departments of Pando and El Beni form the southwestern fringe of the Amazon basin, and, despite a dearth of published data, are highly likely to support active silvatic *T. cruzi* transmission (77). Economic and environmental pressures are now driving communities eastwards, and little or nothing is known concerning the status of domestic disease in this region (94).

High nationwide *T. cruzi* seroprevalence means that non-vector borne disease transmission is an important consideration throughout Bolivia. Congenital transmission is well documented in Cochabamba and Tarija, despite the success of recent vector control in these regions, and rates of transmission are among the highest in South America (46, 98). Similarly, widespread blood bank contamination has been reported, even in western highland departments situated far from endemic areas (93). More recently steps have been taken to screen potential donors but further progress must still be made until Bolivian blood banks are certifiably free from *T. cruzi* infection (67).

1.6   Chagas disease ecology.

T. cruzi infection is a complex vector-borne zoonosis. Several genera and 137 species of haematophagous triatomine vector are known and over half, naturally or experimentally, are capable of transmitting the parasite. (99-102). Infection is maintained in the silvatic environment by a variety of mammalian reservoir hosts. No mammal species, silvatic or domestic, is thought to be refractory to *T. cruzi* infection, but only a limited number are thought to be responsible for the majority of disease carriage (29).

1.6.1   The vector.

1.6.1.1   The taxonomic status of the haematophagous Triatominae

The haematophagous Triatominae are true bugs (Order Hemiptera: Family Reduviidae) and belong to the subfamily Triatominae, one of ~ 30 subfamilies within the Reduviidae, which generally comprise insect-feeding predatory reduviids (102, 103). A robust molecular clock for the group dates their origins to coincide with the final breakup of Gondwanaland 99.8-93.5 MYA (104). Subfamily Triatominae contains six tribes and 18 genera; Triatomini: *Triatoma, Meccus, Dipetalogaster, Mepraia, Eratyrus, Panstrongylus, Hermanlentia, Paratriatoma*; Linschcosteini: *Linshcosteus*; Rhodniini:

Rhodnius, Psammolestes; Cavernicolini: *Torrealbaia, Cavernicola*; Bolboderini: *Bolboderini: Bolbodera, Belminus, Parabelminus, Microtriatoma*; Alberproseniini: *Alberprosenia* (105). Autapomorphies which have traditionally defined the group include their haemotophagous mode of nutrition, an elongate and straightened labium for feeding, as well as the absence of dorsal abdominal scent glands in nymphal stages (101). Aside from these distinct morphological features, the group is exceedingly diverse. Substantial genetic variation, as well as variation in haemostatic salivary secretions and sensorial organisation, has been reported, while several species representing two genera are found outside the New World (105-111). On this basis, the monophyly of the Triatominae is frequently called into question (110). Recent phylogenetic analyses have confirmed, by the inclusion of other non-haematophagous reduviids, that the Triatominae may well be polyphyletic, with haemotophagy arising at least twice during their evolution (Figure 7) (112, 113). Both analyses suggest that the two tribes containing genera of greatest human importance, Triatomini and Rhodniini, emerged independently of one another. Further inclusive analyses of this nature may reveal even greater polyphyly within subfamily Triatominae. After all, the adaptation from a predatory to a haemotophagous lifestyle has many obvious advantages, including a readily available protein rich food source, protection from the elements, and avoidance of seasonal fluctuations in prey populations (110).
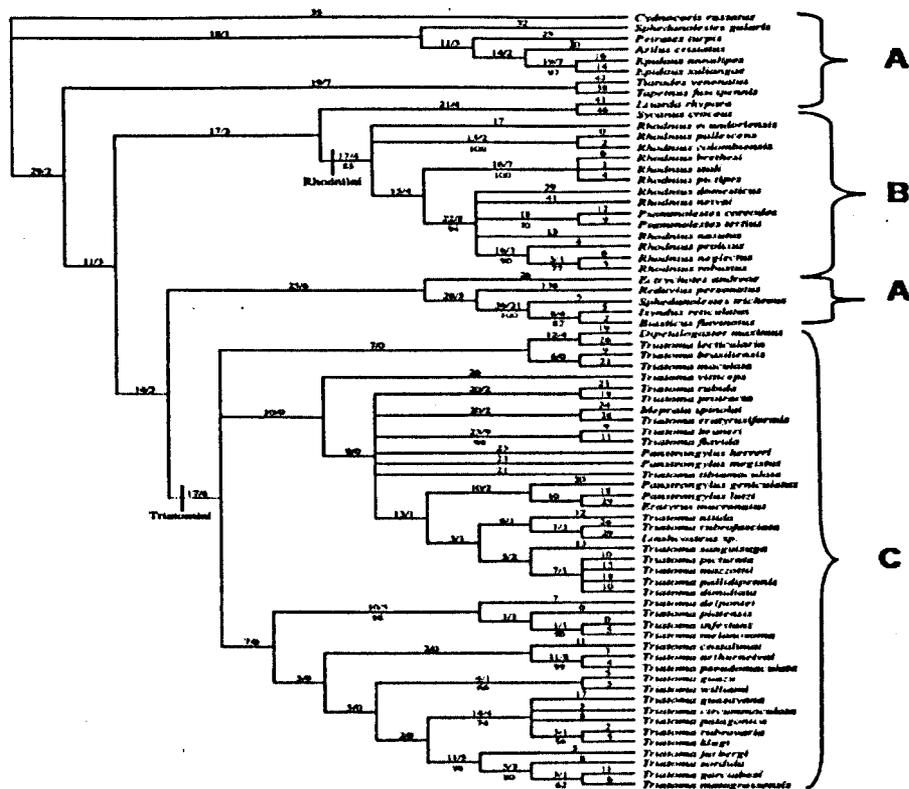
Figure 7 Strict consensus maximum parsimony tree showing polyphyly among the Triatominae based on mitochondrial rDNA sequences. The values above the branches are the branch length/decay index and the values under the branches are parsimony bootstrap values. A – Non-haematophagous reduviids. B – Rhodniini, C- Triatomini. Adapted from de Paula *et al.*, 2005 (113)

1.6.1.2 Domestically important triatomines

Not all triatomines are equally effective at colonising the domestic and/or peri-domestic environment. Most are specialists, exploiting tightly defined ecological niches, the result of ancient associations between vertebrate host and habitat (100). Only a few species have been able to undergo the behavioural and physical adaptation required to capitalise on the recent arrival of humans in South America. A number of features have been identified that characterise domestically important triatomines (101, 114). These include an eclectic host range, anthropophily, adaptation to colonise man-made structures, a relatively short life cycle, fecundity, high dispersal (passive or active), high blood ingestion rate, low host contact time per feeding and a low volume of blood required to support progression from one nymphal stage to another. To a greater or lesser extent all domestically important species must share such characteristics.

The most widespread domestic disease vector, thought to be responsible for almost half all human disease transmission, is *Tr. infestans* (Figure 8) (95). Successfully targeted by recent control campaigns, *Tr. infestans* was once widespread throughout the Southern Cone and as far north as Southern Peru (99). Silvatic populations have been identified in the Bolivian Andes associated with murine and caviomorph rodents among rock piles as well as at sporadic foci in the xerophilous Bolivian Chaco region (43, 44, 115). Population genetic evidence suggests the vector has been passively dispersed by human populations (*Tr. infestans* is capable of only limited independent dispersal), initially in Pre-Incan times throughout the Western Andes, and subsequently, Post-Colombian, eastwards into Argentina, Paraguay, Uruguay, and Brazil (95, 116).

Figure 8 Three important domestic Chagas disease vectors. Left – *Rhodnius prolixus*; centre *Triatoma infestans*; Right – *Triatoma dimidiata*.
(Sources: http://www.weblio.jp/img/dict/esksk/niid/entomology/picture/prolixus/prolixus011.jpg
             http://www.sgfhp.org/kissingbug.jpg
             http://www.rodex.com.mx/fotos/productos/triatoma0.gif)

In Northern and Central South America and southern Mexico, *R. prolixus* and *Triatoma dimidiata* replace *Tr. infestans* as the principal domestic vectors of Chagas disease (Figure 8) (99, 101). *R. prolixus* shares many morphological similarities with the widely distributed silvatic vector *Rhodnius robustus*, and their taxonomic separation has been questioned in the past (117). Genetic analyses have demonstrated that they are distinct species, but that *R. robustus* is paraphyletic, with a clade closely related to *R. prolixus* occurring in the Venezuelan Orinoco region (118). *R. prolixus* and *R. robustus* may, therefore, share a common ancestor in Venezuela. Additionally Venezuela, where this species colonises numerous palm tree species, is the only country within the domestic distribution of *R. prolixus* from which silvatic populations have been identified (42, 65). As with *Tr. infestans* in Bolivia, this is likely to identify Venezuela as the origin of *R. prolixus*. Furthermore, domestic populations of *R. prolixus* throughout North and Central South America show low cytochrome b genetic diversity, suggesting rapid

expansion, possibly coinciding with human demography (118). Recent control campaigns in Central America, where silvatic populations do not occur, have effectively eradicated this species from much of its former distribution (119). By contrast *Tr. dimidiata* is present at silvatic foci throughout much of its domestic distribution and frequent, often seasonal, reinvasion has been reported (36, 120, 121).

Although a few widespread species are responsible for the majority of disease, numerous secondary vector species also play a role in transmission (Table 1). Most cause localised epidemics coincident with their silvatic foci and, with the exception of a few species, widespread passive or active domestic dispersal is rare. In the Southern Cone region, the relatively recent spread of *Tr. infestans* is thought to have displaced many species originally endemic in domestic transmission cycles, notably *Tr. sordida* in Bolivia, Argentina and Paraguay, and *P. megistus* in Brazil (114). Laboratory studies have also demonstrated that *Tr. infestans* is capable of driving *Tr. sordida* to extinction over a six month period when grown together from identical founder populations (122). As domestic *Tr. infestans* populations decline in the face of successful spraying campaigns, the focus of control is now shifting to secondary vector species, which are capable of reinvasion from silvatic foci throughout their historic range (81).

Table 1 Selected secondary Chagas disease vectors.

Vector	Distribution	Epidemiological significance	Reference
Panstrongylus chinai	Ecuador, Peru	Domestic/ Peridomestic	(123, 124)
P. geniculatus	South America (widespread)	Light-attracted/ Domestic/ Peridomestic	(41, 101, 125)
P. herreri	Peru	Domestic/ Peridomestic	(123)
P. megistus	Brazil, Argentina	Domestic/ Peridomestic	(126, 127)
Rhodnius ecuadoriensis	Peru, Ecuador	Domestic/ Peridomestic	(128, 129)
R. pallescens	Panama, Nicaragua	Light-attracted/ Domestic/ Peridomestic	(119, 130)
R. stali	Bolivia	Domestic/ Peridomestic	(40)
Triatoma brasiliensis	Brazil	Domestic/ Peridomestic	(101)
Tr. carrioni	Ecuador, Peru	Domestic/ Peridomestic	(123, 124)
Tr. guasayana	Bolivia, Paraguay, Argentina	Light-attracted/ Peridomestic	(97, 131)
Tr. maculata	Venezuela, Colombia, Brazil	Domestic/ Peridomestic	(65, 132)
Tr. sordida	Bolivia, Paraguay, Argentina	Domestic/ Peridomestic	(96, 127, 133)

1.6.1.3 Silvatic triatomine bugs

Only one fossil haematophagous triatomine has ever been recovered (Section 1.8.2) (134). Nevertheless, molecular and circumstantial biogoegraphic evidence suggest that triatomine bugs date to the formation of South America, 99.8-93.5 MYA (104). A long history of triatomines in South America is supported by their spectacular diversity on

the continent and the exquisite adaptation of many species to their silvatic niches (100). In addition to very close associations between some species and their vertebrate host and niche, broad associations apparently exist between some genera and habitat. Notably, *Rhodnius* species almost exclusively infest palm trees (100, 101). A level of association also exists between certain palm and bug species; *R. ecuadoriensis* primarily infests *Phytelephas aequatorialis*, *R. brethesi* is primarily found in association with *Leopoldinia piassaba*, and *R. colombiensis* seems to prefer *Attalea butyracea* (100, 135). On the other hand, *R. pictipes* and *R. robustus*, two widespread silvatic traitomines, are more eclectic and colonise, often simultaneously, several species of palm (135). The feeding preferences of *Rhodnius* species correspond to their arboreal niche, with didelphid marsupials as well as a number of bird species their most common prey. A further closely related genus within tribe *Rhodniini*, *Psammolestes*, feeds almost exclusively off birds (100).

It has been proposed that, with some exceptions, *Triatoma* and *Panstrongylus* are primarily associated with a terrestrial ecotope (100). A widespread survey of silvatic *Triatoma* habitats found over twenty species associated with terrestrial rodent burrows and/or rocky ecotopes (136). Similarly, with the exception of *P. megistus*, which has been captured in palms and tree holes, most silvatic *Panstrongylus* species have been found infesting mammal burrows, specifically those of Order *Cingulata* (armadillos) (100, 137, 138). In evolutionary terms, *Triatoma* and *Panstrongylus* are much closer to one and other than to *Rhodnius* (105, 107). In the absence of further polyphyly (Section 1.6.1.1), an evolutionary scenario could be envisaged in which the two groups (*Rhodnius* and *Triatoma-Panstrongylus*) radiated to exploit respective arboreal and terrestrial niches early in their history within South America.

While the silvatic associations of the three main triatomine genera associated with human disease are fairly well understood, data on others are fragmentary (100). The niches of a few species have been described (136). These include *Eratyrus mucronatus*, which feeds on the haemolymph of arachnids in its early life stages and later on arboreal mammals; *Belminus herreri* which feeds exclusively off lizards and hides beneath the bark of large forest trees; as well as *Triatoma rubrofasciata* which, although not strictly silvatic, is of note because it is invariably associated with the ship's rat (*Rattus rattus*) and probably owes its presence in the Old World to this phenomenon (100, 101, 139).

1.6.2 The reservoir host

1.6.2.1 Overview of reservoir host diversity, ecology and evolution.

T. cruzi infection is widespread in mammals throughout its silvatic distribution (Section 1.1.3). Parasite has been isolated from numerous species and over 73 genera including members of nine orders: *Artiodactyla* (Even-toed ungulates), *Carnivora* (Carnivores), *Chiroptera* (Bats), *Pilosa* (Sloths and anteaters), *Cingulata* (Armadillos), *Lagomorpha* (Hares and rabbits), *Didelphimorphia* (Opossums), *Primata* (Primates) and *Rodentia* (Rodents) (29, 140).

Geological signatures are easily visible on the global distribution of mammal species, particularly those whose limited capacity for self-dispersal makes them almost entirely reliant on land connections between the continents. The evolutionary history behind the extant diversity of South American mammal fauna is thus tied closely to geological and palaeoclimatological events that occurred in and around the continent over the past 90 million years (141). Approximately 100 MYA, in the mid to late cretaceous, South America had recently broken away from Africa but still remained part of a giant landmass comprising modern-day Antarctica and Australia (Figure 9). Metatherian mammals (marsupials), thought to have first emerged in South America, and basal therian (placental) mammals were widespread (142, 143). As this continent also broke up, mammal populations were separated. 65-80 million years ago the South American continent lost practically all its land connections with its neighbours, save sporadic links with Antarctica via the Scotia island arc, and embarked on c.65 million years of 'glorious isolation' (144).
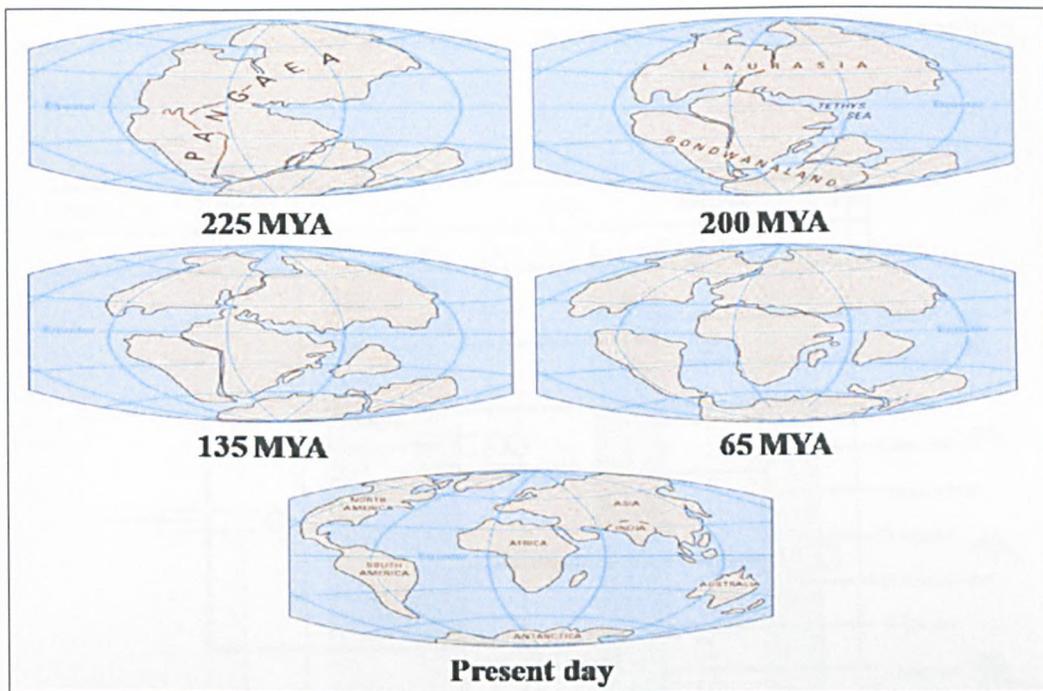
Figure 9 Continental drift from 225 MYA to present showing the almost 65 MYA isolation of South America.
(Source:
http://www.math.montana.edu/~nmp/materials/ess/geosphere/inter/activities/exploration/5globes.gif)

During their isolation South American placental mammals radiated to exploit a full range of environmental niches, while marsupials retained a more modest array of forms (142, 145). The placental group, monophyletic and unique to South America, are collectively termed the Xenarthra and, together with native African fauna (Afrotheria), may be among the most basal of placental mammals (146-148). Of the myriad lineages that evolved over this 65 million year period, only a few remnants survive today. These are the *Edentata*, which include the *Cingulata* (armadillos), the *Vermilingua* (anteaters), and the *Folivora* (sloths). In the face of significant rate variation between lineages, a recent molecular study of three nuclear genes used a Bayesian relaxed molecular clock to time diversification events within the group (Figure 10) (149). The *Pilosa* (Vermilingua /Folivora) appear to have diverged towards the end of the Paleocene, with Myrmecophaga (anteaters) diversifying into the smaller arboreal clade *Cyclopes* and larger, terrestrial clades *Tamandua* and *Myrmecophaga* in the late Eocene (38 MYA). The initial diversification of the modern armadillo lineages appears to have occurred around the same time and points to a very early split between the genus *Dasypus*, the nine–banded armadillos, and the *Tolupeutinae Euphractinae* (149, 150). Also of interest

46

is the recent diversification of the *Euphractinae* in the late Miocene, possibley to exploit dryer climatological conditions throughout Southern Bolivia, Paraguay and northern Argentina (145, 149).
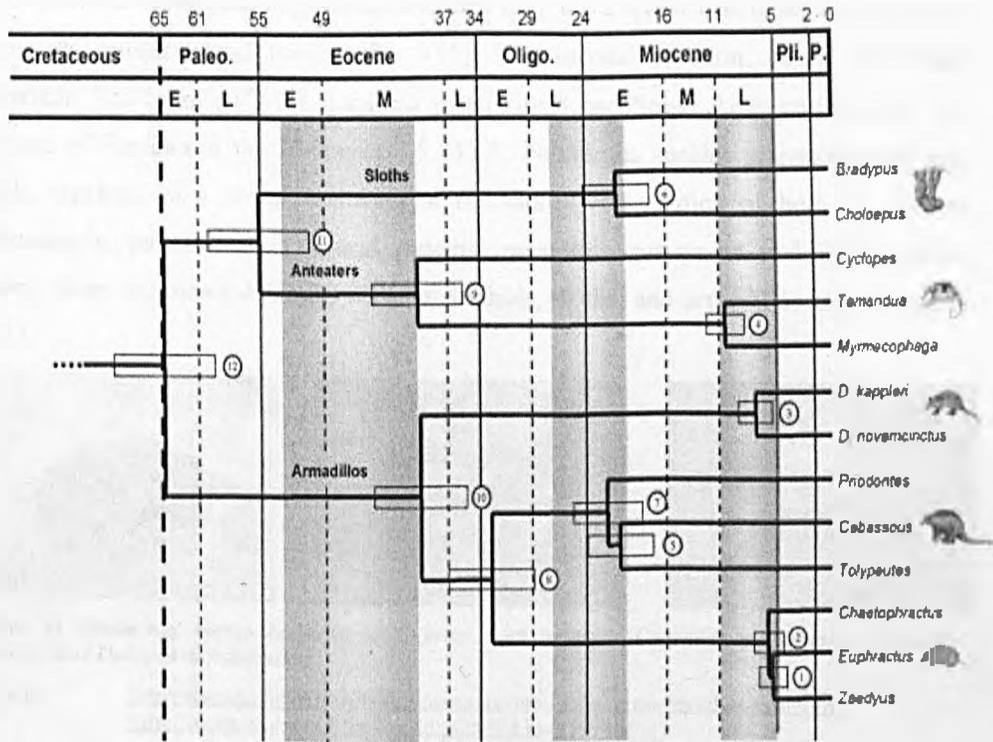
Figure 10 Timing the divergence of the extant Xenarthra using three nuclear genes. Phylogenetic relationships and divergence times are represented by a chronogram, whose branch lengths are proportional to time units. Confidence limits on divergence time estimates are represented by rectangles at nodes corresponding to ± one standard deviation. Tertiary Epoch boundaries follow the 1999 geologic timescale of the Geological Society of America. The Cretaceous/Tertiary transition (K/T) is represented by a vertical dashed bar at 65 Mya. Major Andean tectonic crises 5 are represented by grey shaded areas. Source: Delsuc *et al.*, 2004 (149)

The emergence of the major extant marsupial clades, the *Didelphimorpha*, which encompasses the larger opossums, *Didelphis, Metachirus, Chironectes, Lutreolina*, and *Philander*, as well as the smaller gracile and mouse opossums *Marmosops, Gracilianus* and *Thylamys* may also have occurred around 38 MYA (151). The apparently simultaneous emergence of the didelphids and edentates correlates with climatological and vegetational changes brought about by the rise of the Andes and the formation of the Amazonian river system (152).

From its initial separation from Gondwanaland and subsequent isolation to the present day, two major mammalian faunal invasions took place in South America. The first was that of the new-world primates (the *Platyrhini*) and the hystricognath rodents, of which

47

the latter first appear in the fossil record 31-37 MYA (153). The *Hystricognathi* gave rise to the monophyletic clade. *Caviomorpha*, and molecular analysis is congruent with fossil evidence in dating their emergence 31 MYA while refuting previous claims of a transatlantic invasion and suggesting that they may have actually arrived via Antarctica along the Scotia island route (154, 155). The second invasion, called the Great American Interchange (GAI), occurred when North and South America joined at the Isthmus of Panama in the Pliocene, 2.5 MYA. Numerous species migrated north and south, resulting in a major upheaval in the faunal composition of both continents. Artiodactyls, perissodactyls, murid rodents, mustelids, procynids and felids where among those that headed South, while didelphids, sloths, and armadillos moved North (141).

Figure 11 Three key reservoir hosts of *T. cruzi*. Left to right: *Didelphis marsupialis, Philander opossum, and Dasypus novemcinctus.*

(Sources: http://consejo.bz/belize/images/animals/omnivores/grey-4eyed-opossum.jpg
                           http://cache.eb.com/eb/image?id=62547&rendTypeId=4)

Numerous mammal species carry *T. cruzi*, but the highest prevalence of infection is found among only a few of the more ancient mammalian lineages (29). These include the didelphids, specifically the *Didelphidae* and *Philander*, the edentates, specifically the armadillos, and possibly the caviomorph rodents. The ecologies of the first two groups are fairly distinct. The former, with the exception of the fossorial genus *Monodelphis*, is largely restricted to arboreal or semi-arboreal niches. The latter, armadillos, favor terrestrial habitats, nesting in burrows or among boulders (156). Among these groups the most frequently infected, common and widely dispersed are *Didelphis marsupialis* (the common opossum) and *Dasypus novemcinctus* (the nine-banded armadillo) (29). The ecologies of the *Caviomorpha* are more eclectic and reflect the diversity within the clade. They include the terrestrial *Agoutidae* (agoutis) and *Caviidae* (guinea pigs), the semi-aquatic *Hydrochaeridae* (Capybaras), and the arboreal/terrestrial *Echimydae* (spiny rats).

1.7   T. cruzi: taxonomic status and related species

T. cruzi is a flagellate trypanosomatid parasite belonging to the subgenus Schizotrypanum (Protozoa: Sarcomastigophora: Mastigophora: Zoomastigophorea: Kinetoplastida: Trypanosomatidae: Trypanosoma: Schizotrypanum) (157). Several other species also belong to this group including Trypanosoma rangeli, a non-pathogenic trypanosome of numerous mammals (including humans) in South America primarily transmitted by Rhodnius, as well as some bat trypanosomes, T. vespertilionis, T. dionisi, T. hedricki, T. myoti and T. cruzi marinkelli. The latter is thought to actually be a subspecies of T. cruzi (140, 158, 159). Finally molecular evidence has tentatively placed one Australian trypanosome, isolated from a kangaroo, among subgenus Schizotrypanum, although this sample derives from direct PCR of whole blood and live isolates are required to test the validity of this assertion (159, 160).

T. brucei and Leishmania, the two other trypanosomatid groups responsible for causing human disease, are taxonomically distinct from T. cruzi, and belong to a different subgenus and genus respectively (140). T. brucei belongs to the subgenus Trypanozoon, and, with the exception of T. rangeli, groups phylogenetically with other salivary - transmitted trypanosomes, including T. evansi, T. vivax, and T. congolense (159, 160). Leishmania is a genus in itself, and phylogenetically more distantly related to T. cruzi and T. brucei than they are to one and other.

Parasitism probably emerged many times within the Kinetoplastida (161). Nonetheless genus Trypanosoma is likely to be monophyletic, with Leishmania forming a sister group along with other non-trypanosome Trypanosomatidae (162). While the closest free living relatives of the family Trypanosomatidae are likely to be the Eubonidae, the ancestral route to parasitism is still a matter of debate (157). The paradigm best supported by the molecular evidence involves the accidental transmission of an invertebrate-parasitising kinetoplastid to a vertebrate by a haematophagous insect during feeding (160). This view is supported by the increasingly common isolation of lower trypanosomatids, including those of reptiles as well as a monoxenous dog flea parasite, from immunosupressed HIV/AIDS patients (163).

49

1.8   *T. cruzi*: genetic diversity

1.8.1   Current approaches and outcomes

Genetic diversity within *T. cruzi* populations was first identified by pioneering enzymatic studies undertaken in the late 1970s in north-eastern Brazil (164, 165) Over the following three decades, widespread sampling from silvatic and domestic foci, as well as advances in genetics and genetic analysis, have dramatically improved our overall understanding (99). Progress is still to be made in many areas, however, including the standardisation of genotyping techniques between research organisations, the standardisation of genotype nomenclature, the development of robust *in situ* genotyping techniques, and, importantly, the understanding of the population genetic and evolutionary processes underlying extant parasite diversity. In this section *T. cruzi* genotyping methodologies and their outputs are reviewed to date. A summary of the current genotype nomenclature consensus can be found in Table 2. Both the zymodeme and clonet typing schemes of *T. cruzi* genotype identification were originally defined on the basis of multilocus enzyme electrophoresis (Section 1.8.1.1). In the light of an increasingly confusing array of different typing schemes, a consensus defining six genetic groups was achieved in the late 1990s (166).

Table 2 Summary table of three widely used *T. cruzi* genotype nomenclatures

Genetic group (166)	Zymodeme (Miles) (164, 165, 167)	Clonets (Tibayrenc) (168)
T. cruzi I (TCI)	Z1	1-25
T. cruzi IIa (TCIIa)	Z3	26-29
T. cruzi IIb (TCIIb)	Z2	30-34
T. cruzi IIc (TCIIc)	Z3/Z1 ASAT	35-37
T. cruzi IId (TCIId)	Bolivian Z2	38-39
T. cruzi IIe (TCIIe)	Paraguayan Z2	40-43

1.8.1.1   Multilocus enzyme electrophoresis (MLEE)

MLEE or 'isoenzyme analysis' was first applied to the study of genetic variability in natural populations in the 1960s (169). The technique was adapted to study the genetic diversity of *T. cruzi* in the late 1970s and is still in widespread use today (164, 165, 170). Parasites are first lysed, then spun, and the supernatant, which contains soluble enzymes, collected for analysis. Inter-strain variability is characterised by the differing electrophoretic mobility of the enzyme under study. Although technically phenotypic, the differing charge on enzyme variants can be attributed to amino acid polymorphisms,

50

which have a clear genetic basis. A number of enzymes have been developed to probe *T. cruzi* genetic diversity. These include aspartate aminotransferase (ASAT); glucose phosphate dehydrodgenase (G6PD); glucose phosphate isomerase (GPI); phosphoglucomutase (PGM), 6-phosphogluconate dehydrogenase (6PGD) and malic enzyme (ME).

The presence of distinct circulating isoenzymatic *T. cruzi* groups, termed zymodemes in these early studies, was quickly established (32, 164, 165). Furthermore these groups appeared to have epidemiological significance. In Brazil, zymodeme 1 (Z1/TCI) was principally isolated from silvatic hosts and vectors, including *D. marsupialis* and *R. pictipes*, while zymodeme 2 (Z2/IIb) was found in the domestic environment. Further groups where also identified, zymodeme 3 (Z3/TCIIa) and Z3/Z1 ASAT (TCIIc) from silvatic mammals, including *D. novemcinctus, Nasua nasua* (the coati) and *Monodelphis brevicaudata* (the short-tailed opossum) as well as two heterozygotes from domestic transmission cycles; Paraguayan Z2 (TCIIe) and Bolivian Z2 (TCIId) respectively, defined by multiple heterozygous electrophoretic bands (167, 171).

Remarkably, the strain distinctions made by these early studies have since proved exceedingly robust by comparison to other genotyping methodologies. Studies that employ isoenzyme analysis today are therefore still valid (99, 170). Additionally some studies employing a greater range of enzyme markers have demonstrated their ability to detect variation within the classic Z1, Z2 and Z3 groups (168, 172-174).

Despite its past success, isoenzyme analysis has now been largely replaced by other genotyping techniques. MLEE is highly labour intensive, results are hard to compare directly between studies, good quality quantitative data are difficult to generate (restricting phylogenetic anlaysis), and the phenotypic data generated may not always fully represent the true genotype. The role of post-translational modifications, for example, in influencing the electophoretic mobility of the enzymes in question, has not been adequately addressed.

1.8.1.2   PCR fragment length polymorphism
1.8.1.2.1   Randomly amplification of polymorphic DNA (RAPD)

RAPD genotyping has been applied to several species of parasitic protozoa, including *T. cruzi, Toxoplasma gondii, Leishmania sp., T. brucei* and *Plasmodium spp.* (175-179).

The great advantage of this technique is that informative genetic markers can be obtained in the absence of previous sequence information from the target organism. Short oligonucleotides (8-12 bp) of arbitrary sequence are used to amplify fragments of genome in a standard PCR reaction. Broad primer specificity means that multiple fragments are amplified from each sample. Fragment size profiles can be visualised electrophoretically and compared between isolates.

A reasonably high degree of congruence has been identified between *T. cruzi* isoenzymic zymodemes and RAPD profiles (175, 180). All six major lineages can be distinguished, although the heterozygous nature of TCIId and TCIIe is not apparent. Heterozygotes, however, and putative homozygous parental profiles have been demonstrated among sympatric silvatic TCI isolates by RAPD analysis (181). This, albeit tentative, evidence of active genetic exchange in *T. cruzi* paved the way for numerous future studies (Section 1.8.2) (182, 183).

As with MLEE, RAPD analysis is problematic. The technique is labour intensive, highly sensitive to contamination by foreign DNA (oligonucleotides employed are broadly specific and the risk of cross reactivity is high), a poor source of quantitative data, and results are hard to compare directly between studies. RAPD analysis is still of use as a relatively quick and cheap *in situ* technique, although in reality targeted PCR fragment length polymorphism (henceforth abb. PCR-FLP) and restriction fragment length polymorphism (RFLP) analysis have overtaken it in this capacity.

1.8.1.2.2 Targeted PCR fragment length polymorphism (PCR-FLP)

Targeted PCR-FLP analysis involves the specific PCR amplification of size variable sequence domains within the *T. cruzi* genome. Sequences targeted include some from the kinetoplast DNA as well as non-transcribed regions of the mini-exon, 18s and 24Sα ribosomal RNA gene families (184-188). Mini-exon and rRNA sequences were used successfully in early phylogenetic analyses of the interrelationships between different trypanosomatid species (189). Their use as informative genetic markers to determine *T. cruzi* intraspecific diversity derives from genotype-specific sequence length polymorphisms within these regions. Additionally a number of targets are also able to determine the presence of the closely related trypanosome, *T. rangeli* (185, 187). Unfortunately, no one target is able to distinguish all six lineages as defined by previous

isoenzymatic and RAPD studies. Instead a number of targets are typically used in tandem (e.g. Figure 12)(186).
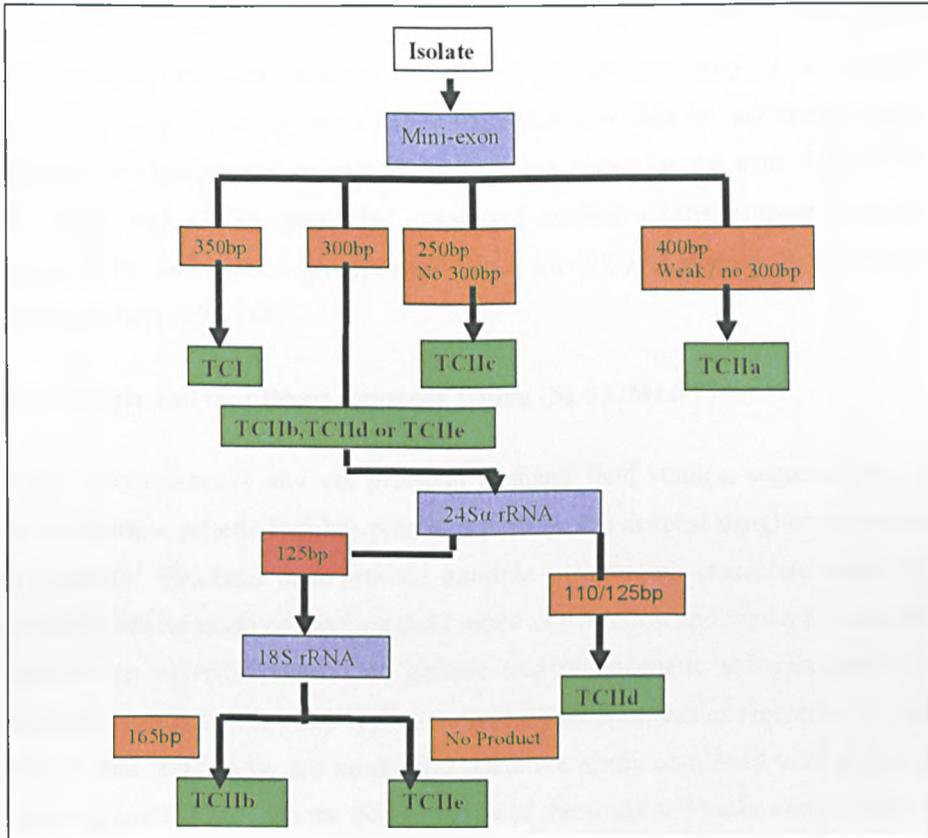
Figure 12 Flow diagram depicting a *T. cruzi* PCR-FLP genotyping strategy. Blue boxes: genomic targets. Orange boxes: PCR – amplified fragment sizes. Blue boxes: *T. cruzi* gentotypes. After Brisse *et al.*, 2001 (186)

PCR-FLP is a relatively cost effective and robust technique for genotyping *T. cruzi* and can be achieved from relatively low quantities of starting material. In addition, the stringency of the relatively long primers used to amplify target regions means that direct genotyping from blood, tissue and triatomine faeces is also possible. (92, 185, 190).

1.8.1.3 Restriction fragment length polymorphism (RFLP)

RFLP, like PCR-FLP, involves the targeted PCR amplification of regions or genes from the *T. cruzi* genome. Although RFLP targets can also be derived empirically through trial and error, the plethora of *T. cruzi* sequence data now available from all six *T. cruzi* lineages has facilitated the identification of fixed single nucleotide polymorphisms (SNPs) between lineages in certain genes (Section 1.8.1.4). Restriction enzymes are

then selected to cut at the informative site and digestion patterns compared between samples to determine lineage (191, 192).

The major draw back of this technique is that a single base mutation can produce a false positive or false negative result. This problem is in part avoided by the choice of conserved target genes and because multiple targets are necessary to resolve all six lineages, providing the researcher with the opportunity to 'flag up' ambiguous samples. Established loci include single-copy house-keeping genes for the heat shock proteins (HSP) *hsp60* and *hsp70*; genes for conserved surface motifs glucose-6-phosphate isomerase (GPI) and surface glycoprotein GP72; as well as a multi-copy mRNA gene, *1f8*, among others (191-194).

1.8.1.4 Single and multilocus sequence typing (SLST/MLST)

Although resource-heavy and not practical in many field studies, sequence typing of single or multiple genetic loci has provided perhaps the deepest insights into extant *T. cruzi* diversity. Sequence data provide multiple informative characters about which assumptions can be made concerning their mode of evolution and which are susceptible to analysis by powerful population genetic and phylogenetic software tools (195). Limitations encountered in other typing methods, including lack of repeatability, lack of resolution, and homoplasy, are minimized. Data are easily combined with pre-existing sequence datasets to broaden the interpretation of the study and make comparisons with related organisms.

Sequence analysis is not without constraints (195). Loci must be carefully selected according to the conclusions that are to be drawn. Significant evolutionary rate heterogeneity exists across the genome of any organism as well as between the genomes of different clades. Mutational saturation at informative sites can lead to homoplasy and the incorrect estimation of genetic relationships between lineages, while inter-clade rate heterogeneity leads to inaccurate estimates of divergence times between them. The presence of multiple copies of many genes complicates the selection of loci that share the same ancestry (orthologues) and can lead to further errors.

Sequence typing shows that *T. cruzi* is exceedingly divergent at a subspecies level, with greater diversity estimated between TCI and TCIIb than that separating five species of *Leishmania*, and four times that of the split between chimpanzees and man (196, 197).

As such, loci selected to examine the deeper roots of the *T. cruzi* tree must be highly conserved to prevent the loss of phylogenetic signal. A number of single locus typing studies have focused on the small subunit rRNA region, known to be highly conserved among many species (196, 198). Highly conserved markers, however, suffer from a lack of informative sites and correspondingly low resolution. More recent studies have sought to circumvent this constraint by generating sequences from multiple, single or low copy number, conserved loci combining both nuclear and mitochondrial targets, a technique similar to MLST (191, 194, 199).

MLST in its true sense was first developed to study bacterial pathogens and is still in widespread use in this capacity (200, 201). Multiple house-keeping genes are sequenced from a representative panel of clones and can be analysed in a number of ways. Firstly, genes can be analysed separately, compared with others at the same locus to identify possible mosaics, congruence assessed between phylogenies generated from different loci, and a super tree accounting for polytomies produced. This technique has been successfully applied to *T. cruzi* and was able to confirm the hybrid status of lineages IId and IIe suggested by other techniques, identifying IIb and IIc as likely parents (Section 1.8.2) (183, 191, 194, 199). Secondly, sequences from multiple loci can be concatenated, boosting the resolution of the analysis. Assuming the rate of recombination between loci is sufficiently low, the resulting phylogenetic tree can be taken as a true representation of the relationships between individuals (202). This technique has yet to be attempted in *T. cruzi* although studies are currently under way (Yeo, M, *personal communication*). Finally, variant sites in a sequence haplotype at a single locus can be designated as an allele and all loci combined to produce an allelic profile for each strain. This technique has been recently applied with some success to diploid organisms and has provided substantial insights into population substructure (203, 204). Again, work is underway to prove its efficacy in examining *T. cruzi* populations.

The principal drawback with using MLST as a means of phylogenetic inference in diploid organisms arises from the impossibility of predicting the full pattern of linkage across haplotypes drawn from different loci. As such, a single concatenated haplotype across several genes may contain a mixture of sequences from either 'parent'. For syntenous loci, multilocus haplotypes can be defined *in silico* using haplotype

simulation software like PHASE(205). However, in *T. cruzi*, where the genome is currently poorly mapped, the prospects for doing so remain distant (6).

1.8.1.5 Multilocus microsatellite typing (MLMT)

Microsatellites are short tandem repeats (STR) consisting of motifs generally 1-6 base pair (bp) in length, tandemly repeated within nuclear DNA e.g. $(GT)_n$ or $(ATA)_n$. These sequences are abundant, ubiquitous, hyper-variable, and co-dominant and have therefore been used in a number of genetic, phylogenetic and population studies in different organisms (206). Estimates from the human genome suggest that microsatellites mutate several orders of magnitude faster than standard sequences (207). As such they have very high discriminatory power, especially when different loci are used in combination (MLMT), and are potentially useful to type related strains that otherwise cannot easily be distinguished with other techniques.

Microsatellites are thought to mutate by a mechanism known as 'slip strand miss-pairing', a process that results in the addition or subtraction of one or more repeat units to/from the sequence (208). While the mechanistic basis of microsatellite evolution is understood, it is not clear how repeat length changes over time. A number of models have been proposed. These fall into two categories: those that address absolute size differences between alleles and assume no underlying mechanism, the Infinite Alleles Model (IAM); and those which assume a step-wise accumulation of additive or subtractive repeat unit changes (SMM) (209, 210). Both models assume that microsatellite loci adhere to strict neutrality, i.e. they are non-coding and therefore not subject to any selective constraints, thus any observed variation is attributable to mutation-drift models alone. The main advantage of the SMM model is that it assumes a linear increase or decrease of repeat number with time, allowing the inference of divergence times between taxa (211). Significant empirical evidence, however, suggests that microsatellite mutations do not always accumulate in a linear fashion, with repeat unit length and population size, among other factors, leading to deviations from this assumption (207, 211). Numerous modifications to the SMM have been proposed to account for this (211, 212). Nonetheless, the SMM still remains highly sensitive to homoplasy, an important phenomenon in rapidly evolving loci, to which the IAM is intrinsically less sensitive. While not permitting divergence time estimates, IAM

distance measures have been shown to be successful in reconstructing phylogenetic relationships between closely related species (211, 213, 214).

MLMT has only been applied to the study of *T. cruzi* population genetic structure in the last ten years. Studies examining natural populations have shown varying degrees of concordance with other typing techniques (215-217). At best, most recent analyses have successfully defined all six lineages as separate clades, and indicate, on the basis of distinct allelic repertoires, that hybrid lineages TCIId and TCIIe may be the result of at least two independent hybridisation events (216). Earlier analyses have been less successful in fully corroborating other typing techniques, although some major lineages can be distinguished, especially TCI and TCIIb (215, 217). These incongruities are likely to arise from a poor choice of mutational model, a limited number of markers, as well as the use of an unrepresentative panel of reference strains. In all cases, however, a high degree of intra-genotype diversity was identified by comparison to sequencing studies (199).

Interestingly, a major observation made by previous microsatellite studies is that of possible multiclonality within strains of the same major genotype (215, 217). Profiles that demonstrated multiple (3+) microsatellite alleles at a single locus were treated as being multiclonal, as after biological cloning no multi-allelic loci were observed. The authors initially claimed that multiclonal strains were absent only among chronic patients by comparison to other sources, an assertion later retracted on examination of a larger panel of strains (215, 217). Additionally, the potential for 'hidden homozygosity' to exist was not examined, such that an apparently heterozygous diploid locus from a non-cloned sample could actually be the product of an equally mixed infection involving two homozygous clones.

1.8.2   Current understanding of *T. cruzi* evolution and diversity

The widespread distribution of *T. cruzi* in the Americas and correspondingly high levels of genetic diversity observed in the parasite all point to a long history on the continent. Evidence of trypanosomes in the fossil record is, as with all other micro-organisms, extremely scarce. PCR-based screening of ≤9000 year old Chilean mummies was able to demonstrate the presence of *T. cruzi* DNA, indicating a long term association with humans in South America (218). The scope of archaeological DNA analysis is, however, limited to thousands of years, depending on the state of tissue preservation

(219). Remarkably, metacyclic trypanosomes have been observed in faecal droplets within fossil amber from the Dominican Republic in association with a 5th stage nymphal triatomine exuvium and bat hairs (134). Dominican amber represents a secondary geological deposit, however, making precise dating problematic. Associated organisms within the same deposit have allowed some tentative dating, and suggest 15-45 MY as minimum age for trypanosomes in South America (220). This date cannot be directly applied to *T. cruzi*, as the age of the specimen precludes genetic analysis.

In the absence of clear fossil evidence, studies have increasingly employed molecular data to study the early evolutionary history of *T. cruzi*. Comprehensive, taxon-rich trypansomatid phylogenies have demonstrated that New World parasitic mammalian trypanosomes are closely related to those from Australian marsupials (Section 1.7) (162). This fits well with the phylogeography of many of the principal mammalian reservoir host species of *T. cruzi* and the pattern of global tectonic movements 65-75 MYA (Figure 9, Section 1.6.2.1). The early trypanosomatid progenitors of *T. cruzi* may have emerged on the South American/Australian/Antarctican super-continent before their subsequent 65 million year isolation within South America. *T. cruzi* is then likely to have diversified to exploit the range of host lineages present during this period, which primarily included early placental and marsupial mammals, followed c.30 million years later by the caviomorph rodents (Section 1.6.2.1).

The extant genetic diversity of *T. cruzi* could provide clues into its evolutionary history within South America. Numerous genotyping methods have supported the existence of six subspecfic groups, TCI and TCIIa-e (Section 1.8.1). A recent SSU rRNA and GAPDH phylogenetic analysis of all available parasitic trypanosomes has demonstrated that, rather than cospeciating with specific hosts or vectors, a phenomenon seen in some parasites and symbionts, these organisms are more eclectic, evolving to exploit a clique of associated vectors and hosts largely governed by ecological niche (195, 221). This trend is termed 'host fitting' and may be applicable to genetic subdivisions within *T. cruzi*. With the exception of some terrestrial rodents found in arid regions, TCI has been shown to be largely associated with arboreal triatomines and mammals, specifically with *Rhodnius* and didelphid marsupials (29, 222, 223). By contrast TCIIc and TCIIa are generally found in the terrestrial niche in association with armadillos, most commonly *Dasypus novemcintus*. Interestingly, TCIIa and TCIIc have been isolated from the terrestrial didelphid marsupial, *Monodelphis brevicuadata*, in this case

suggesting that niche may have a dominant role over host in defining parasite diversity (164, 224).

While silvatic hosts and vectors have been identified, to a greater or lesser extent, for TCI, TCIIc and TCIIa, those of TCIIb, TCIId and TCIIe, all of which are found frequently in domestic transmission cycles, are less well characterised. Phylogenetic analyses of several nuclear and mitochondrial genes suggest a very deep split between TCIIb and TCI/TCIIc/TCIIa, with TCIId and TCIIe as hybrids falling across TCIIb and TCIIc (Figure 13) (191, 196, 198, 199). Divergence estimates between the most distant clades vary between 3-16 MYA and 18-37 MYA (198, 199). Early estimates lead to the erroneous conclusion that TCIIb diverged prior to the isolation of South America, entering recently in conjunction with the caviomorph rodents or even during the GAI via the Isthmus of Panama (196). The absence of this lineage from anywhere outside South America, and its recent isolation from the armadillo *Euphractus sexcinctus* in Paraguay, suggest that it is more likely to be associated with ancient fauna native to the continent (29). TCIIb has also been found in primates (marmosets) in the Atlantic forest region of Brazil, however, some of these primates had been kept in reproductive colonies, with occasional triatomine infestation, and then released (225, 226). As such, some of the primate infections in the Atlantic Forest may be epizootic and do not represent the original sylvatic hosts. The occurrence of TCIId and TCIIe from silvatic transmission cycles is rarer still, although IId has been found associated with armadillos in the Paraguayan Chaco and TCIIe once with *D. marsupialis* in Northern Bolivia (29, 227).
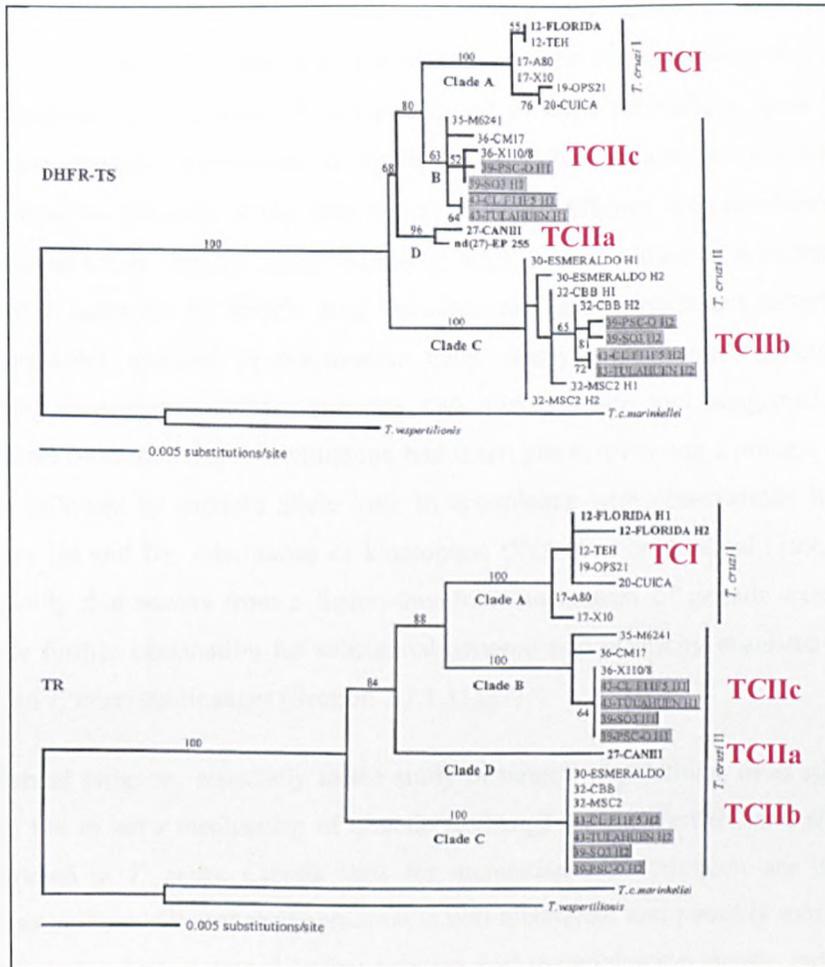
Figure 13 Genealogical relationships among *T. cruzi* strains based on ML analyses of sequences from the nuclear genes dihydrofolate reductase–thymidylate synthase (DHFR-TS) and trypanothione reductase (TR). Grey-highlighted taxa represent haplotypes from hybrid lineages TCIId or TCIIe. Figures above branches indicate % bootstrap values. After Machado & Ayala. 2001 (199)

It is a widely held belief, based on the evidence of strong linkage disequilibrium between numerous genetic markers and the stable genetic heterogeneity observed in natural populations, that *T. cruzi* is a primarily clonally reproducing organism (228-231). Additionally, unlike other parasitic protozoa, including *Plasmodium falciparum* and *Toxoplasma gondii*, there is no evidence that *T. cruzi* has an obligate sexual stage in either host or vector (232-234). The presence of an obligate sexual stage is not, however, prerequisite to genetic exchange. *T. brucei* has recently been shown to undergo meiotic recombination in tetse flies and recombinant genotypes have been identified from field isolates (234-236). Hybrid strains TCIId and TCIIe clearly suggest that recombination has occurred at least once during the evolution of *T. cruzi*. The

extent of current recombination is not fully understood. Putative recombinant and parental RAPD profiles were recovered from silvatic TCI transmission cycles in northeastern Brazil (181). The apparent occurrence of natural recombination in *T. cruzi* and the widespread linkage disequilibrium recorded in most populations were viewed as highly incongruent observations. To investigate this phenomenon, two putative parental clones from the Brazilian study were transfected with different drug resistance markers, co-passaged either through mice, triatomine bugs, axenic culture or a mammalian cell line (183). Selection for double drug resistance revealed recombinant progeny derived from parasites cultured in mammalian cells. Analysis of several genetic markers including isoenzymes, gene sequences and microsatellite loci suggested that non-mendelian (non-meiotic) recombination had taken place, involving a process of genome fusion followed by random allele loss. In accordance with observations from hybrid lineages IId and IIe, inheritance of kinetoplast DNA was uniparental (199, 216). The aneuploidy that results from a fusion-then-loss mechanism of genetic exchange may provide further explanation for substantial genome size plasticity observed within and between *T. cruzi* sublineages (Section 1.1.1.) (237).

Substantial progress, especially in the study of natural populations, must still be made before the *in vitro* mechanism of genetic exchange can be verified and shown to be widespread in *T. cruzi*. Classic tests for mendelian recombination are likely to be fruitless in the event that recombination is non-mendelian and possibly extremely rare. Recent studies have detected further putative past recombination events, indicating that sublineages TCIIa and TCIIc may also be hybrids and that the fusion event(s) that generated TCIId and TCIIe may have occurred in the western Southern Cone, not in Brazil (191, 194, 238). It is clear that mutational processes gradually erase evidence of past recombination but whether the loss of observed heterozygosity in TCIIa and TCIIc is primarily the result of allele loss, as predicted by the current model for genetic exchange in *T. cruzi*, or by gene conversion between homologous chromosomes, remains unproven.
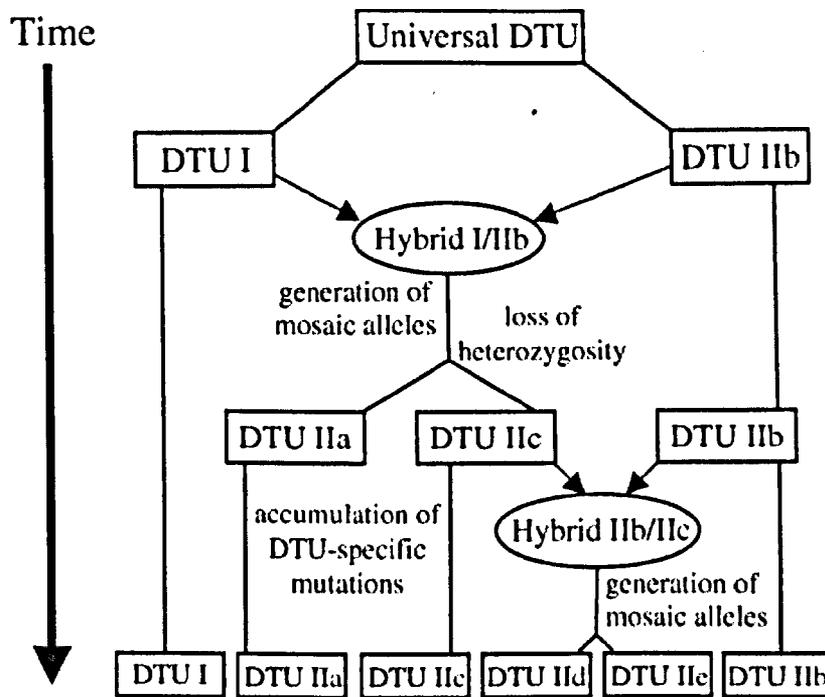
Figure 14 A schematic of recombinatory and mutational processes that may have shaped *T. cruzi* subspecific diversity. DTU: Discrete typing unit. Source: Westenberger *et al.*, 2005 (191).

The fitness advantage of genetic recombination to pathogenic organisms is well established in conferring virulence, metabolic versatility and drug resistance (239). The widespread occurrence of TCIId/TCIIe in human populations across the Southern Cone lead to speculation that heterozygous phenotypes in *T. cruzi* could be similarly advantageous (240). Results from this study and fitness assays in mice models, however, proved inconclusive (241). One study even showed that some naturally occurring hybrid strains were *more* susceptible *in vitro* to treatment with trypanocidal drugs than non-hybrids (242). Comprehensive studies are thus urgently required to provide a clear understanding of the importance of hybrid lineages in *T. cruzi* pathogenesis.

1.8.3 The role of *T. cruzi* genetic diversity in determining disease outcome

Previously it was suggested the relative roles of host and parasite genetics in determining disease outcome in humans were not fully defined in Chagas disease (Section 1.1.2). Two competing hypotheses remain. The first maintains that *T. cruzi* pathology is largely the result of autoimmune damage to affected tissues, especially in the heart and digestive tract (16). There is evidence to suggest that much tissue damage

occurs as a result of the destruction of cells by inflammatory mononuclear cell infiltrates (243). Additionally, the presence of autoantigens in *T. cruzi*, parasite antigens similar to epitopes expressed on host cells, was posited as a possible source of the cross-reactivity of certain antibodies against host tissues (22).

The most frequently cited circumstantial evidence for a role for autoimmunity in Chagas disease is the apparent lack of parasites from histological examination of diseased tissues (16). However, with the advent of more sensitive techniques for detecting parasite in tissue samples, especially PCR, a more direct role for the parasite in pathogenesis has become implicit (244, 245). Additionally, chemotherapeutic intervention, and a concurrent reduction in parasitemia, does result in some reduction in the severity or likelihood of progression to severe disease (246).

While there is reasonable evidence that Chagas disease results at least partly from parasitological factors, the role of parasite genetic diversity in determining the severity of disease is far from clear. The hypothesis that only TCIIb, d and e lineages cause symptomatic disease and that TCI lineages are predominantly associated with asymptomatic infection has been largely discredited (245, 247). Similarly there is no firm evidence that TCIIb, d and e are the exclusive agents of digestive tract megasyndromes and TCI with those of the heart (26, 248, 249). Nonetheless, digestive tract symptoms are more common in the Southern Cone region where TCIIb, d and e predominate in domestic transmission cycles (250). Whether this observation is merely an artefact of more intense domestic transmission in these areas, or the possible result of host genetics, remains to be resolved.

The presence of mixed genotype infections in human patients adds a further complication in establishing a link between parasite diversity and disease outcome. Recent advances have been made, in the form of direct profiling of parasite genotype from infected tissues (251). Here limited evidence differential tissue tropism between major lineages was demonstrated but larger scale clinical studies are now urgently required.

1.9 Population genetics: overview and definitions

In this section the fundamentals of population genetics are considered. More detailed explanations of population genetic theory, tests implemented and software packages used are included in Chapter 3.

1.9.1 Hardy-Weinberg equilibrium

Developed after the rediscovery of Gregor Mendel's 19th century work in the early 1900s by G.H. Hardy and W. Weinberg, the Hardy-Weinberg principle is used to describe the genetic content of diploid populations in terms of allele frequencies (206). The model was originally designed to predict the proportion of two alleles in a population after one generation of random mating (Figure 15). Many genetic loci, however, tend to segregate for multiple alleles in populations. The Hardy-Weinberg principle can be extended to account for this, where the frequency of any allele in a population can be derived as the sum of the frequency of the homozygote and half the frequency of the heterozygote. This extension can be used to calculate the overall expected heterozygosity (H_E) at a single locus in a population adhering to the Hardy-Weinberg principle (Equation 1). Such a population is said to be in Hardy-Weinberg Equilibrium (HWE).

$$p^2 + 2pq + q^2 = 1$$
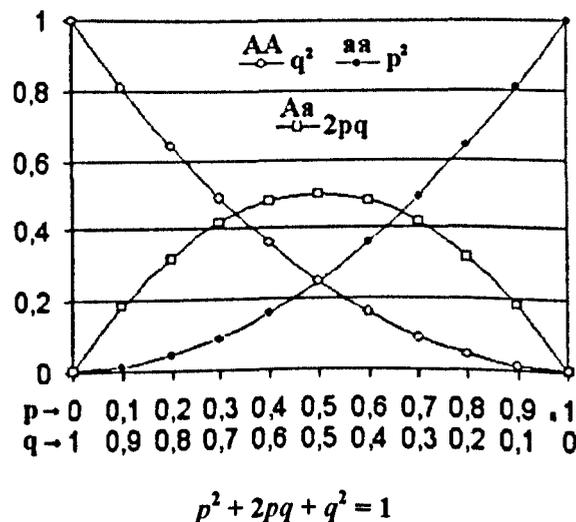
Figure 15 The Hardy-Weinberg principle applied to two alleles. The horizontal axis shows the allele frequencies for p and q, the vertical axis genotype frequencies of homozygotes (AA\aa) and heterozygotes (Aa). Allelic proportions can be summarized in the equation below. After Hedrick, 2003 (206)

$$H_E = 1 - \sum_{i=1}^{k} Pi^2$$

Equation 1 Deriving expected heterozygosity (H_E) of a single locus with multiple alleles at HWE. The sum of the expected frequency of homozygotes (Pi^2) is subtracted from the total frequency of alleles the population (1), to give the expected frequency of heterozygotes. After Hedrick, 2003 (206)

The Hardy-Weinberg principle makes a number of assumptions, many of them fairly unrealistic, about the population under study. The model population is diploid, sexual, exhibits random mating, is infinitely large, suffers no genetic mutation, is under no selection and is subject to neither immigration nor emigration. Deviations from these assumptions will affect the observed heterozygosity (H_O), which can be defined as the number of heterozygotes per locus in a sample from a real population. Various population genetic phenomena can account for significant differences between expected and observed heterozygosity. Excess heterozygosity may result from selection favouring heterozygotes, outbreeding, negative assortative mating and parology. A shortage of heterozygotes may occur through selection against heterozygotes, inbreeding, positive assortative mating or null alleles (e.g. a misidentified microsatellite locus). The applicability of these guidelines to interpreting deviations from HWE in populations of organisms that are either fully clonal, or display non-Mendelian sexuality, is debatable and discussed later in Chapter 3.

1.9.2 Linkage disequilibrium, recombination and the index of association

Under expectations of HWE, the association between any two alleles at different loci within a population should be entirely random, or, in 'linkage equilibrium'(195, 206). If there is a deviation from this expectation and two alleles are linked, physically or otherwise, they are said to be in linkage disequilibrium (LD). LD can be measured, if there is access to gametic phase (haplotype) data, by the term D, among others (Equation 2). If D significantly greater or less than zero, then a degree of linkage does exist between loci.

$$D = P_{AB}P_{ab} - P_{aB}P_{Ab}$$

Equation 2 Calculating the deviation from random association (D) between alleles at two different loci. If all haplotypes are in equal proportion (P), $D = 0$ and the loci are in linkage equilibrium. Positive or negative deviations from zero suggest a degree of linkage disequilibrium. After Page & Holmes, 1998 (195)

Where multi-allelic population genetic data exist for which no haplotype information is available, a second measure of linkage can be derived, the 'Index of Association' or I_A (252). I_A can be calculated from a comparison between observed (V_0) and expected (V_E) values for the variance in the mean difference in number of alleles between two individuals within a population. V_E can be calculated on the assumption that all alleles are independent of one another (linkage equilibrium), if the number of loci and individuals is known. This can be compared to the V_0 which is derived from the data itself. Again, if all loci are randomly associated, then I_A should equal zero. Significant deviation from this value suggests that the population is not truly panmictic and that some linkage between loci exists.

Linkage disequilibrium is a key expectation of a truly clonal organism, where recombination, meiotic or otherwise, does not act to break up associations between alleles. Nonetheless a few other population genetic phenomena have been observed to generate linkage, and must be accounted for before recombination can be ruled out (195, 206, 252). Firstly, genetic drift can act to generate non-random allelic associations is small, isolated populations. Secondly, selection for allelic linkage groups which show epistatic fitness advantage may occur, increasing the overall LD with the populations, even if there is an extant capacity for genetic exchange. Finally population subdivision can lead to the over estimation of LD. If samples are picked from two or more sub-populations biologically, ecologically, or geographically separated, the allelic repertoire between each sub-population will be distinct, even if recombination does occur within each. Thus non-random associations of alleles are identified between some individuals in the global population sample, which actually represent sub-populations, but are misconstrued as circulating clones.

1.9.3 Population sub-structure, gene flow and F-statistics

Natural populations are rarely totally uniform in their distribution and genetic content (206). Most species populations are subdivided through a range of geographical, ecological, and behavioural factors. In Hardy-Weinberg terms, the probability of all individuals in a population mating with one another is rarely ever equal. Frequently of interest to researchers is the level of connectivity between these subpopulations, termed 'gene flow'. Gene flow provides another mechanism, other than mutation, by which new genetic variation can enter a sub-population. It will act to homogenise gene frequencies between sub-populations and inhibit divergence and speciation between sub-populations, while increasing the overall heterozygosity within a sub-population.

Sewall Wright first showed that gene flow and population substructure could be estimated from population genetic data through the hierarchical partitioning of genetic variation observed in a population using F-statistics (253). This technique was later extended by Masatoshi Nei to incorporate multiple loci based on expected heterozygosity (254). Three statistics are typically derived: F_{ST}, which provides an estimation of population subdivision, from which the level of gene flow between populations can be estimated; F_{IS}, which provides an estimate of the level of inbreeding with each respective sub-population; and F_{IT} which provides a measure of both population subdivision and inbreeding (Equation 3).

$$F_{IS} = \frac{\overline{H_S} - \overline{H_O}}{\overline{H_S}} \qquad\qquad F_{IT} = \frac{\overline{H_T} - \overline{H_O}}{\overline{H_T}}$$

$$F_{ST} = \frac{\overline{H_T} - \overline{H_S}}{\overline{H_T}}$$

Equation 3 Calculating F statistics for multiple loci. $\overline{H_O}$ is the average observed heterozygosity within a subpopulation over loci, $\overline{H_S}$ is the average expected heterozygosity within a subpopulation over loci, and $\overline{H_T}$ is the average expected heterozygosity in the total population over loci. After Hedrick P.W. 2003 and Nei M. 1977 (206, 254)

1.9.4 Distance measures

Pairwise genetic distance between individual samples within a dataset have the advantage of allowing the construction of phylogenetic trees from both microsatellite and sequence data. The disadvantage of this type of inference is that assumptions must be made concerning the mode of evolution of different genetic markers. The nature of these assumptions relies strongly on the type of genetic marker employed. Two extreme models have been proposed for microsatellite evolution, the IAM and SMM (Section 1.8.1.5). Within each model a number of measures have been developed to derive genetic distance from microsatellite allele size data (211, 213). Using modelled population genetic data it has been shown that, broadly speaking, IAM distance measures D_S (Nei's standard distance (255)), D_A (Nei's 1983 distance (256)) and D_{AS} (Stephen's inverse allele sharing distance (257)) perform better than SMM measures at recovering true tree topologies, irrespective of the clustering algorithm used (213). All these IAM distance measures make use of the product of the gene frequencies shared between individuals/populations. SMM distance measures, in particular Goldstein's widely used $\delta\mu^2$ (212) tend to overestimate genetic distance with time in heterozygous datasets. Nonetheless they may be useful in estimating divergence times between closely related taxa (Section 1.8.1.5).

Genetic distance measures derived from DNA sequence data must conform to a different set of assumptions altogether (195). A great number of different models have been proposed, the details of which will not be discussed here. All share the same objective in general terms: they aim to model the true number of nucleotide substitutions between sequences. If two homologous sequences differ by few substitutions, the likelihood that multiple substitutions have occurred at the same site (i.e. saturation (Section 1.8.1.4)) is fairly low. As the number of base differences between homologous sequences increases, however, so does the probability of saturation, a phenomenon which must be corrected for by distance measures. Additionally not all base changes occur with the same probability. Purine-purine (e.g. A-G) and pyrimidine-pyrimidine (e.g. C-T) mutations, known as transitions, are more likely to occur than purine-pyrimidine mutations, known as transversions. Finally, within protein-coding sequences, the rate of nucleotide changes that do not affect amino acid composition (synonymous mutations) may well be different to those that do effect amino acid composition (non-synonymous mutations) if the protein is under selection.

1.10 Population genetic models in parasitology.

The advent of multilocus genotyping techniques and associated analytical methods has considerably improved the understanding of parasite population genetics, allowing the formation of generalised models of population structure applicable across a range of pathogenic organisms, including bacteria, fungi and protozoa (Figure 16) (252, 258). Pathogenic micro-organisms have traditionally been classified as either showing predominantly panmictic or clonal population structure (259). Led by substantial research into the population genetics of pathogenic bacteria, it is increasingly apparent that pathogenic organisms actually lie along a full spectrum of population genetic structures between clonality and panmixia (Figure 16) (260).
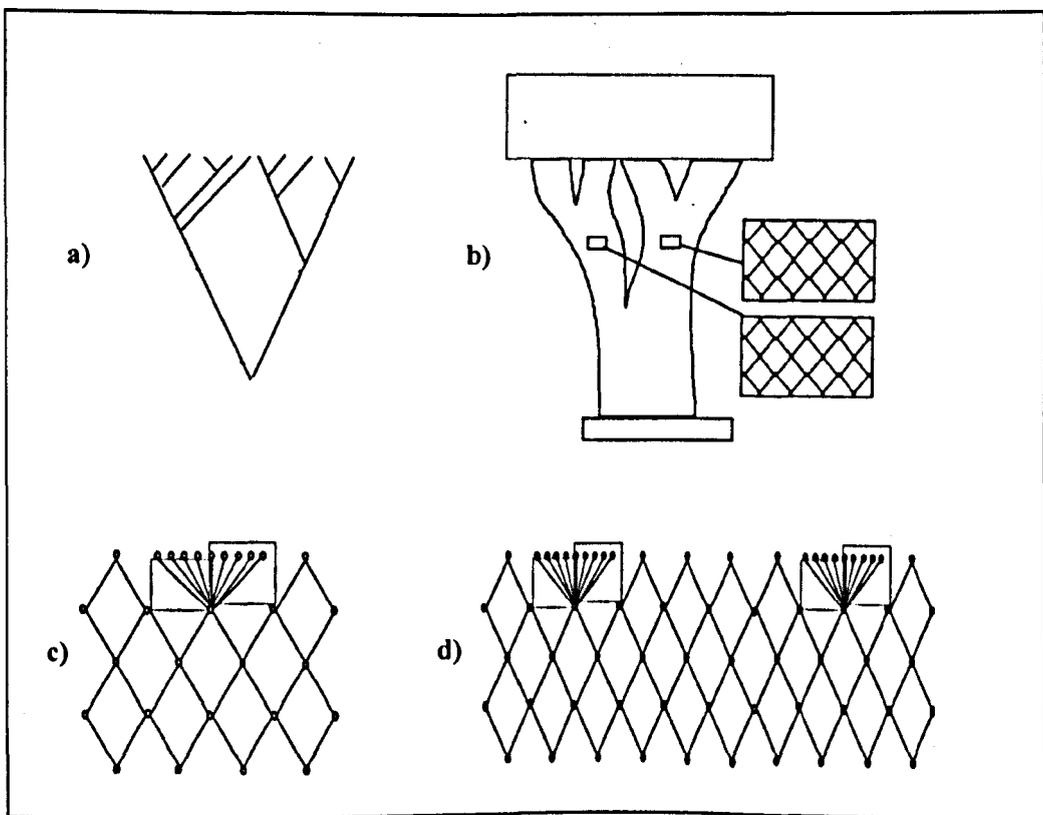
Figure 16 Population genetic models for pathogenic micro-organisms. a) Evolution is exclusively clonal and strong linkage disequilibrium is observed between loci. Extant gene frequencies accurately mirror phylogeny. b) High levels of recombination occur and the population is almost entirely panmictic. Some population differentiation is observed in allele frequencies which results of restricted gene flow due to biological, ecotypical or geographical barriers. c) Population is generally panmictic apart from occasional emergence of epidemic clones that show substantial linkage disequilibrium. d) Multiple epidemic clones emerge from a panmictic population at different endemic foci. After Macleod *et al.*, 2001 (261)

There is substantial population genetic evidence to suggest that many protozoal pathogens, including *To. gondii, T. brucei, Leishmania sp, T. cruzi*, and *P. falciparum* undergo clonal reproduction, in many cases limited to certain populations associated with different hosts and/or geographic regions (230). In all cases, however, evidence for a past or extant capacity for genetic exchange has been demonstrated (Section 1.8.2) (258). The evolutionary explanation for this observation in not clear. Human-associated *To. gondii* isolates are subdivided into three epidemic clonal lineages whose origin is thought to correspond with a switch in transmission strategy to an oral infection route (262). A lack of LD between some loci in Brazilian populations, however, suggests that a degree of recombination does take place (263). LD observed between some physically unlinked loci in the same population, however, suggests that secondary effects, possibly inter-locus epistatic fitness interactions, may be occurring. Epidemic clonality has also been observed in *T. brucei* (Diagram C, Figure 16), and while this could represent a founder affect, epistasis observed in *To. gondii* underlines the possibility that the host or niche may select for particularly virulent strains (236). Instead of allowing the emergence of novel, advantageous genotypes, recombination in these populations may break up advantageous allelic combinations that correspond to infectivity or virulence. Thus, although sex in these organisms may play an important role in generating new genotypes to exploit different hosts or niches, it later becomes largely redundant. This provides a cogent model for *T. cruzi* evolution, where a small number of sexual or parasexual events have given rise to a limited number of predominantly clonal lineages associated with distinct ecological niches (Section 1.8.2). However, some aspects of *T. cruzi* population genetics are at odds with long term clonality, even in populations where widespread LD is observed. Most importantly, a number of recent studies have identified excess homozygosity as a feature (199, 215, 217). As mentioned previously, a standard Mendelian interpretation of this observation, e.g. positive assortative mating or inbreeding, cannot be applied (Section 1.9.1). Furthermore, observations from Bdelloid rotifers, among other species, suggest that diploid organisms exhibiting long term clonality should demonstrate excess heterozygosity as gene copies within the same organism diverge irreversibly in the absence of genetic exchange (the Meselson effect) (264, 265). This is apparently not the case in *T. cruzi* (199, 215). Recent large-scale MLMT analyses of *L. tropica* and the *L. donovani* complex demonstrate that these organisms too demonstrate excess homozygosity in the majority of populations studied (266, 267). Perhaps these data provide circumstantial evidence that some genetic

process, either mitotic or sexual, may be occurring at sufficient frequency in natural populations of all three species to generate deviation from the Meselson effect. Potentially a further, more widespread, mitotic process other than gene conversion could be envisaged. Birky (1996) proposed that serial chromosomal duplication events, followed by chromosomal loss, could result in loss of observed heterozygosity in clonal organisms (268).

2 Aims and Objectives

2.1 Aim

The overall aim of this project is to examine the potential of rapidly-evolving microsatellite markers as a tool to evaluate *T. cruzi* molecular epidemiology in the Americas, with particular emphasis on silvatic transmission, but including some domestic isolates. To this effect, fieldwork will be undertaken to collect new isolates from Venezuela and Bolivia, and basic genotype information combined with a retrospective analysis of all published records of silvatic *T. cruzi* genotypes, to provide a 'state-of-the-art' with respect to the current understanding of *T. cruzi* genotype abundance, diversity, ecology and distribution. By comparison to nuclear sequence data, the power of microsatellites to resolve all six major *T. cruzi* lineages will first be assessed. Latterly, an expanded panel of microsatellites will be developed to examine the extent of diversity within two major silvatic lineages: TCI and TCIIc, again by comparison to nuclear sequence data, and with particular reference to ecological, biogeoraphical, and, in the context of a limited number of domestic TCI strains, epidemiological correlates.

2.2 Specific objectives

- Conduct fieldwork in Venezuela and Bolivia with a view to identifying the prevalence and genotypic diversity of circulating silvatic *T. cruzi* strains.

- Evaluate the status of domestic *T. cruzi* transmission in Northern Bolivia, an area of unknown endemicity.

- Contextualise *T. cruzi* major genotype diversity identified from the field with an exhaustive survey of published silvatic *T. cruzi* major genotype records.

- Develop a panel of microsatellite markers and asses their ability to resolve a) inter-lineage and b) continent-wide intra-lineage (TCI + TCIIc) population genetic structure in *T. cruzi*.

- Undertake single-locus sequence typing of a representative subset of strains for comparison with microsatellite analysis.

- Address the phenomenon of multiclonality in selected reservoir host-derived TCI strains using microsatellites in the context of infrapopulation diversity and possible recombination.

3 Materials and methods.

3.1 Historical data collection and analysis

Subgroup-specific *T. cruzi* genetic diversity data derived from silvatic mammals and triatomines were collated from all available published material as well as from unpublished records held by Miles MA. Data were indexed and catalogued by host, vector, subgroup, date, and geographic origin. Several genotyping methodologies are present in the literature including MLEE, RAPD, and PCR-FLP (Section 1.8.1). To be included in this analysis, results for individual isolates had to fulfil a number of criteria. 1) Importantly, whatever the genotyping methodology, comparison with suitable reference strains had to be made in relation to field caught isolates. Reference strains were as those described in Brisse *et al.*, 2001 (186). 2) Strains included were demonstrably original, and care was taken to avoid the inclusion of re-published isolates. 3) Only a single clone per isolate was included. Data were visualized and mapped using GIS mapping software (ARCVIEW v3.3, ERSI). Isolates for which no, or only national level, geographic location was provided were excluded. Statistical analysis of genotype associations was achieved using Pearson's χ^2 contingency test. A full list of source references is included in appendices.

3.2 Fieldwork and collaborating institutions.

Fieldwork was undertaken in 2004, 2005 and 2007 in Bolivia and Venezuela. Collaborative support in Venezuela was provided by the Universidad Central de Venezuela (UCV) and in Bolivia by the Centro Nacional de Enfermedades Tropicales as well as the Servicio Departamental de Salud, Ministerio de Salud y Desportes.

3.2.1 Study site selection
3.2.1.1 Bolivia

Bolivia is ecologically highly diverse. The Amazon basin lies to the north and the drier Southern Cone region to the south, while to the west the Andean highlands rise up in stark contrast to the moist eastern lowlands. The result is the existence if at least four highly distinct 'ecoregions' within Bolivia, bordered by transitional zones. Detailed maps of these regions is found in Figure 26 and Figure 30 (269). Study sites within Bolivia were chosen to represent this diversity as fully as possible. Selection criteria excluded the high Western

74

Andes and Altiplano in view of the very low likelihood of active silvatic disease transmission in these areas (Section 1.5). In 2004/2005 seven study sites were identified and sampled (Figure 17): San Antonio de Parepeti (20°01'S 63°01'W), Guitterez (19°41'S 63°12'W), and Mora (18°39'S 63°14'W) in the dry southern Chaco region; Quatro Canadas, in the lowland dense dry forests of western Santa Cruz department (17°30'S 61°30'W); Cotopachi, in the dry 'Puna' of the transitional Andean foothills (17°26'S 66°17'W); Montequilla, again a transitional ecotope, in the moist semi-tropical Yungas region occupying the Andean foothills of eastern La Paz department (16°27'S 67°31'W); and San Juan de las Aguas Dulces, situated in the moist tropical lowlands of El Beni department (14°49'S 64°36'W). In 2007 further study was undertaken in El Beni department. Three sites were identified to encompass a greater proportion of that region: San Juan de Mocovi in southern Beni (15°07'S 65°19'W); Santa Maria in northern Beni (14°08'S 65°22'W); and San Cristobal to the northeast (14°08'S 65°55'W). A full description of each study site is included in Chapter 4.
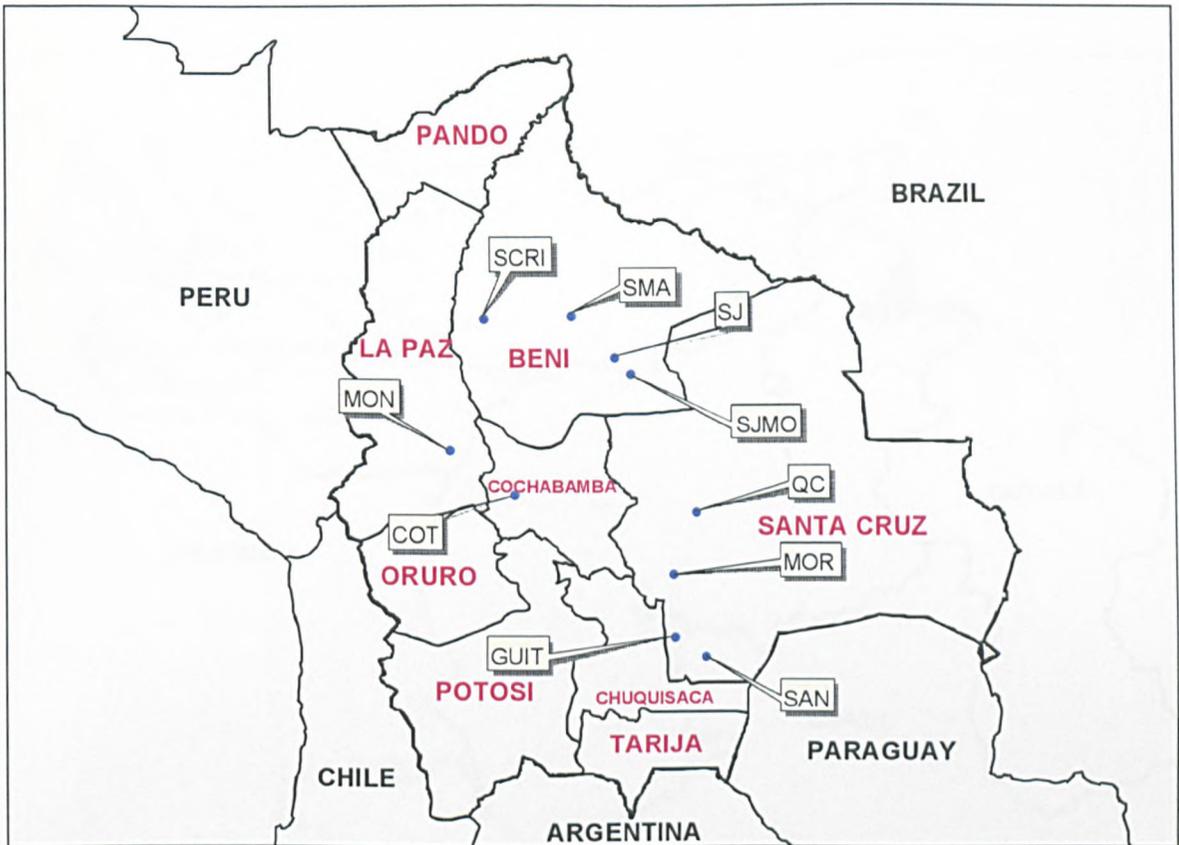
Figure 17 Map of Bolivia with study sites. MON – Montequilla, Dept La Paz; **COT** – Cotapachi, Dept. Cochabamba; **MOR** – Mora, Dept. Santa Cruz; **GUIT** – Guitterez, Dept. Santa Cruz; **SAN** – San Antonio, Dept. Santa Cruz; **QC** - Quatro Canadas, Dept. Santa Cruz; **SJMO** - San Juan de Mocovi; **SMA** – Santa Maria de Apere; **SJ** – San Juan de las Aguas Dulces; **SCRI** – San Cristobal.

3.2.1.2 Venezuela

In Venezuela study sites were chosen to focus on Barinas state, an area known to be endemic for both silvatic and domestic Chagas disease transmission (Figure 18) (65, 82). Barinas borders the northern Andean Cordillera, a mountain range which extends along the western border of the state. The densely forested western foothills give way to seasonally flooded plains or llanos and sporadic dry forest to the east. A detail map of these regions is found in Figure 34. Three study sites were identified in the western region: San Rafael de Catalina (8°30′N 70°44′W); Curbati (8°26′N 70°33′W); and Pinolito (7°30′N 71°14′W) which, although in a western location, was more typical in aspect of the eastern portion of the state. A full description of each study site is included in Chapter 4.

76

Figure 18 Map of Venezuela with study sites. QP – Querbradon del Pescado; SRC – San Rafael de Catalina; PIN – Pinolito. Administrative departments are labelled in red.

3.2.2 Mammal collection

3.2.2.1 Trapping

Small mammals were trapped using collapsible spring-door cage and box traps (Figure 19). A range of sizes was employed: small rodent traps (7.6 x 10x 30.5cm, H. B. Sherman Traps, Inc., Tallahassee USA), larger kangaroo rat traps (7.6 x 10 x 22cm H. B. Sherman Traps, Inc., Tallahassee USA), Tomohawk traps (50.8 x 18 x 18 Tomahawk Live Trap co. Wisconsin, USA) and larger custom-made traps (40 x 40 x 80 cm), based on the Tomahawk design.
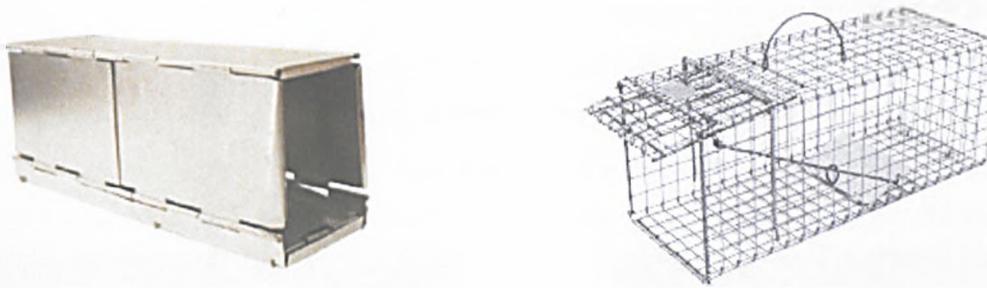
Figure 19 Sherman small rodent (left) and Tomahawk (right) live traps.
(Sources: www.shermantraps.com/images/product/icon/35.jpg
http://www.livetrap.com/cgi/.......shtml&0_option=1&0=204)

Two bait mixes were employed. The first, to attract primates and marsupials, consisted of mashed banana, pineapple and sardines. The second, designed to attract rodents, was adapted from Gascon *et al.*, (2000) and comprised rolled oats, honey and peanut butter (270). Trap stations typically consisted of a large and small trap (to account for size variation in the mammals) within two metres of one another, ten paces apart and hidden by foliage. Trap lines were marked with fluorescent material and laid in linear transects. Pineapple and sardine juice was distributed in the immediate vicinity of trap stations to increase the likelihood of capture. Arboreal trap stations, where set, were set in lines approximately 30 ft above those on the ground and lashed to trees using arboriculture roping techniques. Trap lines were re-baited and reset daily at dusk and checked at dawn. All trap lines were located within a 10 kilometre radius of each community. Trap effort and trap size employed varied considerably between sample sites, largely dependent on local faunal diversity and trap availability. In El Beni, Bolivia and Barinas, Venezuela, on the basis of reported local hantaviral endemicity, small rodents, the known reservoirs of this group of viruses, were not handled and peanut-based bait mixes not employed (271).

3.2.2.2 Capture

Not all mammals were equally likely to be captured in box or cage traps. Burrowing mammals, particularly armadillos, possess powerful forelimbs, and are likely to escape from all but the most robustly designed traps. Such traps (e.g. the Armadillo Live Trap - Tomahawk Live Trap co. Wisconsin) weigh approximately 20 kg, compared to 1.5 kg as

the maximum weight of trap employed in this study, and are thus impractical to transport. The capture of species with idiosyncratic dietary niches, especially myrmecophagous mammals, is similarly unlikely using standard traps. In the light of these limitations, local hunters were hired to capture live mammals at each study site (Figure 20). Appropriate protective precautions were taken to prevent infectious contact between animal and man.

Figure 20 Mammal capture. Left: Excavating an armadillo from its burrow with the aid of local hunters. **Right:** An opossum *(D. marsupialis)* captured in a live trap.

3.2.3 Triatomine collection

Triatomines were collected in a variety of ways. Silvatic triatomine bugs were obtained from their natural ecotopes by microhabitat dissection (272). Among those habitats sampled were animal holes and nests identified during hunting expeditions as well as by tracking mammals after their release using the spool and line technique (273). Palm tree felling and dissection was employed at some sites and live-baited Noireau traps were also used (Figure 21) (274). These traps were constructed from a screw-capped plastic container with wire mesh incorporated into the lid. Over this is placed a fine mesh of synthetic material to prevent ants from entering. A single mouse was placed inside the container with cotton wool as bedding, and a piece of fruit added to provide moisture and nutrition. When sealed, the exterior of the trap was covered with double-sided sticky tape. Triatomine bugs, attracted to heat and carbon dioxide emitted by the mouse, become trapped on the tape. Traps were laid in the evening, collected in the morning and placed in a variety of locations including palm trees, bromeliads, rock piles, tree holes and under fallen logs.

Figure 21 Trapping triatomines. Clockwise from bottom left: Setting Noireau traps in palm crowns; a *Rhodnius* nymph captured by a Noireau trap; a typical dwelling in northern Beni; extensive staining of the inner walls of a dwelling by triatomine faeces.

Domestic and peridomestic triatomines were captured by manual searches of houses and peridomestic structures including chicken coops, pig pens and dog kennels. Additionally, at the beginning of each study period, information and sample flasks were distributed among the community, and any triatomines found by locals collected. In cases where triatomines were suspected to be sequestered deep within adobe walling, an aerosol solution of mild irritant was applied (10% (v/v) N, N-diethyl-m-toluamide (DEET)) to flush the insects out. All searches were coordinated with the relevant community authority.

3.2.4   Isolation of parasite from mammals

Mammals were anaesthetised by the intramuscular administration of Ketamine (100mg/kg body weight), sexed, photographed and an ear snip taken and stored in 70% ethanol (v/v) for possible later analysis as well as to identify recaptures. Blood was taken by cardiac

puncture and 200–300 ul inoculated *in situ* into sterile biphasic culture medium. Each tube contained 3 ml gel base (1.4% blood agar (w/v), 0.5% trypticase (w/v), agar 0.5% (w/v), NaCl 0.6% (w/v), gentamicin 150ug/ml, 5-fluorocytosine 150ug/ml, 10% defibrinated rabbit blood (v/v)) with a 0.75 ml saline overlay (0.9% NaCl, gentamicin 150ug/ml, 5-fluorocytosine 150ug/ml) (275). Tubes were examined at 7, 14, 28, 42, 56, 70 and 84 days for trypanosomes. Positive tubes were sub-cultured.

The parasite was also isolated by xenodiagnosis. Typically six bugs (colony reared *Tr. infestans* in Bolivia or *R. prolixus* in Venezuela), of nymphal stage three or four were used for each animal. Bugs were allowed to feed off the mammal for one hour, kept in well-ventilated, dark, dry, ant-free conditions and dissected four weeks later. All mammals were released alive once sampled.

3.2.5   Isolation of parasites from triatomines

The surface of the triatomine was first disinfected by immersion for ten minutes in White's solution (0.025g $HgCl_2$, 0.65g NaCl, 0.125 ml conc. HCl (sp. gr. 1. 18), 25 ml abs. ethanol, and 75ml H_2O), and then washed in sterile saline overlay (0.9% (w/v) NaCl, 400ug gentamicin, and 400ug 5-flourocytosine). Under aseptic conditions within a laminar flow hood, the triatomine was then transferred to a sterile microscope slide on which had been placed a drop of sterile saline overlay prepared as above. Using fine forceps the gut contents were then drawn out through the apex of the abdomen and homogenised using a blunt microspatula. After the addition of ~100ul more overlay, a drop was examined microscopically to detect the presence of parasite. If flagellates were present, 50-100 ul of homogenate was passaged into culture. Remaining or microscopically negative homogenate was stored for later analysis.

3.3   Reference strains and donated isolates

In addition to isolates obtained from the field, further parasite cultures and whole genomic DNA were also obtained from other sources. In collaboration with Dr. Sebastiao Aldo Valente, silvatic mammal isolates stored in liquid nitrogen at the Instituto Evandro Chagas, Belem do Para, Brazil were recovered, cultured, and whole genomic DNA extracted. Parasite cultures were also provided by Professor Michel Tibayrenc, Institut de Recherche

pour le Developpement, France, and whole genomic DNA by Dr. Patricio Diosque, Instituto de Patologia Experimental, Facultad de Ciencias de la Salud, Universidad Nacional de Salta, Argentina and Dr. Hernan Carrasco, Facultad de Medicina, UCV. In the United Kingdom isolates and reference strains were provided by Professor Michael Miles, Dr. Matthew Yeo and Nidia Acosta. A full list detailing the original references of all the samples used in this thesis is included in the appendices.

3.4 Parasite culture

3.4.1 Liquid culture

Positive biphasic cultures from the field were repassaged in biphasic culture as well as transferred to 10ml liquid RPMI 1640 medium (Sigma) supplemented with 0.5% (w/v) trypticase (BBL), 0.5% (w/v) HEPES, 0.03M haemin, 10% (v/v) foetal calf serum (FCS, heat-inactivated), 2mM sodium glutamate, 2mM sodium pyruvate, gentamicin (150ug/ml) and 5-fluorocytosine (150 ug/ml). Trypanosomes from frozen culture were allowed to thaw at room temperature before inoculation into biphasic and subsequently liquid medium. Once ready for re-storage, 0.9ml of log-phase culture was mixed with 100ul glycerol, cooled to -80°C for 24 hours, before being replaced in liquid nitrogen.

3.4.2 Solid phase cloning

Solid phase cloning was undertaken as in Yeo *et al.*, (2007) (276). Plates were prepared by the addition of 12ml of gel base identical to that described in Section 3.2.4. Log-phase uncloned culture density was measured using a haemocytometer, and $10^3 - 10^4$ cells mixed with 2.4ml of RPMI 1640 medium, supplemented as in Section 3.4.1, and 0.6ml of 3% (w/v) low melting point agarose containing 0.9% (w/v) NaCl. This overlay was poured onto the gel base and allowed to set.

The time taken for clonal colonies to become visible varied between isolates. Generally colonies were visible to the naked eye 2-4 weeks after inoculation. Colonies were picked off into 1-2ml of supplemented RPMI medium using a 200ul pipette tip, before repassage to a larger volume for DNA extraction and/or cryopreservation.

3.5   *T. cruzi* genotyping

Basic *T. cruzi* genotyping was carried out using a combination of PCR-FLP and RFLP techniques. As reported by Yeo (2003), certain PCR-FLP methods proved ineffective at typing some strains, in particular the 18s rRNA PCR-FLP developed by Brisse *et al.*, (2001), as well as the mini-exon PCR-FLP when applied to lineage TCIIc (186, 277). In light of these shortcomings, a new strategy was developed, a schematic of which can be seen in Figure 22. Specific genotyping methods are detailed below.

3.5.1   DNA extraction

Liquid cultures were monitored using inversion microscopy and genomic DNA extracted from 10ml in exponential growth phase (3–5 x 10^7 cells/ml). Cells were harvested by centrifugation at 3000g at 4°C for 10 minutes and washed once in chilled PBS (pH 7.2). High purity genomic DNA was subsequently extracted using a Qiagen DNAeasy® kit following the manufacturer's instructions.

DNA was extracted from triatomine gut homogenate (Section 3.2.5) with DNAzol® solution (Invitrogen) using a modified protocol. 50-100ul of homogenate were mixed with 1ml of DNAzol® solution, inverted twice, and incubated at room temperature for three minutes. DNA was precipitated by the addition of 0.5 ml absolute ethanol (room temperature) and ten further inversions. DNA was then pelleted by centrifugation at 13 000 rpm for 4 minutes, washed twice with 70% ethanol (v/v) before re-suspension in 50 ul 8mM NaOH.
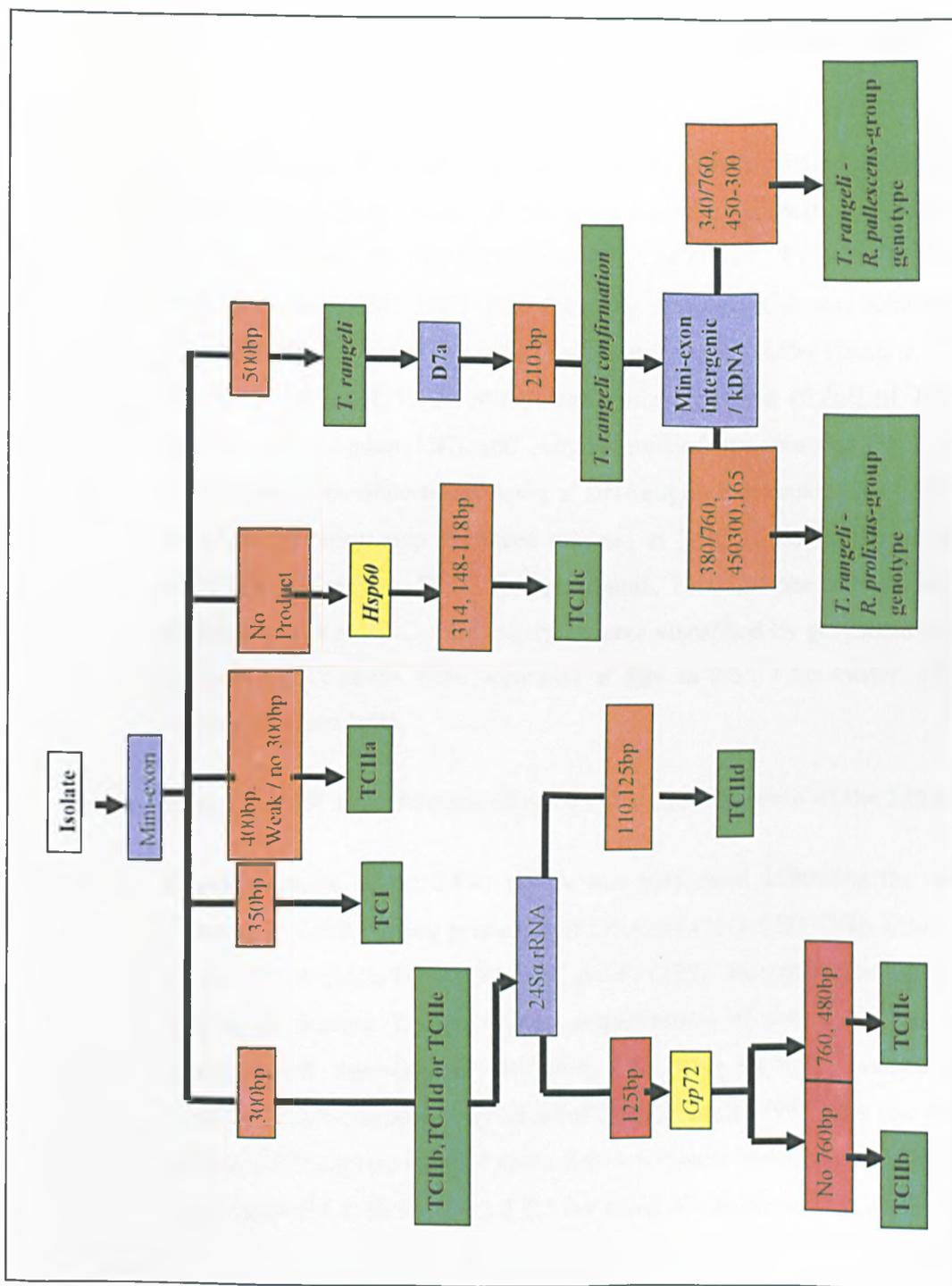
Figure 22 Schematic showing the modified *T. cruzi* genotyping strategy employed in this study: Blue boxes represent PCR-FLP typing systems, yellow boxes RFLP typing systems, green boxes strain identities and orange boxes PCR product sizes. See text for details of specific typing methodology

3.5.2 PCR-FLP typing

3.5.2.1 PCR amplification of the non-transcribed spacer region of the mini-exon gene

Characterisation of the non-transcribed spacer region of the mini-exon gene was performed as described in Souto *et al.*, (1996) using a pool of primers TC (5'-CCC CCC TCC CAG GCC ACA CTG), TC1 (5'-GTG TCC GCC ACC TCC TTC GGG CC), and TC2 (5'-CCT GCA GGC ACA CGT GTG TGT G) (184). Amplification was achieved in a final reaction volume of 20ul containing 1x *Taq* polymerase NH_4^+ buffer (Bioline, UK, 1.5 mM $MgCl_2$, 200uM of each dNTP, 20 pM of each primer, 1 unit (0.2ul) of *Taq* DNA polymerase (Bioline Ltd., London, UK), and 20ng of purified trypanosome DNA. Cyclic amplification of the locus was undertaken using a DNAengine® thermocycler (Bio-Rad, UK) with an initial denaturation step for three minutes at 94°C, followed by 30 amplification cycles (94°C for one minute, 55 °C for one minute, 72°C for one minute) and a final ten minute elongation step at 72°C. PCR products were visualised by gel electrophoresis using a 1.5% agarose gel. Products were separated at 80v in 0.5x TAE buffer, stained, and observed under ultraviolet light.

3.5.2.2 PCR amplification of the D7 divergent domain of the 24Sα rRNA gene

Characterisation of the 24Sα rRNA was performed following the method described by Souto *et al.* (1993) using primers D71 (5'-AAG GTG CGT CGA CAG TGT GG) and D72 (5'-TTT TCA GAA TGGCCG AAC AGT) (278). The amplification mixture was identical to that in Section 3.5.2.1. Cyclic amplification of the locus was undertaken using a DNAengine® thermocycler (Bio-Rad, UK) with an initial denaturation step for three minutes at 94°C, followed by 30 amplification cycles (94°C for one minute, 60 °C for one minute, 72°C for one minute) and a final ten minute elongation step at 72°C. PCR products were separated as in Section 3.5.2.1 but using 4% NuSieve™ (Cambrex, UK) agarose gel.

3.5.3 RFLP typing

3.5.3.1 PCR amplification and enzymic digestion of the *hsp60* gene.

PCR amplification of the *hsp60* gene was carried out as described in Sturm *et al.*, (2003) using primers *hsp60*f (5'- GTG GTA TGG GTG ACA TGT AC) and *hsp60*r (5'- CGA GCA GCA GAG CGA AAC AT) (194). Amplification was achieved in a final reaction volume of 30ul containing 1x *Taq* polymerase NH_4^+ buffer (Bioline, UK)), 2 mM $MgCl_2$, 400uM of each dNTP, 30 pM of each primer, 1.5 units of *Taq* DNA polymerase (Bioline, UK), and 30ng of purified trypanosome DNA. Cyclic amplification of the locus was undertaken using a DNAengine® thermocycler (Bio-Rad, UK) with an initial denaturation step for three minutes at 94°C, followed by 30 amplification cycles (94°C for 30 seconds, 60 °C for 30 seconds, 72°C for 30 seconds) and a final ten minute elongation step at 72°C. Restriction digestion was carried out at 37°C overnight using the enzyme *Eco*RV (Promega, UK) according to the manufacturer's instructions. Restriction products were visualised as in Section 3.5.2.1 using a 1.5% agarose gel.

3.5.4 *T. rangeli* PCR-FLP genotyping

T. rangeli is endemic throughout much of the geographic range of *T. cruzi* (Section 1.7). Three PCR-FLP protocols described in the literature were employed to avoid confusion with *T. cruzi*, confirm the presence of *T. rangeli*, and, in the absence of a suitable panel of reference strains, attempt to identify *T. rangeli* genetic lineage.(158, 185, 279).

3.5.5 PCR amplification of the D7a divergent domain of the large subunit ribosomal RNA gene

Strains identified as possibly being *T. rangeli* by amplification of the non-transcribed spacer region of the mini-exon gene were confirmed by PCR amplification of the D7a divergent domain of the large subunit ribosomal RNA gene as described by Souto *et al.*, (1999) using primers D75 (5'-GCA GAT CTT GGT TGG CGT AG) and D76 (5'-GGT TCT CTG TTG CCC CTT TT)(185). The amplification mixture and conditions were identical to those described in Section 3.5.2.2. Products were visualised as in Section 3.5.2.1 using a 3% agarose gel.

3.5.5.1   kDNA minicircle PCR amplification

kDNA PCR amplification was undertaken as described in Urrea *et al.*, (2005) using a pool of primers: S35 (5'-AAA TAA TGT ACG GGT GGA GAT GCA TGA); S36 (5'-GGG TTC GAT TGG GGT TGG TGT); KP1L (5'-ATA CAA CAC TCT CTA TAT CAG G)(158). Amplification was achieved in a final volume of 20 ul containing 1x *Taq* polymerase reaction NH_4^+ buffer (Bioline, UK)), 200uM dNTPs, 3.5 mM $MgCl_2$ 20 pM of each primer, 1 unit of *Taq* DNA polymerase (Bioline, UK) and 20ng of parasite DNA. Cyclic amplification of minicircle loci was undertaken using a DNAengine® thermocycler (Bio-Rad, UK) with an initial denaturation step for five minutes at 95°C, followed by 35 amplification cycles (95°C for 30 seconds, 60 °C for 30 seconds, 72°C for 30 seconds) and a final five minute elongation step at 72°C. Products were visualised as in Section 3.5.2.1 using a 1.5% agarose gel.

3.5.5.2   Mini-exon gene intergenic region PCR amplification

T. rangeli mini-exon gene intergenic region (TrINT) PCR amplification was undertaken as described by Grisard *et al.*, (1999) using primers TrINT-1 (5'-CGC CCA TTC GTT TGT CC) and TrINT-2 (5-TCC AGC GCC ATC ACT GAT C)(279). Amplification was achieved in a final volume of 20ul containing 1x *Taq* polymerase reaction NH_4^+ buffer (Bioline, UK)), 1.5 mM $MgCl_2$, 200 uM dNTPs; 10 pM of each primer, 1 unit of *Taq* polymerase and 20ng of parasite DNA. Cyclic amplification of the locus was undertaken as in Section 3.5.2.1. Products were visualised also as in Section 3.5.2.1 using a 1.5% agarose gel.

3.5.6   Summary of identification criteria.

PCR product size was estimated by comparison to DNA ladders: Hyperladder I, IV and V (Bioline, UK). Strain identification was achieved by comparison to reference strains X10 Clone 1 (TC1), CAN III (TCIIa), Esmeraldo-Clone 3 (TCIIb), X9/3 (TCIIc), Sc43 (TCIId), CLBrener (TCIIe) and *T. rangeli* (R1271). A summary of the profiles presented by each strain and the extent to which trypanosome species and subgroups are distinguishable are shown in Table 3 and Table 4.

Table 3 PCR product and restriction fragment size criteria for identification of major *T. cruzi* lineages and trypanosome species (*T. cruzi* vs *T. rangeli*). Fragments unique to genotypes and/or species are in bold.

Technique	T. cruzi TCI	TCIIa	TCIIb	TCIIc	TCIId	TCIIe	T.rangeli	Resolution
Mini-exon	350bp	400bp	300bp	None/250bp	300bp	300bp	500bp	TCI/TCIIa/T.rangeli
Hsp60	432-462bp	432-462bp	432- 462bp	314, 118-148bp	432-462, 314, 118-148bp	432-462, 314, 118-148bp	-	TCIIc
24S rRNA	110bp	120bp	125bp	110bp	110/125bp	125bp	-	TCIIa/TCIId
Gp72	440, 760bp	440, 760bp	500, 440, no 760bp	440, 760bp	440, 580, 760bp	440, 480, 760bp	-	TCIIb/TCIId/TCIIe
D7a	250bp	265bp	265bp	265bp	265bp	265bp	210bp	TCI/TCII/T. rangeli

Table 4 PCR product sizes criteria for distinguishing major *T. rangeli* lineages.

Technique	T. rangeli R. pallescens group	R. prolixus group
TrINT	340bp	380bp
kDNA	760, 300-450bp	760, 340-450, 165bp

3.6 DNA sequence analysis

3.6.1 Amplification and sequencing of the glucose-6-phosphate isomerase (*gpi*) gene fragment.

Amplification of a 1038bp fragment of the glucose-6-phosphate isomerase (*gpi*) gene was achieved according to Gaunt *et al.*, (2003) using primers *gpi*.for (5'-CGC ACA CTG GCC CTA TTA TT) and *gpi*.rev (5'-TTC CAT TGC TTT CCA TGT CA) (183). Amplification was achieved in a final volume of 25ul containing containing 1x *Taq* polymerase reaction NH_4^+ buffer (Bioline, UK)), 2 mM $MgCl_2$, 200 uM dNTPs; 25 pM of each primer, 1.25 units of *Taq* polymerase, and 35ng of parasite DNA. The reaction cycle involved an initial denaturation step for five minutes at 94°C, followed by 28 amplification cycles (94°C for 30 seconds, 60 °C for 30 seconds, 72°C for 30 seconds) and a final ten minute elongation step at 72°C. Products were visualised as in Section 3.5.2.1 using a 1.5% agarose gel.

PCR products were purified using a Qiagen MinElute® 96-well purification kit in conjunction with a vacuum manifold according to the manufacturer's instructions. DNA

was prepared for sequencing with a BigDye® v3.1 sequencing kit (Applied Biosystems, UK), also according to the manufacturer's instructions. In addition to forward and reverse external primers, one internal primer was also employed, *gpi.*1 (5'TGT GAA GCT TTG AAG CCT TT), designed using OligoPerfect™ Designer (Invitrogen, UK).

3.6.2 Molecular cloning of sequence haplotypes

Samples demonstrating two or more heterozygous sequence profiles at individual nucleotide sites were cloned individually using the pGEM T easyVector® system (Promega, UK) to empirically derive sequence haplotypes. Owing to the high reported occurrence (c.20% e.g (280)) of artifactual recombinant sequence haplotypes derived from *Taq* DNA polymerase template switching during PCR amplification, ten different clones were sequenced from each sample. Minority recombinant sequence artefacts were identified and excluded from the analysis.

3.6.3 Automated sequence determination, sequence assembly and alignment

DNA sequence was determined using a 48-capillary 3730 DNA analyzer (Applied Biosystems, UK). Sequence output was visualised and consensus sequences derived using BioEdit sequence alignment editor software (Ibis Biosciences, USA). Sequences were aligned using ClustalW multiple alignment software.

3.6.4 Data analysis
3.6.4.1 *In silico* haplotype reconstruction

In silico haplotype reconstruction was achieved using PHASE v2.1 software for comparison with those derived experimentally (See Section 3.6.2) (205, 281). PHASE implements a Bayesian statistical method of haplotype inference using a Markov chain Monte Carlo (MCMC) algorithm. As such, multiple iterations are required to achieve reliable results, whereby the MCMC algorithm converges on the most likely combination of haplotypes and haplotype pairs within the population. To avoid the problem of convergence of the algorithm on non-optimal haplotype identities and/or combinations, phase estimation must include a number of independent runs ($N \geq 5$), each starting from a

different random number seed. The probability of haplotype distributions across different samples can then be compared between runs to check for consistency.

Although the population genetic model used to calculate haplotype distributions implemented by PHASE assumes Hardy-Weinberg allele frequencies, simulations demonstrate that the model is fairly robust when a degree of population substructure exists (281). The efficacy of the model in predicting haplotypes for clonal or partially clonal organisms remains to be formally tested. Furthermore, the accuracy of haplotype estimation relies on the multiple occurrence of individual haplotypes within a dataset. Haplotype estimation for small datasets, or for datasets containing rare haplotypes, has an intrinsic tendency towards inaccuracy.

3.6.4.2 Phylogenetic analysis

Phylogenetic analysis was undertaken from a sequence alignment of all experimentally derived haplotypes. Initially, tree topology was defined using Kimura-2-parameter distances and reconstructed through Neighbour-Joining (NJ) in the PHYLIP v3.67 software package (282). The tree topology was then used, in conjunction with sequence data, to estimate maximum likelihood parameters and branch lengths, as in Machado & Ayala 2001, but using baseml, part of the paml v4 package (199, 283). The most appropriate mutational model can be derived in baseml by comparing the likelihood of different parameter combinations, and selecting the model with the highest (least negative) -ln likelihood score. A thousand bootstrapped datasets were generated in SEQBOOT, analysed using Kimura-2-paremeter distances, and the resultant trees assessed for congruence in CONSENSE, all in PHYLIP v3.67 (282).

3.6.4.3 Breakpoint analysis

Sequences were statistically screened to detect possible recombinant regions and/or break points using MaximumChi, Chimera, and Bootscan tests, all implemented in the RDP2 software package (284). MaximumChi and Chimera are both based upon a technique first described by Maynard-Smith (1992) to detect mosaic gene sequences (285). MaximumChi is limited to the analysis of pairs of sequence and exhaustively works through every pair in the dataset. Once invariable nucleotide sites have been excluded, a χ^2 contingency table is

constructed describing the variance between the two sequences either side of a virtual partition along the sequence. A separate χ^2 statistic is calculated for a virtual partition at each consecutive site. At a breakpoint, the variance between the two sequences on either side partition theoretically reaches a maximum, resulting in a maximum value for χ^2 that can be compared to others at different sites by plotting the statistic against nucleotide position. A modified strategy is employed by Chimera, but also involving a multiple χ^2 contingency table test (286). In this case sequence triplets are scanned exhaustively with the aim of identifying parent and hybrid sequences.

Bootscan employs a tree-based strategy to detect recombination (287). A 'sliding-window' of fixed width (e.g. 250 bp) is moved along the sequence alignment. For each sequence fragment a NJ tree is constructed and bootstrapped in PHYLIP. Incongruence with >70% support between the tree topology defined from each section of sequence is taken as possible evidence of recombination.

3.7 Microsatellite analysis

3.7.1 Screening the *T. cruzi* CL Brener strain (TCIIe) genome sequence for microsatellite repeat sequences.

Microsatellite repeat sequences were identified from the draft, unpublished, sequence of the *T. cruzi* genome by MWG, available at www.gcncdb.org. Four mega bases of sequence, including at least 13 syntenous sequence fragments from both coding and non-coding regions were screened for di- and tri-nucleotide repeat sequences. To achieve this objective, a pattern matching script (regular expression) was written in SED. An extension of the algorithm was also included to extract the up and downstream flanking regions of the microsatellite sequence (~200bp). Primer design was achieved in PRIMER (Daresbury Laboratories, UK)

3.7.2 Testing the microsatellite panel across TCI and TCIIc

In collaboration with MWG, a representative subset of TCI and TCIIc samples was used to test the efficacy of microsatellite loci identified. Using a standard PCR protocol - 20ul reaction volume: 1x ThermoPol Reaction Buffer (New England Biolabs (NEB), UK), 4 mM MgCl$_2$, 200 uM dNTPs; 10 pM of each primer, 1 unit of *Taq* polymerase (NEB, UK),

and 20ng of parasite DNA – 200+ candidate loci (including those from previous publications) were amplified (183, 215). After modification of annealing temperatures to balance stringency with PCR product yield, the following reaction cycle was implemented across all loci: a denaturation step of 4 minutes at 95°C, followed by 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. Products were visualised as in Sections 3.5.5.2 and 3.6, using 1.5% agarose gel. Loci that failed to amplify for either TCI or TCIIc were discarded, resulting in a panel of 86 working loci suitable for use in subsequent analysis.

3.7.3 Generation of microsatellite panels for multiplex analysis

Approximate fragment sizes derived in Section 3.7.2 were used to guide the array of fluorescent dye labels chosen for the forward flanking region primer used to amplify each of the eighty-six robust loci identified. Five fluorescent dyes were employed – 6-FAM and TET (Proligo, Germany) as well as NED, PET & VIC (Applied Biosystems, UK). Primer stock solutions were arranged in a 96-well plate format for transfer to reactions in order to maximise throughput. Amplification of microsatellite alleles occurred in separate reaction wells prior to the multiplexing of up to six reaction products per capillary for subsequent fragment analysis.

3.7.4 Microsatellite PCR amplification reaction conditions for fragment analysis.

Reaction conditions for robust loci identified after screening across TCI and TCIIc were similar to those described in Section 3.7.2. To achieve optimal PCR product concentration for fragment analysis, and to avoid laborious dilution of PCR products, primer and dNTP concentrations were minimised. The resulting amplification mixture, with a final volume of 10 ul, was as follows: 1x ThermoPol Reaction Buffer (New England Biolabs (NEB), UK), 4 mM MgCl₂, 34 uM dNTPs; 0.75 pmols of each primer and 1 unit of *Taq* polymerase (NEB, UK). The quantity of DNA added was also reduced, in this case to 1ng/reaction, in light of the large number of loci to be amplified per sample.

Loci which failed to amplify across only a limited subset of samples were re-run using a modified amplification protocol employing a stepped annealing temperature in order to maximise the quantity of PCR product generated. The modified protocol was as follows: a

denaturation step of 4 minutes at 95°C, followed by two initial high stringency amplification cycles (95°C for 20 seconds, 57°C for 20 seconds, 72°C for 20 seconds) and a further 26 low stringency cycles (95°C for 20 seconds, 54°C for 20 seconds, 72°C for 20 seconds) as well as a final elongation step at 72°C for 20 minutes.

3.7.5   Automated fragment analysis.
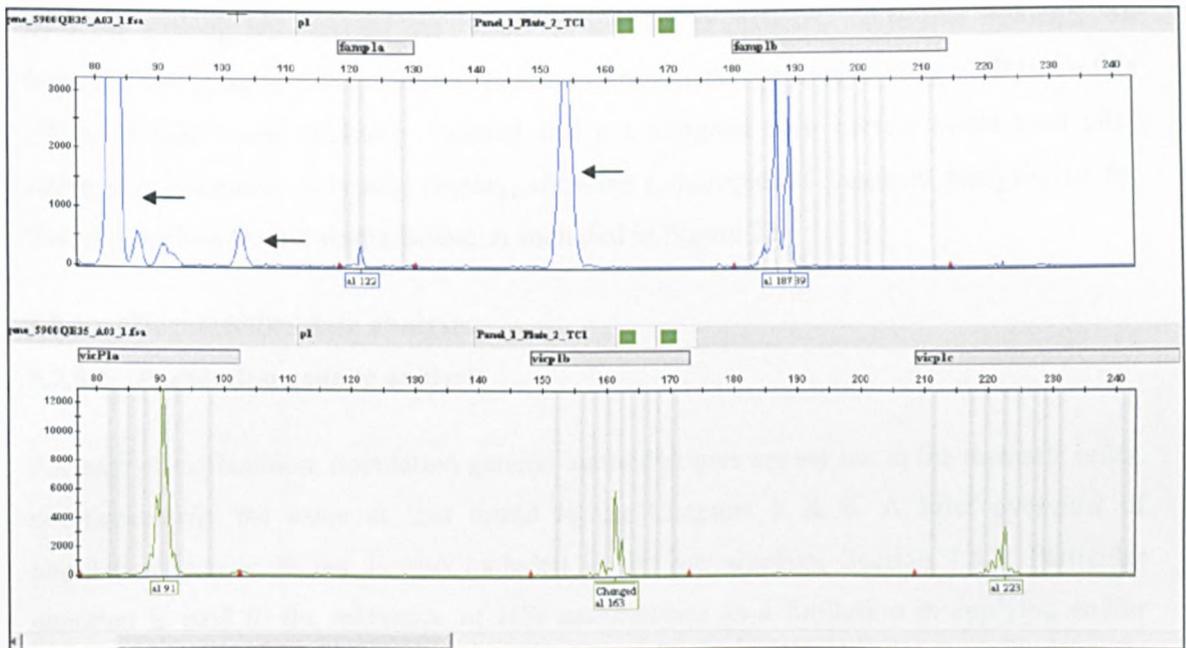
Figure 23 Typical output for multiplexed allelic fragment-size analysis in GeneMapper (Applied Biosystems, UK). Allelic size profiles are displayed for five loci, including one heterozygous locus, indicated by two distinct peaks. Vertical axes represent peak intensity (concentration of product), horizontal axes product size (base pairs). Colours indicate dye types (Green=VIC, Blue=6-FAM). Multiplexed loci were chosen such that allelic profile size ranges did not overlap for those PCR products tagged with the same fluorescent marker. Allelic size scoring was achieved by assigning peaks to specific 'bins' (vertical grey bars). Manual editing of all profiles was necessary to avoid confounding stutter patterns and sequencing artefacts (indicated by black arrows).

Fragment analysis was achieved using a AB3730 48 well DNA sequencer (Applied Biosystems, UK). Multiplexed fluorescently labelled PCR products were loaded into each capillary in conjunction with a G500 LIZ (Applied Biosystems, UK) size standard. PCR products were automatically sized according to the LIZ size standard in GeneMapper (Applied Biosystems, UK). Allelic size scoring, carried out automatically in the GeneMapper software, was also checked manually for errors and inconsistencies associated with abnormal stutter patterns and other artefacts. As a control, a subset of isolates from each genotype was subjected to two or more analyses on different sequencing runs, and a

high degree of consistency in allele size calls observed. Where non-stepwise allele size shifts were observed (e.g. 1bp), the consistency of this allele was checked across the dataset as well as with controls. Single base-pair shifts, when not due to sequencing error, can represent insertions in the flanking region. Equally, while the microsatellites extracted from the ClBrener (TCIIe) gene sequence were apparently perfect, it is possible that occasional imperfections (e.g. Poly-A repeats) may exist at the same loci in TCI and/or TCIIc. Finally, all allele scoring was carried out 'blind' to prevent prejudicial, subjective influence on fragment sizing decisions, a common practice in microsatellite analysis (e.g. (288)). To this effect, samples were randomly recoded and not assigned their correct codes until after sizing was complete. A typical display, showing a multiplexed fragment analysis run for five distinct loci from a single isolate, is included in Figure 23.

3.7.6   Microsatellite data analysis
3.7.6.1   Population genetic analysis

For ease of explanation, population genetic methodologies are set out in the thematic order, not necessarily the same as that found in the Chapters 5 & 6. A brief overview of population genetic theory is also included in the introduction, Section 1.8.3. Particular attention is paid to the relevance of HW assumptions as a limitation in applying and/or interpreting some tests. A summary, including tests for which the assumption of HWE is of a particular importance, is supplied in Table 5 at the end of this section.

3.7.6.1.1   Multi-clonality vs aneuploidy.

As detailed in later sections (Chapters 5&6) a number of isolates displaying multiple (3+) alleles at one or more loci were identified (Figure 24). This phenomenon is inconvenient in terms of standard population genetic analysis. Most software packages are limited to the analysis of diploid and/or haploid organisms only (289). The relative possible roles of aneuploidy and multiclonality in generating this phenomenon are a subject of discussion in later chapters (Chapters 5&6). In order to circumvent practical analytical problems arising from multiple alleles, and to assess their impact, as far as possible, on the inference on population genetic structure from the dataset, a software approach was developed.
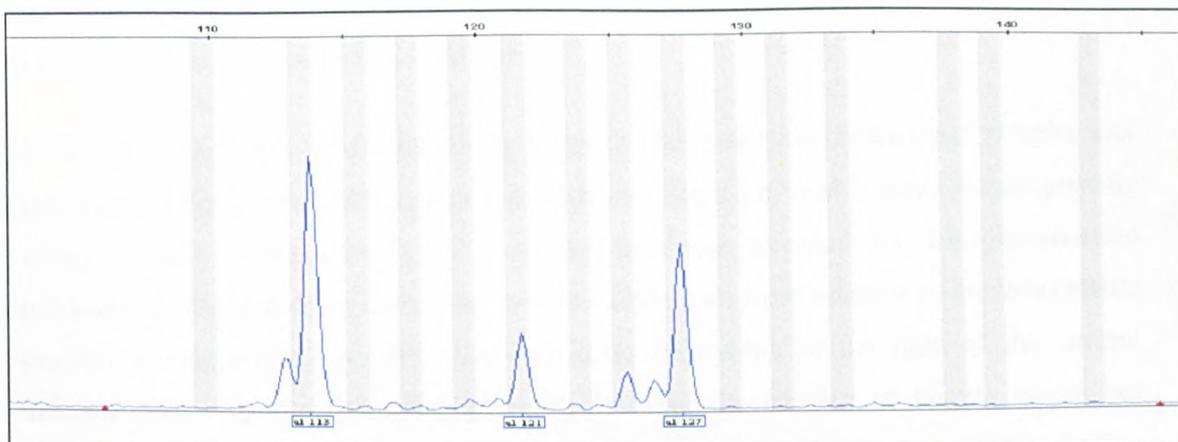
Figure 24 A sequence trace showing three allele peaks at a single locus, indicative of either trisomy or multi-clonality.

Written in Microsoft Visual Basic in collaboration with John Rivett-Carnac (Chorus Consulting Ltd), a script was designed to make multiple random diploid re-samplings of the mutliclonal/ aneuploid dataset, in a method similar to bootstrapping, but including all samples on each run. These outputs were linked via a batch script file written by the author in DOS to the MICROSAT v1.2 software package (290). Congruence between outputs was tested as in Section 3.7.6.1.2, and a single diploid dataset was picked at random from those generated for use in all subsequent analyses.

Even though the instability introduced by multi-allelic/aneuploid loci on estimates of between-isolate genetic distance can be assessed using the technique described above, the necessity of taking only a single dataset on for later analysis introduces a level of immeasurable error, and complicates interpretation of associated parameters. The extent and nature of this error can be approximately estimated. Firstly, minimal loss of data was associated with this approach (<0.6% for both TCI and TCIIc, Chapter 5). Secondly, the frequency of multiple alleles per population is included in Chapter 5, and highlights where more caution in interpreting results is necessary. Thirdly, if un-cloned samples containing mutli-allelic loci represent mixed clonal infections rather than aneuploid clonal isolates, then one might expect that diploid re-sampling of these profiles will act to over-estimate the level of heterozygosity present. This expectation may also be true for un-cloned isolates comprising mixed infections but demonstrating no multiple alleles ('hidden homozygosity,' see Chapters 1.5&6). Finally, multi-clonality may also act to artificially mimic panmictic

allele combinations, and caution must be used when interpreting linkage disequilibrium statistics in affected populations (Chapter 5).

A deductive approach to distinguishing between aneuploid and mutliclonal samples can also be envisaged. Multi-allelic loci demonstrating tight physical linkage would provide strong support of aneuploidy, and both datasets were screened for this phenomenon (Chapter 5). The deductive extraction of entire diplotypes from putative mutliclonal allelic profiles is theoretically possible but practically impossible in the light of the current datasets, and was not attempted, primarily because the number of known clones per population was too low to provide a common reference point. Additionally, this technique assumes that mutliclonal allelic profiles represent a perfect mix between different diploid profiles, an assumption that is demonstrably false (Chapter 6).

3.7.6.1.2 Individual pairwise distance comparisons

Pair-wise distance comparisons between individual samples were made in MICROSAT v1.2 under both an IAM (D_{AS}) and a SMM ($\delta\mu^2$) (212, 257). The relative merits of these two models of microsatellite evolution are discussed in Chapter 1. 1000 diploid re-samplings of the dataset were made, as in Section 3.7.6.1.1. To attain the final pair-wise distance matrix, the mean across those derived from each re-sampled dataset was taken, and a phylogenetic tree constructed in PHYLIP v3.67 using a Neighbour-Joining (NJ) clustering algorithm. Congruence between trees derived from 1000 re-sampled datasets was assessed under majority–rule consensus, also in PHYLIP v3.67, and used to assess the overall impact of multiple alleles on the dataset.

A further test of the robustness of tree topology was carried out by majority-rule consensus analysis in PHYLIP v3.67 of 10,000 bootstrap trees collated by combining 100 bootstraps, made in MICROSAT v1.2, each drawn from 100 respective randomly re-sampled datasets.

3.7.6.1.3 Isolation by distance (IBD)

Isolation by distance was calculated using a modification of Mantel's test for matrix correspondence in GENALEX v6 (291, 292). The statistic allows the testing of the relationship between matching elements in two different matrices, in this case pairwise genetic distance (D_{AS}) and pair-wise physical distance (km) between all samples. The output of the test is in the form of a linear regression, whereby the correlation (R_{XY}) between two variables is equivalent to the gradient of the regression line, and a measure of the 'goodness of fit' of the data to the regression line given by the R^2 value. The significance of any correlation is tested in GENALEX by a random permutation procedure (292).

Genetic distance matrices between individual isolates were calculated as in Section 3.7.6.1.2. Pairwise geographic distance matrices, in kilometres, were generated in ARCVIEW v3.3 (ERSI, USA) using a distance matrix extension under a Lambert Azimuthal Equal-Area Projection of the western hemisphere (293). Comparisons were made for both TCI and TCIIc datasets.

3.7.6.1.4 Estimation of population parameters

Several population parameters were calculated, a number of which are self-explanatory. Mean number of alleles per locus (MNA), calculated in Microsatellite Toolkit, was used as a measure of overall genetic diversity in populations (294). Mean observed and expected heterozygosity was calculated across loci in ARLEQUIN v3.1 (295). The method for calculating expected heterozygosity is detailed in Section 1.9.1, and relates to the gene diversity, defined by Nei (1973) as the probability of choosing two non-identical alleles at random at a locus in a population (296). Observed heterozygosity is a simpler measure, representing the proportion of individuals heterozygous at a given locus over the total number of individuals (206). In a population in HWE, the probability of choosing two different alleles at random from a population at a single locus equates directly with the expected number of heterozygotes in that population. A population that is not at HWE may either show an excess of observed heterozygosity, where a greater proportion of loci are heterozygous than expected at HWE, or a deficit, where a greater proportion of loci are

homozygous. A global test across loci per population was implemented in GENEPOP v4.0 for both heterozygous deficit and excess (297). Additionally, loci were tested individually for deviation from Hardy-Weinberg proportions in ARLEQUIN v3.1, and significance assessed after sequential Bonferroni correction, to limit the potential of making a type I error (295, 298).

Population specific estimates of the inbreeding co-efficient (F_{IS}) were calculated in ARLEQUIN v3.1 (295). F_{IS} provides a measure of the inbreeding of individuals within a subpopulation that produces deviation from panmixia as a result of non-random union of gametes. Values for F_{IS} typically vary between -1 (all loci are heterozygous), 0 (alleles are randomly distributed between individuals-HWE), and +1 (all loci are homozygous) (299). ARLEQUIN v3.1 calculates a p-value for each F_{IS} by randomly permuting alleles between individuals within a population. Significantly positive F_{IS} values are identified by those that fall in the upper 5% of the null distribution ($p<0.05$), while significantly negative ones fall into the lower 5% ($p>0.95$).

Two measures of inter-allelic linkage were calculated in MULTILOCUS v1.3 (300). The first of these was Maynard-Smith's (1993) Index of Association (I_A) (See Section 1.9.2) (252). I_A essentially assesses whether the probability of a pair of isolates being different at one locus influences the probability of their being different at another. To achieve this, the distance between all pairs of individuals is calculated, and the variance (V_O) of these values compared to that expected at HW equilibrium (V_E). If I_A deviates from zero, this can be taken as evidence of linkage between loci. MULTILOCUS assesses the significance of any deviation from zero by generating a null distribution using a random permutation procedure assuming an infinite level of recombination between individuals. Additionally, MULTILOCUS provides the option of predefining known multilocus linkage groups (i.e. loci of the same chromosome or, in this case, contig) and subpopulations of individuals, to inform the randomisation procedure and prevent the overestimation of linkage, as discussed in Section 1.9.2. In this study, loci on the same contig were assigned to the same linkage group (Table 35 and Table 42, Chapter 5 and Table 49, Chapter 6). A minority of loci, for which no information concerning their physical linkage was available, were excluded. In conjunction with linkage groups, where applicable populations were also defined, to prevent overestimation of linkage due to genetic subdivision (Section 1.9.2)

The I_A is a fairly conservative statistic, and may well fail to detect low levels of recombination in a dataset. As such a second statistic was also calculated, also in MULTILOCUS v1.3, for each possible locus pair, r^d (300). This statistic is calculated under the same theoretical framework as I_A, but was chosen because, in calculating linkage only between locus pairs, it is unbiased by the possible presence of strong linkage between some loci or groups of loci but not between others.

3.7.6.1.5 Population sub-division: F_{ST} and R_{ST}.

The level of subdivision was assessed initially for each pair of populations. Pair-wise F_{ST} values were estimated in ARLEQUIN v3.1 using θ, and represent the proportion of variation accounted for by the sub-division between two populations by comparison to the total level of variation across both populations (206, 295). As such, if no subdivision exists, variation is randomly distributed between the two populations $F_{ST}=0$, whereas if variation is discretely partitioned between two populations ($F_{ST}=1$) then total subdivision exists. Pairwise F_{ST} p-values are generated in ARLEQUIN v3.1 by comparison to a null distribution in which genotypes are randomly permutated between populations in the pair.

In reality, integer values for F_{ST} (1 or 0) are never obtained. Porter (1990) provides a general guide to the interpretation of F_{ST} estimates in nature, whereby $F_{ST}<0.2$ indicates negligible inter-population subdivision, $0.2<F_{ST}<0.33$ indicates moderate subdivision and values for $F_{ST}>0.33$ are indicative of effective isolation (little or no gene flow) (301). F_{ST} was also estimated across all sub-populations (non-pairwise) in the context of an AMOVA (Section 3.7.6.1.8).

As F_{ST} makes no explicit assumptions about the mechanism of genetic mutation, dealing instead with absolute allele identity, it can be viewed as conforming to the IAM of microsatellite evolution. R_{ST}, defined by Slatkin (1995), is a measure of population subdivision analogous to F_{ST} designed to account for variation in allele size, conforming to the SMM (302). While F_{ST} can be calculated as H_T-H_S/H_T, where H_T is the expected heterozygosity across both populations and H_S is the average expected heterozygosity within each population, R_{ST} is calculated as $S-S_W/S$, where S_W is twice the estimated average variance in allele size between alleles within each subpopulation and S is twice the

estimated average variance between alleles across both populations (302, 303). R_{ST} was calculated in MICROSAT v1.2, and while no random permutation procedure is implemented in this package, the dataset was subjected to 1000 bootstrap re-samplings to obtain the standard error of each pair-wise estimate (290). Neither R_{ST} nor F_{ST} explicitly rely on a HW allele frequencies within or between populations, indeed a departure from HWE resulting from population subdivision is what both tests are designed to measure. Because F_{ST} relies on H_E, not H_O, deviations from HWE do not effect the estimated probability of finding two identical alleles at the same locus in a sub-population vs the probability of finding the same across both populations. Simulations with clonal data indicate that F_{ST}, and by implication R_{ST}, is a fairly robust estimator of subdivision, so long as identical multilocus genotypes are not excluded from the analysis (299).

3.7.6.1.6 Statistical population assignment.

Statistical population assignment, based on allele frequency distributions within populations, was carried out in GENALEX v6 (292). The aim of this analysis was to test how robust population assignments of individual samples were made on an *a priori* basis (e.g. due to geography or transmission cycle). The algorithm implemented involves calculating the approximate distribution of genotype likelihoods within a population, leaving out the individual to be assigned, and estimating the likelihood of this individual belonging to that population by comparison to all others (304, 305). Frequency based assignment allows the estimation of the exact number of probable migrants to and from populations assigned on an *a priori* basis. By comparison, other estimates of immigration, e.g. *Nm*, based on *F*-statistics, provide only a relative measure (206). Additionally, this frequency based assignment algorithm intrinsically takes into account *a priori* population subdivision i.e. user-defined subdivisions are those for which genotype likelihood distributions are calculated. STRUCTURE (See Section 3.7.6.1.7), on the other hand, uses no *a priori* information, and groups individuals on the basis of genetics alone. Population parameters (Section 3.7.6.1.4) were therefore calculated on the basis of *a priori* population subdivisions once any clear 'migrants' had been re-assigned. This approach was seen as the attractive 'middle ground' between assigning populations on the basis of either genetics or geography alone.

Unfortunately, one key flaw in applying statistical assignment to *T. cruzi* populations is the reliance of this method on HW assumptions i.e. linkage equilibrium between loci and random mating between individuals. As such, allele frequencies within a population are used to generate the expected proportions of diplotypes and levels of heterozygosity within a randomly sampled population of the equivalent size at HWE. Deviations from HWE across the dataset may act to marginally lower the -ln likelihood of all individual assignments to populations, while population-specific allelic identity will still act to group similar isolates. To limit the effect of erroneous re-assignments, decisions to move samples between populations were also informed by the topology of pairwise NJ trees created as in Section 3.7.6.1.2 .

3.7.6.1.7 Analysis in STRUCTURE.

STRUCTURE v2.1 was used to define the probable number of sub-populations present in the assembled dataset in the absence of any *a priori* information regarding parasite geographical or ecological origin (306). Given a predefined number of populations, K, the software uses a Bayesian clustering approach to simultaneously assign individuals to populations using multilocus allele frequencies, while also defining the identity of these population groups on the basis of the allele frequency distribution. Furthermore, STRUCTURE allows a single multilocus phenotype to be assigned to 2+ populations (admixture), facilitating the identification of hybrid and/or transitional genotypes.

The probability of each value for K representing the 'true' number of populations in a sample is given by the value ln Pr(X/K), where X denotes the genotypes of the sampled individuals. As the value for K approaches the true number of populations, ln Pr(X/K) stabilises towards a plateau. Increasing the value of K beyond the 'true' number for K will produce no improvement in the value for ln Pr(X/K).

Like PHASE, STRUCTURE uses an MCMC algorithm to implement the clustering procedure. As such, multiple iterations, including a burn-in period, are required to allow the MCMC to converge on the optimal assignment of individuals between populations. The length of the burn-in, as well as number of subsequent iterations, is broadly dependent on the size of the dataset. As with PHASE, the accuracy of results can be assessed by

comparing the consistency of the estimation of ln Pr(X/K) between runs. Up to 10 runs per value of K are recommended, and, in general terms, if the variance in ln Pr(X/K) within runs is greater than that between runs, a longer burn-in and/or greater number of iterations per run is likely to be appropriate.

Like the statistical population assignment implemented in GENALEX, STRUCTURE is based around a model of HWE within populations, and complete linkage equilibrium between markers (292, 306). Again, this poses grave difficulties when applying the algorithm to organisms or populations of organisms that contravene these assumptions. Nonetheless, there is a precedent, and two recent studies have successfully applied STRUCTURE to the population genetic analysis of species from the *Leishmania donovani* complex, a related trypanosome to *T. cruzi* and an organism thought to have a predominantly clonal mode of reproduction (267, 307). In these studies, however, STRUCTURE was used as the sole basis for population assignment, and population parameters calculated from STRUCTURE-defined populations. In this study STRUCTURE is merely used as a further technique to estimate the level of hidden population substructure within predefined populations, and to inform the interpretation of population parameters drawn from them.

3.7.6.1.8 Analysis of Molecular Variance (AMOVA)

AMOVA was carried out in ARLEQUIN v3.1 (295). The analysis was used in the context of this study to test appropriateness of different methods of population subdivision in explaining the variance in the data. AMOVA is a hierarchical analysis that describes how molecular variance is apportioned within a dataset (308). Levels of hierarchy can be defined as within individuals within populations (for haplotypic data), between individuals within populations, between populations, and between groups of populations. In this case haplotypic data were not available, and the aim was to assess the appropriateness of population subdivision, in the absence of any higher level population grouping. As such the AMOVA was carried out at across two levels, between individuals within populations and among the populations themselves. The overall F_{ST} is calculated simply as the variance attributed to among population differentiation (σ^2_a) over the total variation present (σ^2_T). As

in Section 3.7.6.1.5, a p-value can be generated for the overall F_{ST} by comparison to a null distribution in which genotypes are randomly permutated between populations.

3.7.6.1.9 Tests for population bottlenecks.

A population bottleneck is the term applied to an event that drastically reduces the effective size of a population (N_e) over a relatively short period of time. Such an occurrence is frequently linked in conservation biology to rapid change or reduction in habitat, over-fishing, over-hunting etc. (e.g. (309, 310)). In parasite populations a bottleneck event might be associated with a selective sweep, or the exploitation of a new host species (e.g. (311, 312)). Population bottlenecks are often conceptually indistinguishable from a 'founder effect', whereby a small founder population becomes spatially and/or temporarily isolated from the parent population, best exemplified by studies of populations recently colonising oceanic islands (e.g. (313)).

At selectively neutral loci, the effect of population bottlenecks and founder effects are also theoretically indistinguishable (313). In neutral terms, both events involve taking a small 'random' sample from a large population, and the subsequent expansion of a secondary population from this limited number of individuals. Typically the impact of the rapid reduction in N_e on the genetic composition of a population dwindles over time, with bottleneck severity (size and speed of N_e reduction) positively correlating with a more lasting effect (314-316). A number of genetic phenomena have been identified that should be present in populations that have undergone a recent bottleneck. The first of these is reduced genetic diversity in the 'bottleneck' population with respect to the 'source' population (313, 317). While conceptually simple, this test necessitates prior knowledge on the identity of the 'source' population, a task that can be complicated in practice. Secondly, as small populations are inherently more sensitive to random allelic loss (genetic drift), than larger ones, and rare alleles - those at low frequency (1-10% in a population) - are more likely to be lost, a population with a large, stable N_e (at mutation-drift equilibrium) should show a greater number of these alleles than one that has recently undergone a bottleneck. Luikart et al., (1998) propose a simple graphical method to test this (315). Under an IAM, allele frequency classes in a population at mutation-drift equilibrium should follow a Poisson distribution (Figure 25, left) (318). Using real and simulated data, a modal shift in

allele frequency class from 1-10% to 10-20% or 20-30% was noted in bottleneck populations, declining over 10-40 generations in simulations, and rate of signature loss negatively correlating with founder population size.
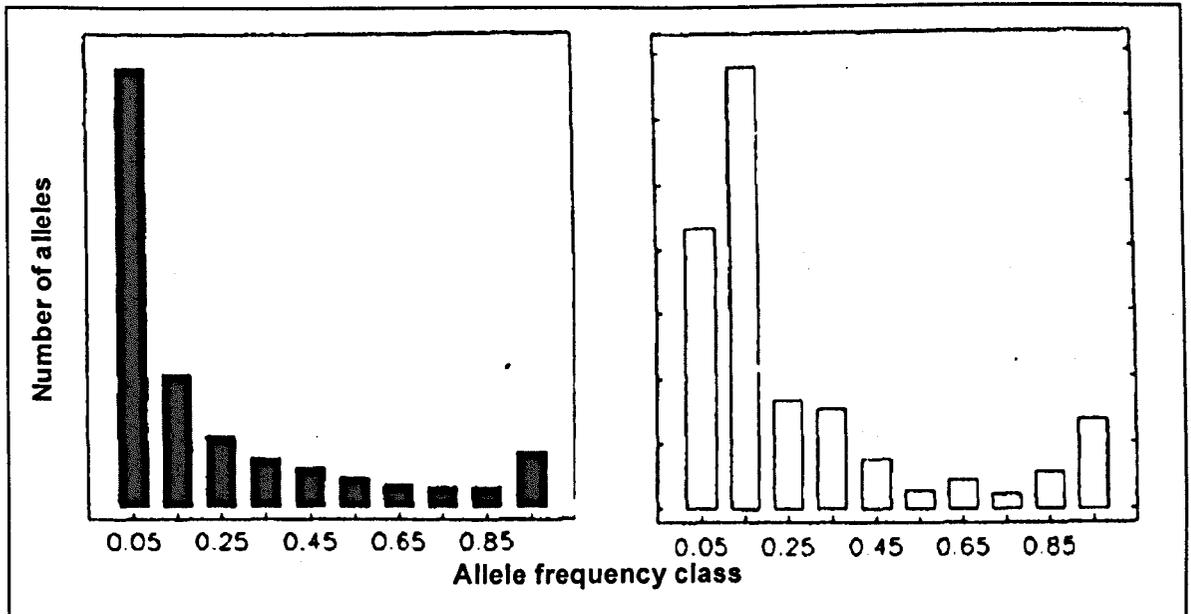
Figure 25 Theoretical modal shift in rare allele frequency after a population bottleneck. Left: Poisson distributed allele frequencies in a population at mutation-drift equilibrium, **Right:** Simulated allele frequencies in a population that has undergone a recent bottleneck. After Luikart *et al.*, (1998) (315)

Thirdly, while the observed heterozygosity of a group of individuals that have been subject to a population bottleneck should theoretically decline, it has been noted that, immediately after an event, a transient excess may be observed with respect to that expected (314). This effect occurs because gene diversity (or expected heterozygosity in an HW population (see Section 3.7.6.1.4)) is thought to decline more rapidly than observed heterozyosity, assuming the 'parent' population existed at HWE. If random mating occurs in the bottleneck population, then the observed level of heterozygosity will then decrease to HW proportions over a number of generations. The rate at which this decline occurs is thought be a function of bottleneck severity, such that the greater ratio of N_e before a bottleneck to the N_e after, the greater the number of generations over which it can be detected.

The final test, the Garza-Williamson index, applies only to microsatellite data (316). The underlying prediction of the Garza-Williamson index is that microsatellite allelic diversity should decline more rapidly than allelic size range in population bottleneck. As such, the M

ratio (allelic diversity vs size range) should approach zero in bottleneck populations. Simulations indicate that, again, the time taken for M to return to pre-bottleneck values depends on the severity of the bottleneck.

The application of most of these tests to populations that do not conform to HW expectations is debatable. While the Garza-Williamson test does not make any specific assumptions regarding allelic frequencies, it does assume a SMM, and has only been tested in sexually reproducing populations (316). Tests for transient heterozygosity excess assume HWE in the 'parent' population, and the Poisson distribution of allelic frequency classes for a population at mutation-drift equilibrium was developed under the assumption of random mating (314, 318). Additionally sample size is critical, and a minimum of 20 individuals is recommended for most tests. As such only three of the four methodologies were implemented (Garza-Williamson, Modal shift, and Allelic diversity) where descriptive values could be derived in the absence of any statistical test.

Table 5 Summary table describing selected key tests and their assumptions.

Aim	Test/ Statistic	Description	Assumptions & disadvantages	Reason for use	Precedents in partially or fully clonal organisms.
Measure of population subdivision	F_{ST} Section 3.7.6.1.5	Measures population subdivision by comparing the distribution of variation within populations with that between.	• Implicitly assumes IAM. • Developed in a framework of HWE, but reliant on H_E not H_O, and only marginally affected by deviations from HWE	Provides pair-wise measure of gene-flow $(1 - F_{ST})$ between population pairs.	e.g. *Leishmania sp* (267, 319) *Candida. albicans(320)*, *T. cruzi(321)*
	R_{ST} Section 3.7.6.1.5	Measures population subdivision as a function of the variance in allele sizes within and between populations.	• Explicitly assumes SMM of microsatellite mutation. • Not reliant on within population HWE. • At the population level, less sensitive to minor deviations from SMM. • Sensitive to saturation.	Provides pair-wise measure of population differentiation under SMM for comparison with IAM.	e.g *Leishmania sp(266)*
Assignment	Statistical population assignment (GENALEX) Section 3.7.6.1.6	Calculates within-population allele frequencies and assigns individual samples across predefined populations.	• Based on the assumption of within-population HWE. H_O of assigned diplotypes must fit with H_E calculated on the basis of observed allele frequencies.	Used as a means of genetic population assignment that incorporates a level of *a priori* population information.	No precedent but HW assumptions similar to those employed in by STRUCTURE.
	Bayesian population assignment (Structure) Section 3.7.6.1.7	Simultaneous calculation of population identity and population assignment.	• As above, assigns diplotypes to populations on the basis of their expected frequency under HWE.	Used as a means of genetic population assignment in the absence of any *a priori* information.	e.g. *Leishmania sp (266, 267)*. *P. vivax(322)* *C. krusei(323)*

Table 5 Continued from overleaf.

Bottleneck	Test for HW excess Section 3.7.6.1.9	Identifies transient excess in H_O with respect to H_E in bottleneck populations.	• Assumes HWE in source population and random mating in bottleneck population. • Sample size N>20. • Bottleneck recent event.	Not used (see text).	No precedent
	G-W Index Section 3.7.6.1.9	Identifies transient increase in allelic size range with respect to low genetic diversity.	• Assumes SMM • No assumption of HWE, but see text. • Sample size N>20 • Bottleneck recent event.	No test employed, descriptive statistic compared between populations.	No precedent
	Modal shift Section 3.7.6.1.9	Identifies transient shift in rare allele frequency.	• Assumes mutation-drift equilibrium and random mating in source population • Sample size N>20 • Bottleneck recent event.	Graphical output compared between populations	No precedent

4 Field observations, *T. cruzi* sub-lineage/host/ecotope associations.

4.1 Study site observations

4.1.1 Bolivia

4.1.1.1 El Beni department

El Beni department is a sparsely populated province in lowland eastern Bolivia. The region comprises approximately 20% of the Bolivian landmass, covering an area of 213,564 km^2, but is home to only 362,521 people, 4.3% of the total population(324). El Beni has a principally agrarian economy, with large sections of the region given over to cattle ranching, limited soya production on the southern border with Santa Cruz Department, and subsistence agriculture of maize, manioc, plantain and rice around rural communities. The region is hot and humid with mean annual temperatures of 25-26°C and rainfall of 1500-2550mm (325, 326). Precipitation is seasonal, with a wet season from November to April. The majority of El Beni lies 130-235 metres above sea level and is subject to intense seasonal flooding. During the 2006/2007 wet season, the region experienced the worst flooding in 25 years, causing an estimated US$115 million of damage to the local economy, and principally affecting cattle ranching in the lower lying north (327). Ecologically, El Beni is a patchwork of two principal vegetation types (Figure 26). The majority of the department is covered by lush savannah grassland known collectively as the 'Llanos de Moxos'. Along riverine alluvial plains and to the northern and western borders of the region, this ecotope is supplanted by dense Amazonian moist forests. To the east, El Beni borders another moist forest, the Madeira-Tapajos forest, which extends east into Brazil and south into north-western Santa Cruz department.

Seasonal outbreaks of both *P. vivax* and *P. falciparum* malaria, especially in the north-western portion of the department, have been widely reported in El Beni, and the Amazonian region as a whole accounts for over 50% of all cases in the Bolivia (83, 328). Dengue fever is also present, especially in urban areas, and occurs seasonally in response to vector abundance (83). Sporadic outbreaks of viral haemorrhagic fever have also been reported and are attributed to a rodent-borne arenavirus (271). By contrast, little is known concerning the status of Chagas disease in El Beni. One study was able to demonstrate the presence of active domestic disease transmission in Alto Beni, on the western fringe of the

department, and implicated *Rhodnius stali* as the primary vector species (40). *R. stali* was also found in nearby silvatic ecotopes infesting palm trees (Genus *Atallea*).

Fieldwork was undertaken in El Beni during 2004 and 2007 in three principal areas. A detailed map of these regions is provided in Figure 26. In the following sections, fieldwork observations for each site, mammal, triatomine, and trypanosome collections are described.

Figure 26 Map of study sites within El Beni department. State boundaries are defined by heavy black lines. Colours represent distinct ecotopes (See legend). Study sites: San Juan de Mocovi and region - **SJMO** - San Juan de Mocovi; **LORI** – Loma Rica; **POZ** – Poza Onda; **BELSEL** – Bella Selva; **CER** – Cerrito; **IBIA** – Ibiato; **ABAC** – Abacuya; **NUAL** – Nueva Allianza; **SJ** – San Juan de las Aguas Dulces; Santa Maria and Region - **MERC** – Mercedes; **SMA** – Santa Maria de Apere; **SMIG** – San Miguel; San Cristobal and region - **PIC** – Picaflores; **SANHEL** – Santa Helena; **SCRI** – San Cristobal. Ecotope divisions defined in Olson *et al.*, 2001 (269).

1.1.1.1.1 San Juan de Mocovi and region, Southern El Beni

San Juan de Mocovi is small community of ~500 inhabitants, situated in the southern portion of El Beni, close to the major road link to the city of Santa Cruz and ~60km south of Trinidad, the capital of El Beni department. Ecologically this region represents the border between the Llanos de Moxos and the moist forests of north-western Santa Cruz. Vegetation consists of patches of savannah grassland interspersed between 4-10m high stands of large evergreen tropical trees including *Ceiba* and *Coccoloba* species as well as large number of *Atallea* species palms (Figure 27). Agricultural land use transformation,

including timber exploitation, is moderate, with some clearing occurring to facilitate cattle grazing, and many of the larger hardwood species (e.g. *Dimorphandra mora*) having been removed many years previously. This region is somewhat higher in altitude (~160m above sea level) than the north of Beni, and less affected by flooding.

Figure 27 Typical lush southern El Beni department vegetation

Mammal captures were made around two communities in this region, San Juan de Mocovi (15°07'S 65°19'W) in 2007 and San Juan de las Aguas Dulces (14°49'S 64°36'W) in 2004, a slightly smaller community ~50km north of SJMO. A total of 85 silvatic mammals were captured in both communities. Mammal species and infection prevalence are shown in Table 6

Table 6 Silvatic mammal captures and associated trypanosomes in southern El Beni.

Species	Capture number	Infected*	Prevalence %	Trypanosome species
Didelphis marsupialis (Common opossum)[a,b]	36	27	75	26 *T. cruzi* / 1 *T. cruzi – T. rangeli* mix
Dasypus novemcinctus (Nine-banded armadillo)[a]	12	4	33.3	4 *T. cruzi*
Tayassu tajacu (Collared peccary)[a]	1	0	0	n/a
Philander opossum (Grey four-eyed opossum)[a,b]	9	5	55.6	4 *T. cruzi* / 1 *T. rangeli*
Tamandua tetradactyla (Lesser Anteater)[a]	3	3	100	3 *T. rangeli*
Dasyprocta agouti (Agouti)[a]	7	0	0	n/a
Saimiri sciureus boliviensis (Squirrel monkey)[b]	1	1	100	1 *T. rangeli*
Eumops / Molossus sp. (Mastiff bat)[a]	2	0	0	n/a
Phyllostomus sp. (Large fruit-eating bat)[a]	2	2	100	2 Unidentified trypanosomes
Euphractus sexcinctus (Yellow Armadillo)[a]	1	1	100	1 *T. cruzi*
Sciureus spadiceus (Red squirrel)[a]	2	1	50	1 *T. cruzi*
Unidentified rodents[b,c]	9	n/a	n/a	n/a
				37 *T. cruzi*, 6 *T. rangeli*,
Total	85	44	51.8	2 Unidentified trypanosomes

*By blood agar culture, xenodiagnosis, or direct PCR of xenodiagnositic triatomine faeces.
[a] Caught with the aid of local hunters
[b] Caught in cage traps
[c] Small rodents were not handled due to safety concerns (See Section 3.2.2)

This study was undertaken during the dry season and traps were placed within 10km of both communities (Total trap nights = 540). In San Juan de las Aguas Dulces, as well as on the ground, traps were also placed in arboreal trap lines and resulted in the capture of a number of *D. marsupialis* and one *S. sciureus boliviensis*. Other mammals caught in traps, in San Juan de las Aguas Dulces and San Juan de Mocovi, included *P. opossum* and a number of small rodent species which were not handled (Section 3.2.2). All other mammals, including bats, were captured with the aid of local hunters. The *Ta. tajacu* sampled had been domesticated by the community (San Juan de las Aguas Dulces) and lived in close proximity to humans - apparently a common local practice, also observed at other study sites (Section 4.1.1.1.2). Two other mammal species caught, *D. novemcinctus* and *D. agouti,* were routinely hunted for food in both communities. Global infection prevalence, irrespective of trypanosome species, was ~50%. Comparisons of infection prevalence between species are problematic as sampling was skewed towards a relatively small number of abundant species. Nonetheless, it is perhaps of note that all three *Tam. tetradactyla* captured exclusively harboured *T. rangeli*.

In conjunction with mammal trapping, 128 domestic, peridomestic, and silvatic triatomines were also captured (Table 7). Domestic and peridomestic searches were made in 124 dwellings in nine different communities (Table 8). Four species of triatomine were recovered: *Psammolestes sp.*, *R. robustus*, *R. pictipes*, *Tr. sordida* and associated nymphal instars. Peridomestic foci, principally chicken nests and enclosed roosts, were the greatest source of infestation associated with human habitation (52 triatomines). Triatomines recovered from truly domestic foci (N=20), including cracked adobe walling, under bedding, as well as from chicken nests under beds, were fewer in number. Silvatic foci yielded a further 56 triatomines. *Panstrongylus geniculatus* adults and nymphs were recovered from microhabitat dissections undertaken among the leaf bedding of two different species of fossorial mammal: *D. novemcinctus* and *D. agouti,* tracked to their burrows using the spool-and-line technique (Section 3.2.3). *Psammolestes sp.* were found among bird nests in thorny scrub. *R. robustus* and *R. pictipes*, as well as numerous *Rhodnius* nymphs not reliably identifiable to species, were co-recovered from species of *Atallea* palm surrounding communities using Noireau traps (Section 3.2.3). Spool-and-line tracking of didelphid mammals proved problematic, as threads were lost in the tree canopy.

Nonetheless Noireau traps were also set in palms and tree holes close to terrestrial cage trap lines, with a view to sampling possible arboreal mammal nesting sites. *R. robustus* and *R. pictipes* were recovered.

Table 7 Triatomine captures and associated trypanosomes in southern El Beni.

Species	Total	Infected Microscopy	Direct PCR	Transmission cycle *(Infected)* Domestic	%	Peridomestic	%	Silvatic	%	Trypanosome species
Panstrongylus geniculatus	2	0	0	0	n/a	0	n/a	2	0	n/a
Panstrongylus nymph	7	1	0	0	n/a	0	n/a	7(1)	14.3	*T. rangeli*
Psammolestes sp.	13	n/a	n/a	0	n/a	11	n/a	2	n/a	n/a
Rhodnius pictipes	8	5	6	0	n/a	4(3)	75	4(4)	100	*T. cruzi*
Rhodnius robustus	6	6	4	1(1)	100	2(2)	100	3(3)	100	*T. cruzi*
Rhodnius nymph	60	18	22	3(3)	100	19(2)	10.5	38(23)	60.5	*T. cruzi* *T. cruzi / 2xT. cruzi - T. rangeli* mix
Triatoma sordida	10	0	5	4(2)	50	6(3)	50	0	n/a	*T. cruzi*
Triatoma nymph	22	0	4	12	0	10(4)	40	0	0	*T. cruzi*
Total	128	30	41	20(6)	30	52(14)	26.9	56(31)	55.4	

Notable in its absence from silvatic foci sampled in this study was *Tr. sordida*. This species exclusively infested peridomestic and domestic foci. It was also the only species found sequestered in cracks in adobe walling, especially in IBIA. Conversely, neither *R. pictipes*, *Psammolestes sp.*, nor *P. geniculatus* were found domestically, although the presence of *R. pictipes* in houses cannot be categorically refuted owing to the difficulty of identifying nymphs to a species level. At peridomestic foci, the only species absent was *P. geniculatus*. All other species were found, frequently co-infesting the same locality. The total prevalence of domestic/peridomestic triatomine infestation across all communities studied was 12.9% (Table 8). Triatomine distribution was patchy, however, and four communities, including San Juan de las Aguas Dulces, were apparently vector free, while in Loma Rica, for example, over half of all houses and outbuildings searched showed signs of infestation. IBIA, which yielded exclusively *Tr. sordida* and *Triatoma* nymphs, showed signs of the highest density of infestation within individual dwellings. The walls of most houses were heavily streaked by triatomine faeces, although the insects themselves were buried deep within the adobe and difficult to remove, even where irritant solution was applied (Section 3.2.3). No systematic fumigation strategy was reported in this region by local authorities or the communities themselves.

Table 8 Communities in southern El Beni: Prevalence of domestic and peridomestic triatomine infestation

Community	Code	Longitude	Latitude	Houses searched	Houses Infested	Prevalence (%)
San Juan de Mocovi	SJMO	15°07'S	64°19'W	24	3	12.5
Loma Rica	LORI	15°05'S	64°13'W	10	6(2)	60.0
San Juan de las Aguas Dulces	SJ	14°50'S	64°36'W	4	0	0.0
Cerrito	CER	14°43'S	64°05'W	7	0	0.0
Poza Honda	POZ	15°00'S	64°15'W	19	0	0.0
Nueva Alianza	NUAL	15°02'S	64°20'W	15	1	6.7
Abacuya	ABAC	15°00'S	64°23'W	15	1(1)	6.7
Ibiato	IBIA	14°50'S	64°26'W	20	5(5)	25.0
Bella Selva	BESE	14°58'S	64°20'W	10	0	0.0
			Total	124	16	12.9

Numbers in parentheses indicate proportion of houses infested domestically.

Parasite prevalence among all triatomines, irrespective of trypanosome species, vector species, and transmission cycle, was 39.8% (Table 7). On the basis of molecular data (Section 4.2), one silvatic *Panstrongylus* nymph showed evidence of intestinal infection with *T. rangeli*, and two silvatic *Rhodnius* nymphs were apparently co-infected with both *T. cruzi* and *T. rangeli*. Silvatic triatomines showed the highest prevalence of *T. cruzi* infection - comparable to that of silvatic mammals - with 55.4% of insects showing evidence of parasite by either direct microscopic examination of gut contents and/or direct amplification of parasite DNA. Peridomestic and domestic triatomines showed lower levels of infection, 26.9% and 25% respectively. In almost all cases direct PCR from DNAzol® - extracted gut homogenate DNA proved more efficient in identifying the presence of parasite than microscopic examination of triatomine faeces. In some instances microscopically identified parasite did not corroborate with a positive PCR result. In these cases it is likely that the quantity of homogenate available for extraction after passage to biphasic culture was too low to facilitate PCR analysis.

4.1.1.1.2 Santa Maria de Apere and region

Situated in northern El Beni department, ~125km northwest of Trinidad (Figure 26), Santa Maria de Apere (14°08'S 65°22'W) is a very small community of <100 inhabitants that borders the Apere river. This region is lower lying than the south (~143m above sea level), and was subject to intense flooding in 2007. Floodwaters, even in July, still prevented

113

passage further north by land, the original intention of the expedition. Riverine forest prevails along the borders of the Apere, again typified by *Attalea spp.* palms, and buttressed-trunked trees (e.g. *Ceiba spp.*). Native forest bordering the community had been considerably transformed. Endemic species were interspersed by coffee, plantain and fruit tree plantations. The depth of forest along the river boundary was no greater than 2km, however, and the dominant vegetation type was open savannah grassland with occasional lone-standing trees (Figure 28). The region is usually subject to some flooding and roads are impassable during the wet season in most years. Santa Maria de Apere is thus extremely remote, sees little human traffic for much of the year, and the population density of the entire region is much lower than in southern El Beni department.
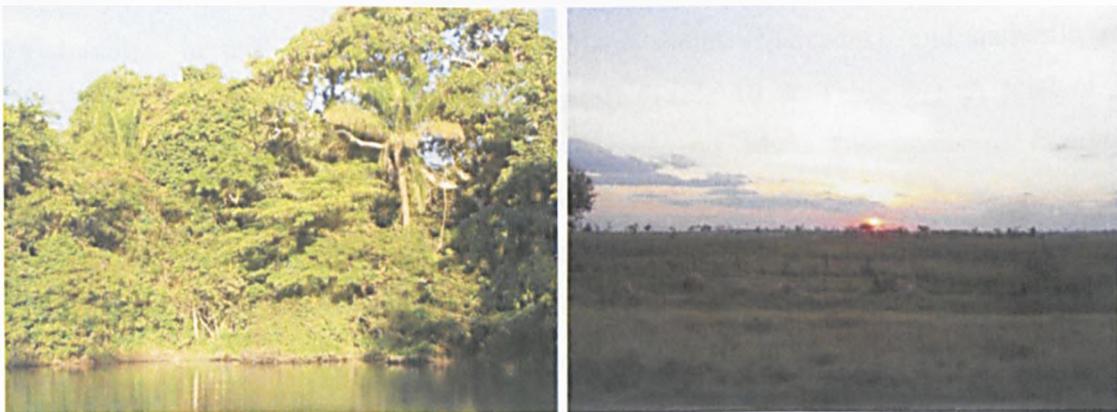
Figure 28 Vegetation types around Santa Maria de Apere, northern El Beni. Left: Riverine moist forest, right: open savannah grassland.

Cage traps were set for silvatic mammals within 5km of the community in low secondary forest. Local hunters were also employed to trap further animals. 24 mammals were captured, representing only three species: *D. novemcinctus, D. marsupialis,* and *P. opossum* (Table 9, Total trap nights = 120). *T. cruzi* prevalence in two species: *D. novemcinctus* and *D. marsupialis,* was greater than that observed in San Juan de Mocovi and San Juan de las Aguas Dulces, while that in *P. opossum* was broadly comparable (Table 6) Observational evidence indicated that many other mammal species were present in the region, including *Tam. tetradactyla*, giant anteaters (*Myrmecophaga tridactyla*), and *Sciureus spp.*, but these were not trapped. As in San Juan de las Aguas Dulces, wild-caught *Ta. tajacu* were reared domestically for food, and *D. novemcinctus* was frequently hunted for the same purpose. One coati, *Nasua nasua,* was also kept as a pet.

114

Table 9 Mammal captures and associated trypanosomes, Santa Maria de Apere, northern El Beni.

Species	Captured	Infected*	Prevalence %	Trypanosome species
D. novemcinctus*	13	12	92.3	T. cruzi
D. marsupialis*	5	5	100	T. cruzi
P. opossum*	6	4	66.6	T. cruzi
Total	24	21	57.5	

*By xenodiagnosis and/or direct blood culture.
*Caught with the aid of local hunters
*Caught in cage traps

Domestic, peridomestic, and silvatic searches were made in Santa Maria de Apere for triatomines, in one community within a 30km radius (Mercedes), and domestic and peridomestic searches in a third (San Miguel) (Table 10 & Table 11). A total of 77 triatomines were captured, including two *R. robustus* adults and numerous *Rhodnius* nymphs. Peridomestic infestation rates were very low - only one house harboured triatomines in its outbuildings - a dog kennel and a chicken roost in Santa Maria de Apere. Domestic infestation was not detected in any dwellings. Widespread infestation of silvatic foci by *Rhodnius sp.*, principally *Attalea spp.* palms around the community and adjacent to mammal trap lines, was identified from two palm dissections as well as using Noireau traps. Where possible, *D. novemcinctus* were tracked using the spool-and-line technique but no infested burrows were detected.

Table 10 Communities in southern El Beni: Prevalence of domestic and peridomestic triatomine infestation.

Name	Code	Longitude	Latitude	Houses examined	Houses infested	%
Santa Maria de Apere	SMA	14°08'S	65°22'W	8	1(0)	12.5
San Miguel de Apere	SMIG	14°00'S	65°21'W	3	0	0.0
Mercedes	MERC	14°44'S	65°44'W	19	0	0.0
			Total	30	1	3.3

Numbers in parentheses indicate proportion of houses infested domestically.

Overall parasite prevalence in triatomines (51.7%), irrespective of transmission cycle and species, was higher than that discovered in southern Beni (39.8%, Section 4.1.1.1.1). Of 77 triatomines captured, however, it was only possible to examine 33 insects. Again, PCR identification of infected faeces proved more sensitive in detecting parasite. Difficulties

encountered returning from this study site resulted in the desiccation of many of the triatomines *en route*, which rendered their analysis extremely problematic.

Table 11 Triatomine captures and associated trypanosomes in northern El Beni.

Species	Total	Examined	Infected		Transmission cycle *(infected / examined)*						Trypanosome species*
			Microscopy	Direct PCR	Domestic	%	Peridomestic	%	Silvatic	%	
R. robustus	2	2	1	1	0	n/a	0	n/a	2 (1/2)	50	*T. cruzi*
Rhodnius nymph	75	31	6	14	0	n/a	6(4/4)	100	69(14/27)	51.9	*T. cruzi*
Total	77	33	7	15	0	n/a	6(4/4)	100	71(15/29)	51.7	

4.1.1.1.3 San Cristobal and region

San Cristobal, a community of ~200 people bordering the Alto Beni region of La Paz department, is situated ~210km west of Trinidad (Figure 26). As with other study sites in El Beni, the vegetation type was a mixture of savannah and moist forest (Figure 29). This region was more affluent than others studied in El Beni, and excellent trade links exist with La Paz city. Extensive agricultural land-use transformation has occurred, with substantial clearing of native forest for cattle grazing, as well as plantain and manioc cultivation. Unseasonably high rainfall severely hampered fieldwork at this site.

Figure 29 Typical vegetation in and around San Cristobal, western El Beni department.

Mammal cage traps were set around the community (Total trap nights = 80), and local hunters engaged to trap further mammals, although none had the economic incentive to do so. Ten mammals were trapped, including seven *P. opossum* and three *D. marsupialis* (Table 12). Despite the small sample size, overall parasite prevalence in the two silvatic

116

mammal species captured (66.6% and 57.1% respectively) was broadly comparable to other sites in El Beni.

Table 12 Mammal captures and associated trypanosomes, San Cristobal, northern El Beni Dept.

Species	Captured	Infected*	Prevalence %	Trypanosome species
D. marsupialis	3	2	66.6	*T. cruzi*
P. opossum	7	4	57.1	*T. cruzi*
Total	10	6	60	

*By xenodiagnosis and/or direct blood culture

Silvatic, domestic and peridomestic searches were made for triatomines in San Cristobal, while domestic and peridomestic searches were made in two nearby communities, Picaflores and Santa Helena (Table 13). Nine triatomines were recovered, including *R. robustus*, *R. pictipes* and associated nymphs (Table 14). No domestic infestation was evident in any communities searched. One peridomestic *R. pictipes* was found in a chicken coop in San Cristobal, but no nymphs were present, suggesting that active colonisation of this locality was unlikely. All eight remaining triatomines were captured with the aid of Noireau traps placed in *Attalea spp.* palms around the community and near mammal trap lines.

Table 13 Communities in western El Beni: Prevalence of domestic and peridomestic triatomine infestation

Name	Code	Longitude	Latitude	Houses examined	Houses infested*	%
Picaflores	PIC	13°54'S	66°47'W	10	0	0.0
Santa Helena	SHELE	14°34'S	66°49'W	10	0	0.0
San Cristobal	SCRIS	14°10'S	66°54'W	13	1(0)	7.7
			Total	33	1	3.0

Numbers in parentheses indicate proportion of houses infested domestically.

Overall parasite prevalence in triatomines was low (22.2%) by comparison to other sites in El Beni (Table 14). The low sample size, largely a result of poor weather, precludes any meaningful inference from this observation.

Table 14 Triatomine captures and associated trypanosomes in western El Beni.

Species	Total	Infected		Transmission cycle (infected)						Trypanosome species*
		Microscopy	Direct PCR	Domestic	%	Peridomestic	%	Silvatic	%	
R. robustus	1	0	0	0	n/a	0	n/a	1	0	*T. cruzi*
R. pictipes	1	1	1	0	n/a	1(1)	n/a	0	n/a	*T. cruzi*
Rhodnius nymph	7	1	1	0	n/a	0	n/a	7(1)	14.3	*T. cruzi*
Total	9	2	2	0	n/a	1(1)	100	8(1)	12.5	

4.1.1.2 Santa Cruz department

Santa Cruz department is the largest regional province in Bolivia covering 370,621km^2 and home to ~2,400,000 people, 26% of the total population (324). Over half, ~1,500,000 people, live in the state capital, Santa Cruz de la Sierra. Ecologically Santa Cruz department is highly diverse (Figure 30). The eastern portion of the province is covered by Chiquitania dry forest, typified by broad-leaved trees including *Schinopsis brasiliensis*, *Anadenanthera macrocarpa*, and *Machaerium scleroxylon*. Primary forest has a canopy height of approximately 20m with emergent trees standing up to 30m tall. Average annual rainfall in the Chiquitania is 936mm, with a marked dry season occurring from April to October (329). Mean annual temperatures are 24.3°C, marginally lower than those in El Beni. The southern portion of Santa Cruz department comprises the northern extreme of the drier Chaco region, which extends a further ~1500km into Paraguay and Argentina. Mean annual rainfall here is around 835mm and mean annual temperatures 26°C, although daytime temperatures in the dry season can soar to 45°C. Vegetation consists of low thorny scrub and cacti, interspersed by areas of denser thorn forest and patches of dry savannah grassland. The west of Santa Cruz department borders the Chapare region and dense temperate montane forests in the westernmost foothills of the Andes. To the east of Santa Cruz the Chiquitania gives way to the southern fringes of the Brazilian Pantanal wetlands.

Santa Cruz department is the agribusiness capital of Bolivia. Extensive cultivation occurs in other departments, but none are major exporters. By contrast, in Santa Cruz, cultivation of soya, sorghum, oilseed, sugar cane, rice, fruit and vegetables accounts for approximately

85% of total crop production in Bolivia (330). Forestry, legal and illegal, is also widespread. Of key economic importance, Santa Cruz contains 75% of all Bolivia's oil and natural gas reserves, estimated at 890 billion cubic metres, and ranked the second largest on the continent (331).
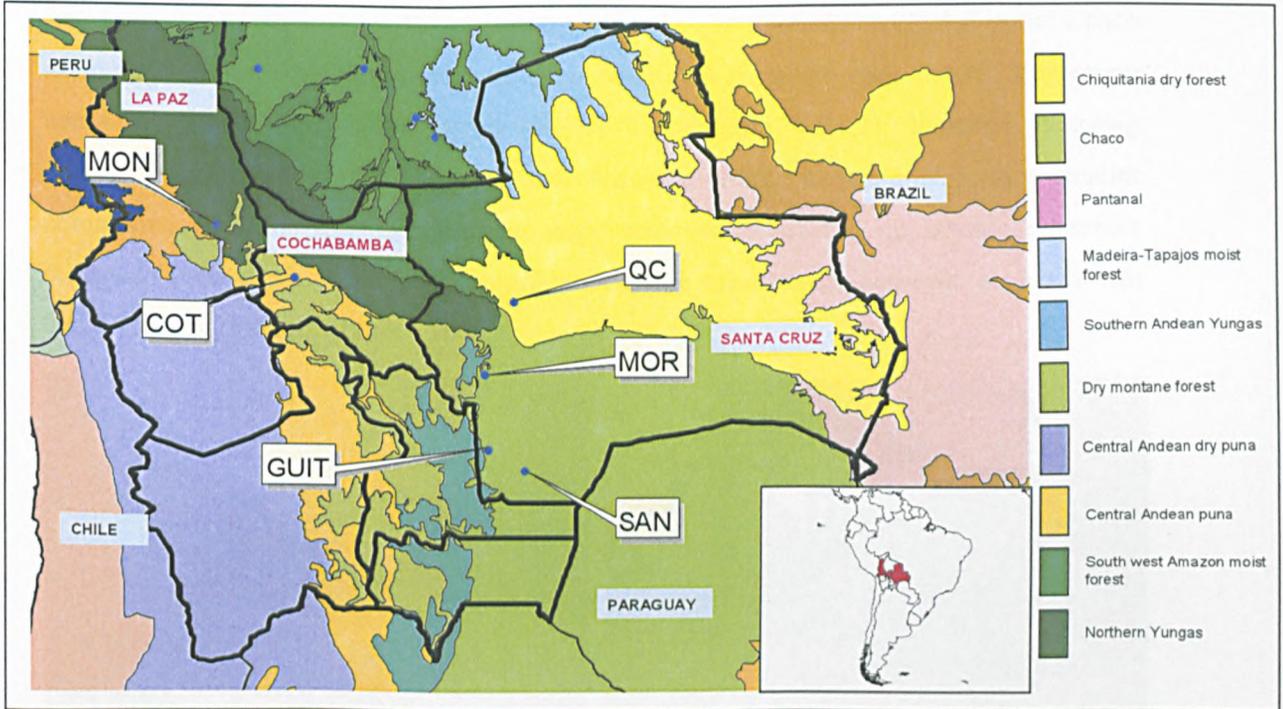
Figure 30 Map of study sites within La Paz, Santa Cruz, and Cochabamba departments. State boundaries are defined by heavy black lines. Colours represent distinct ecotopes (See legend). Study sites: **MON** – Montequilla, Dept La Paz; **COT** – Cotopachi, Dept. Cochabamba; **MOR** – Mora, Dept. Santa Cruz; **GUIT** – Guitterez, Dept. Santa Cruz; **SAN** – San Antonio, Dept. Santa Cruz; **QC** - Quatro Canadas, Dept. Santa Cruz. Ecotope divisions defined in Olson *et al.*, (2001) **(269)**

Much of the wealth generated by the abundant natural resources of Santa Cruz department is concentrated in the state capital, while rural poverty indices are directly comparable to those in resource-poor departments like Cochabamba, Potosi, and Chuquisaca (332). Domestic triatomine infestation, principally by *Tr. sordida* and *Tr. infestans*, and correlated with inadequate housing, is widespread (188, 333). A number of studies have demonstrated the presence of silvatic vectors in the region, principally *Tr. guasayana*, *Tr. infestans*, and *Tr. sordida*, but none have conducted comprehensive surveys of possible silvatic mammalian host species (97, 115). In 2004 fieldwork was undertaken in four areas with the primary goal of elucidating this aspect of silvatic *T. cruzi* ecology.

119

4.1.1.2.1 San Antonio, Dept. Santa Cruz

San Antonio (20°01'S 63°01'W) is a community of ~500 inhabitants situated in the Chaco of southern Santa Cruz department ~240km due south from the state capital. Vegetation comprises thorny xerophilous scrub (Figure 31, Section 4.1.1.2). Mammal sampling occurred within a 10km radius of San Antonio, which included a further smaller community, Pueblo Nuevo. Numerous cage traps were set but no silvatic mammals captured (Total trap nights – 248). Significant capture success was, however, achieved with the aid of local hunters (Table 15).

Figure 31 Left: Dry xerophilous scrub in the Chaco, right: typical adobe dwellings in the Chaco.

A total of 35 mammals were captured, including four species of armadillo: *Chaetophractus vellerosus*, *D. novemcinctus*, *E. sexcinctus* and *Tolypuetes matacus*; as well as one species of nocturnal primate (*Aotus sp.*). Parasitic prevalence was extremely low (2.8%) by comparison to that observed in El Beni department, although the prevalence of parasite in *D. novencintus* (50%) is consistent with other study sites, notwithstanding the very low sample size (N=2). Other mammal species were observed at this study site, including skunks (*Conepatus sp.*), *M. tridactyla*, and puma *(Puma concolor)*. These species either evaded capture or capture was deemed to hazardous to attempt. Excepting *Aotus sp.*, no other arboreal mammals were observed at this study site. In particular, neither *Philander* nor *Didelphis* species were present, despite considerable effort employed to find or trap them.

120

Table 15 Mammal captures and associated trypanosomes in the San Antonio, Chaco region, Dept Santa Cruz.

Species	Captured	Infected*	Prevalence %	Trypanosome species
C. vellerosus (Screaming hairy armadillo)	15	0	0	n/a
D. novemcinctus (Nine-banded armadillo)	2	1	50	*T. cruzi*
E. sexcinctus (Yellow armadillo)	4	0	0	n/a
Tol. matacus (Southern three-banded armadillo)	11	0	0	n/a
Aotus sp. (Owl monkey)	3	0	0	n/a
Total	35	1	2.8	

* By xenodiagnosis and/or direct blood culture

While triatomine capture was not a primary goal of this expedition, a limited number were recovered. Three adult *Tr. infestans* and five associated nymphs were found infesting a house in Pueblo Nuevo, despite reports of recent fumigation. None were infected. Three silvatic *Triatoma sp.* 5th instar nymphs were also recovered by a villager from an unidentified silvatic, reportedly terrestrial location. All three were infected with trypanosomes, although one isolate failed to grow in culture. These were identified as *T. cruzi*. Attempts were made to trace fossorial mammals back to their burrows using the spool-and-line technique to indentify possible triatomine vectors. The depth of burrows in this region was, however, far greater than that observed in El Beni, and the soil around them hard- packed, preventing access to the nests within.

4.1.1.2.2 Guitterez and Mora, Dept. Santa Cruz

Guitterez (19°41'S 63°12'W) and Mora (18°39'S 63°14'W), ~210km and ~100km south of Santa Cruz city respectively, lie on a linear transect between San Antonio and Santa Cruz city. Both sites are technically in the Chaco but occupy transitional ecotopes between two biomes. Neither site represents a community but their names refer instead to exclusively silvatic locations. Guitterrez lies ~30km to the northwest of San Antonio in a transitional zone between dry Chaco type vegetation similar to that observed in San Antonio, and the more humid, mountainous Southern Andean Yungas. This ecotype is typified by higher mean annual rainfall than the Chaco (2500mm) and correspondingly lush vegetation, including Andean alder (*Alnus acuminate*) as well as smaller trees of family Lauraceae and Myrtaceae.

Mora lies in transitional zone between the dry Chaco and montane forests to the west. This region is more densely wooded than the open thorn scrub of the Chaco but subject to similar annual precipitation on account of a rain shadow effect generated by the Andes. Mammal captures were undertaken at both of these sites with the aid of local hunters and in collaboration with Nidia Acosta (LSHTM). No silvatic mammal cage traps were set. In total 25 mammals were captured, 14 in Mora and 11 in Guitterez (Table 16 & Table 17), comprising six different species: *D. agouti, C. vellerosus, D. novemcinctus, E. sexcinctus, Nasua nasua,* and *M. tridactyla.*

Table 16 Mammal captures and associated trypanosomes in Mora, Chaco region, Dept. Santa Cruz

Species	Captured	Infected*	Prevalence %	Trypanosome species
D. agouti	1	0	0	n/a
C. vellerosus	3	0	0	n/a
D. novemcinctus	7	4	57.1	*T. cruzi*
E. sexcinctus	2	2	100	*T. cruzi*
Nasua nasua (Ring-tailed Coati)	1	0	0	n/a
Total	**14**	**6**	**42.8**	**n/a**

*By xenodiagnosis only

Table 17 Mammal captures and associated trypanosomes in Guitterez, Chaco region, Dept. Santa Cruz

Species	Captured	Infected*	Prevalence %	Trypanosome species
C. vellerosus	3	1	33.3	n/a
D. novemcinctus	2	1	50	*T. cruzi*
E. sexcinctus	5	0	0	n/a
M. tridactyla	1	0	0	n/a
Total	**11**	**2**	**18.2**	**n/a**

*By xenodiagnosis only

D. novemcinctus abundance was notably higher in these two study sites by comparison to the open Chaco in San Antonio, accounting for sample size. In species composition, mammals sampled in Guitterez more closely mirrored those recovered from San Antonio. Neither *D. agouti* nor *N. nasua* (caught in Mora) had been observed at any point inhabiting open Chaco. *T. cruzi* prevalence in mammals from both study sites was greater than that in San Antonio. In common with San Antonio, however, was the particularly high prevalence of *T. cruzi* infection observed in *D. novemcinctus.*

4.1.1.2.3 Cuatro Canadas, Dept. Santa Cruz

Cuatro Canadas (17°41'S 62°31'W), again a silvatic location, not a community, lies ~60 km due west of Santa Cruz city in the Chiquitania dry forest that predominates in the western portion of Santa Cruz Department (Section 4.1.1.2). Human land use transformation, in the form of forest clearance for soya and sorghum, has had a strong impact on this area. Nonetheless large patches of Chiquitania-type vegetation still remain. 27 mammals were captured with the aid of local hunters and comprised four species, *D. novemcinctus, E. sexinctus, Ta. tajacu,* and *Tol. matacus* (Table 18). No cage traps were set, nor were any triatomines recovered.

Table 18 Mammal captures and associated trypanosomes in Quatro Canadas, Chiquitania, Dept. Santa Cruz

Species	Captured	Infected*	Prevalence %	Trypanosome species
D. novemcinctus	11	7	63.6	*T. cruzi*
E. sexcinctus	13	2	15.4	*T. cruzi*
Ta. tajacu	1	0	0	n/a
Tam. matacus	2	0	0	n/a
Total	27	9	33.3	

*By xenodiagnosis only

As with all other study sites in Santa Cruz, *D. novemcinctus* showed the greatest prevalence of *T. cruzi* infection. This species was also marginally more abundant in this region than in the Chaco to the south.

4.1.1.3 Montequilla, Dept. La Paz

The city of La Paz lies on a dry plateau, 3,800m in elevation, high among the central Andes. To the east the terrain drops sharply, falling 3,400m in only 200 kilometres, to join the tropical lowlands of El Beni department. Below ~2500m the landscape is dominated by steep-sided semitropical valleys known as the northern Andean Yungas. The study site at Montequilla (16°27'S 67°31'W), a community of ~200 inhabitants, is situated ~70k due east of La Paz in the mid Yungas, 1,800m above sea level (Figure 30). Vegetation comprises dense broad leaf moist forest including numerous tree fern, orchid and bromeliad species. Mean annual rainfall is high, ~1500mm, and run-off from this region is responsible for much of the flooding which occurs in El Beni department (334). The annual

temperature fluctuates between 10°C and 30°C. Local agricultural land transformation is primarily attributed to the growing of coca leaves (*Erythroxylum coca*) and stepped plantations cover much of the central valley slopes. The steeper upper and lower extremes of the valleys remain densely forested and could be important foci for silvatic *T. cruzi* transmission (Figure 32).

Figure 32 Montequilla in the Northern Andean Yungas. Left: typical steep sided valleys with Coca plantations in the foreground, **right:** dense semitropical forest.

Widespread domestic infestation by *Tr. infestans* has been reported in the northern Yungas (188, 335). A study has also demonstrated the potential for *Panstronglyus rufotuberculatus*, thought to be a primarily silvatic vector, to actively colonise local domestic foci (336). In line with study objectives in Santa Cruz department, fieldwork was primarily undertaken to identify possible silvatic mammalian reservoir hosts. Domestic foci were searched for triatomines, but the region has been subject to a concerted vector eradication campaign in recent years, and none were found.

Cage traps were set for silvatic mammals on forested upper and lower valley slopes within 5km of the community (Total trap nights = 350). Local hunters were also engaged to trap further animals. A total of 28 mammals were captured, including *Coendou sp.*, *D. marsupialis*, *M. nudicaudatus*, and *S. spadiceus* (Table 19). Both *D. marsupialis* and *M. nudicaudatus* were tracked using the spool-and-line technique but no evidence of triatomine infestation was found in their nests. *M. nudicaudatus* behaved predominantly terrestrially, never entering the canopy and nesting on the forest floor. *D. marsupialis* did enter the canopy but was tracked to both arboreal and terrestrial nests. Noireau traps were

124

placed in silvatic locations, including tree holes and other putative nesting sites, but no triatomines were recovered. *Attalea spp.* palms were absent from this locality.

Table 19 Mammal captures and associated trypanosomes in Montequilla, Yungas region Dept. La Paz

Species	Captured	Infected*	Prevalence %	Trypanosome species
Coendou sp. (Porcupine)*	1	0	0	n/a
D. marsupialis[j]	6	0	0	n/a
Metachirus nudicaudatus (Brown four-eyed opossum)[b]	20	0	0	n/a
S. spadiceus[a]	1	0	0	n/a
Total	28	0	0	n/a

*By blood agar culture and/or xenodiagnosis.
[a]Caught with the aid of local hunters
[b]Caught in cage traps

No infection of silvatic mammals with any trypanosome species was identified at this site.

4.1.1.4 Cotopachi, Cochabamba Department

Cotopachi (17°26'S 66°17'W) is an open area of dry Andean puna located ~14km south west Cochabamba city at an elevation of ~2600m. The nearest human dwellings occur ~2km from the site in the town of Quillacollo. Puna vegetation consists of open thorn scrub interspersed with rocky outcrops and large, spiny, cacti (Figure 33). Average rainfall in this region is low - 800mm. Temperatures fluctuate both daily and seasonally, with day time highs of 35°C and night time temperatures approaching freezing. By contrast to the arid, exposed hilltops of Cotopachi, the surrounding valleys are sheltered, humid, systematically irrigated, and intensively cultivated by the local inhabitants (Figure 33). The human population density around this study site was the highest of any sampled in Venezuela or Bolivia.
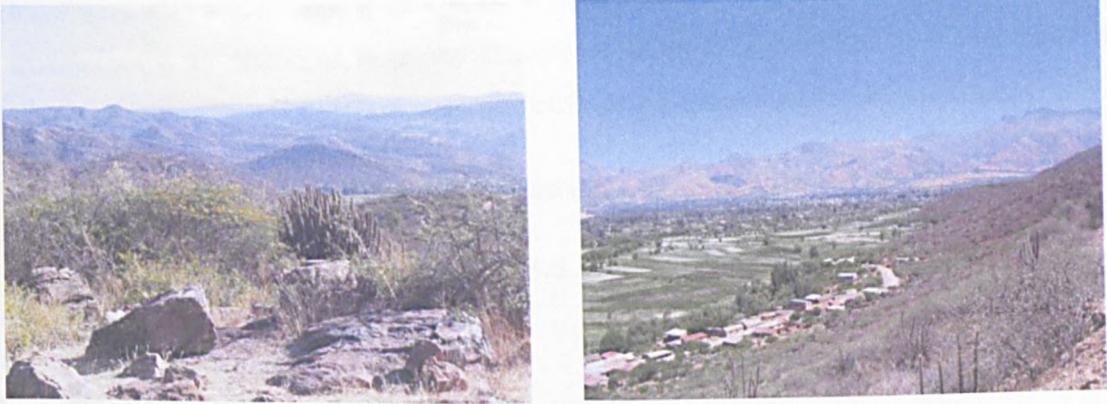
Figure 33 Left – Dry Andean puna vegetation at the Cotopachi study site, **right –** the lush Caramarca valley situated beneath the Cotopachi site.

The valleys around Cochabamba are long established as a region hyper-endemic for domestic Chagas disease transmission. In the early 1990s studies demonstrated seroprevalence as high as 74% in communities in the south of the department (90). The primary domestic vector species is *Tr. infestans*. Recent vector control strategies have successfully eradicated domestic *Tr. infestans* from much of this region, while drawing attention to locally occurring wild populations of the same species (44). Cotopachi is the site of three recent studies assessing the distribution wild *Tr. infestans* and their putative silvatic mammalian host species (43, 116, 222). Work was undertaken at Cotopachi in 2004 to corroborate some of their key findings, again with the primary focus on establishing the identity of wild mammalian reservoir species. A full discussion of data gathered here in context with results from previous studies in included in Section 4.4.

Mammal cage traps were set in three trap lines around the Cotopachi site and baited with two different mixes on alternate nights, one to attract rodents, which have been previously shown to carry *T. cruzi* at this site, and one to attract marsupials (Section 3.2.2) (222). Some traps were also placed around the village of Caramarca ~3km from the Cotopachi site in the lush valley bottom (Figure 33). Total number of trap nights was 640.

61 mammals were captured, including five species: *Phyllotis ocilae*, *Akodon boliviensis*, *Didelphis albiventris* and *Galea musteloides* (Table 20). *D. albiventris* and *G. musteloides* were both captured in Caramarca while the rest were captured in Cotopachi. The most frequently encountered species were *P. ocilae* and *A. boliviensis*. Mammals were tracked to

126

their burrows using the spool-and-line technique and Noireau traps placed inside. Noireau traps were also placed among rock piles in Cotopachi and in tree holes in Caramarca. No triatomines were recovered from any location. A number of peridomestic mammals were observed in and around the Cotopachi site, including goats and packs of feral dogs.

Table 20 Mammal captures and associated trypanosomes in Cotopachi, Dept. Cochabamba.

Species	Captured	Infected*	Prevalence	Trypanosome species
Phyllotis ocilae	23	3	10.3	*T. cruzi*
Akodon boliviensis (Bolivian grass mouse)	33	2	6.1	*T. cruzi*
Didelphis albiventris (White-eared opossum)	2	0	0	n/a
Galea musteloides (Quis)	3	0	0	n/a
Total	61	5	8.2	

*By blood agar culture and/or xenodiagnosis

Overall parasite prevalence regardless of mammal species was 8.2%. No infection was attributed to *D. albiventris* or *G. musteloides*, although both these species were caught in very low numbers.

4.1.2 Venezuela

4.1.2.1 Barinas State, Venezuela

A brief topographical description of Barinas state is included in Section 3.2.1.2. The province covers 35,200km², 3.8% of Venezuela, and is home to an estimated 756,600 people (337). The state spans two major biomes. In the west, descending eastwards from elevations of ~800m in the foothills of the western Andean Cordillera, to ~150m at the beginning of the Llanos region, the landscape is dominated by dry forest and patches of savannah grassland. Tree species are similar to those encountered in El Beni (Section 4.1.1.1), and include semi-deciduous *Ceiba* and *Cordia* species as well as numerous *Attalea* species palms. At elevations below ~150m in the eastern and south eastern portions of Barinas state, patchy woodland gives way to seasonally flooded grassland plains known as the Llanos. The wet season lasts from November to April and mean annual rainfall in Barinas is ~1600mm.
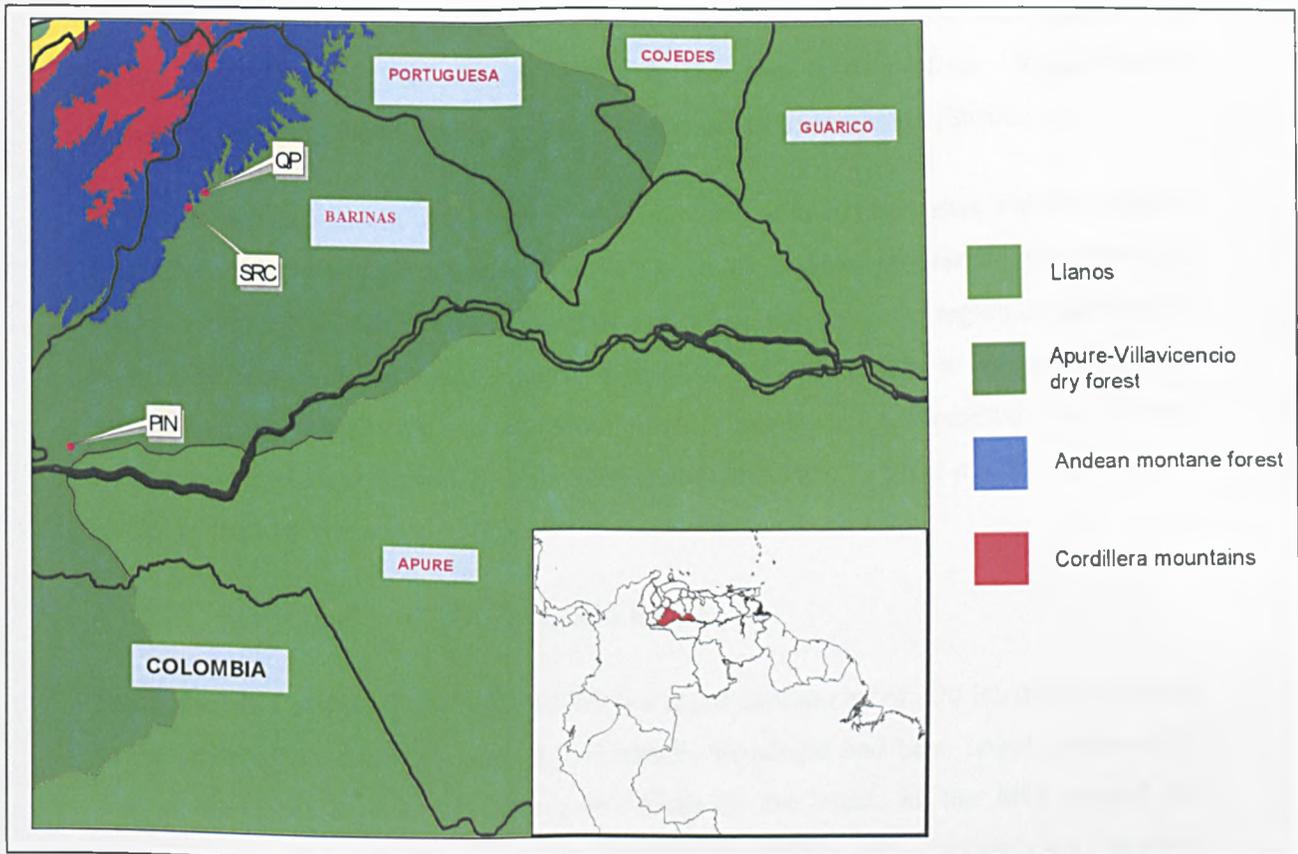
Figure 34 Map of study sites within Barinas state, Venezuela. State boundaries are defined by heavy black lines. Colours represent distinct ecotopes (See legend). Study sites: **QP** – Querbradon del Pescado; **SRC** – San Rafael de Catalina; **PIN** – Pinolito. Ecotope divisions defined as in Olson *et al.*, (2001) (269)

Little of the original expanse of semi-deciduous woodland in Barinas state survives today. Extensive deforestation for timber (2066.1km^2 between 1997 and 2001 alone) and widespread land clearance for cattle have radically transformed the landscape (75). Patches still remain in the mountainous west, although much of the valuable timber has been selectively removed. Most of the state is given over to large cattle ranches. Land ownership is highly inequitable, with over 60% of land in the hands of just 2% of the population (338). The majority of the rural inhabitants of Barinas live in a band of communities along the western fringe of the state coincident with land poorly suited to rearing cattle. Most live hand-to-mouth as subsistence farmers, growing plantain, manioc, rice and other food crops.

Barinas is one of the few remaining states in Venezuela still endemic for Chagas disease (82). The incidence of disease is low by comparison to hyper-endemic regions in Bolivia (Section 4.1.1.4, Section 1.5). Nonetheless a recent study demonstrated seroprevalence of

128

3.3% among individuals sampled across ten different localities and 8% infestation in 125 houses searched (75). The primary vector in the area is *R. prolixus* although other secondary vectors are also thought to play a lesser role in transmission (Section 1.4).

Fieldwork was undertaken in Barinas in 2004 and 2005 at three study sites with the primary goal of identifying silvatic mammal reservoir hosts of *T. cruzi*. Substantial gene flow has been reported between domestic and silvatic vector populations in the region on the basis of both DNA sequence and morphometric data (42). It is reasonable to assume that these vectors may be introducing parasite from silvatic mammals into domestic foci. A clear understanding of local mammalian abundance and associated *T. cruzi* infection prevalence is thus an epidemiological necessity.

4.1.2.1.1 San Rafael de Catalina, Estado Barinas

San Rafael de Catalina (8°30'N 70°44'W) is a small community of ~70 inhabitants situated in the north-west of Barinas state. Local natural woodland had been largely replaced by subsistence crops typical of the region, while to the west, in the hills around the community, forest remains relatively undisturbed (Figure 35). Peridomestic animals, primarily goats, pigs, and chickens, were kept in small quantities. Mammal traps were limited in availability in Venezuela, but a small number were set in nearby forest (Total number of trap nights = 24). Local hunters were also engaged to trap silvatic mammals.

Table 21 Mammal captures and associated trypanosomes in San Rafael de Catalina, Barinas State

Species	Captured	Infected*	Prevalance %	Trypanosome species
*D. novemcinctus*ᵃ	6	3	50	*T. cruzi*
Proechimys semispinosus (Spiny rat)ᵇ	2	0	0	n/a
Rattus norvegicus (Brown rat)ᶜ	1	1	100	*T. cruzi*
Total	**9**	**4**	**44.4**	

*By blood agar culture and/or xenodiagnosis.
ᵃCaught with the aid of local hunters
ᵇCaught in cage traps
ᶜCaptured peridomestically

A total of nine mammals were recovered including three species: *D. novemcinctus, P. semispinosus* and *R. norvegicus* (Table 21). *R. norvegicus* was captured peridomestically from an outbuilding by villagers. Villagers also reported the local presence of *D. marsupialis* in this region but this could not be corroborated. *D. agouti* was also sighted on

hunting expeditions to the forest but not captured. One juvenile *D. agouti* was kept by the villagers as a pet and both *D. agouti* and *D. novemcinctus* were regularly hunted for food. Overall trypanosome abundance in mammals captured was 44%, although neither of the two *P. semispinosus* captured harboured parasite.

Figure 35 The landscape around San Rafael de Catalina, Barinas showing lowlands cleared for agriculture surrounded by forested uplands

Although triatomine capture was not the primary goal of the expedition, some Noireau traps were set around the village and mammal burrows searched. A total of seven triatomines were recovered, five *Rhodnius sp.* nymphs from palms, and two *Panstrongylus sp.* nymphs from different armadillo burrows. All of the *Rhodnius* nymphs and one *Panstrongylus* nymph were infected with *T. cruzi*.

4.1.2.1.2 Quebradon del Pescado, Barinas State

Quebradon del Pescado (8°28'N 70°35'W) lies ~15km north east of San Rafael de Catalina in similar hilly terrain. Querbradon del Pescado is a smaller and more diffuse community than San Rafael de Catalina, containing approximately ~50 inhabitants living in small farmsteads dispersed among patches of forest and cultivated land.

Figure 36 The landscape around Quebradon del Pescado, Barinas. Left - the Western Cordillera mountains in the background; **right** - entering a patch of forest in a valley bottom adjacent to land cleared for grazing.

Mammal traps were set in forest patches around farmsteads and hunters were also engaged to trap mammals. Further mammal traps were also set 6km east of the community, on lower lying land around *Attalea spp.* palms in an attempt to catch marsupials. A total of 42 mammals were recovered including five species: *D. novemcinctus*, *D. marsupialis*, *Conepatus sp.*, *Caluromys sp.*, and *Tam. tetradactyla* (Table 22). Total number of trap nights was 140.

Table 22 Mammal captures and associated trypanosomes in Quebradon del Pescado, Barinas State

Species	Captured	Infected*	Prevalence %	Trypanosome species
Caluromys sp. (Woolly tailed opossum)°	1	0	0	n a
D. novemcinctus°	33	4	12.1	*T. cruzi*
D. marsupialis°	4	3	75	*T. cruzi*
Conepatus sp. (Skunk)°	2	0	0	n a
Tam. tetradactyla°	2	2	100	*T. rangeli*
Total	42	9	21.4	

*By blood agar culture and/or xenodiagnosis.
° Caught with the aid of local hunters
° Caught in cage traps

D. novemcinctus was by far the most abundant species captured in this region. Over half of these (19/33) were juvenile, weighing less than one kilogram. Despite a greater trap effort employed at this site than at San Rafael de Catalina, no *D. marsupialis* were captured in the upland region. Successful trapping of *D. marsupialis* was only achieved in more low-lying riverine forest 6km east of the community, near *Atallea spp.* palms, where this species was more abundant.

131

Trypanosome prevalence, regardless of mammal or trypanosome species, was 21.4% (Table 22). Highest *T. cruzi* prevalence was noted in *D. marsupialis* (75%). Interestingly, as in El Beni, Bolivia, all *Tam. tetradactyla* captured were found to be infected with *T. rangeli*.

Where possible, mammals were tracked to their nesting sites or burrows using the spool-and-line technique, but no triatomines were recovered. Triatomine capture was not a primary goal of this expedition and no Noireau traps were set in palms, nor were comprehensive searches of houses made.

4.1.2.1.3 Pinolito, Nr. Santa Barbara, Barinas State

Pinolito (7°30'N 71°14'W) is situated in the south-western extreme of the low-lying Llanos region of Barinas state beside a tributary of the Apure River. As with the rest of the Llanos region, the landscape in Pinolito consists of low lying seasonally flooded plains, interspersed with patches broad leafed forest abundant with *Attalea* species palms. Pinolito as a community exists to provide labour for large cattle ranches that dominate the region's agricultural economy. As in Querbradon del Pescado, houses dwellings widely dispersed across large expanses of grassland and forest. Mammal traps were set in forest patches around the community and local hunters engaged to trap further mammals.

Table 23 Mammal captures and associated trypanosomes in Pinolito, Barinas State

Species	Captured	Infected*	Prevalence %	Trypanosome species
Aloutta seniculus (Red howler monkey)*	1	0	0	n/a
*D. novemcinctus*ª	2	1	50	*T. cruzi*
*D. marsupialis*ᵇ	7	6	85.7	*T. cruzi*
Total	10	7	70	

*By blood agar culture and/or xenodiagnosis.
ªCaught with the aid of local hunters
ᵇCaught in cage traps

Ten mammals were captured including three species: *D. marsupialis, D. novemcinctus,* and *A. seniculus* (Table 23). As with the lowland region sampled in Querbradon del Pescado (Section 4.1.2.1.2), *D. marsupialis* was relatively more abundant at this site than in upland areas. The low capture success of *D. novemcinctus* was largely attributable to the paucity of local hunters. Although only one primate was captured, *A. seniculus*, this was observed in large numbers feeding on the fruit of *Atallea spp.* palms around the site.

Mammals were tracked using the spool-and-line technique but no triatomines were recovered. Six adult *R. prolixus* were gathered by a householder living nearby a large stand of *Atallea* palms. These triatomines were apparently attracted to the light. The lack of triatomine colonies within the house corroborated this observation. All six bugs were infected with *T. cruzi*.

Overall reservoir host *T. cruzi* infection at this site was relatively high regardless of species (70%). Infection rates within *D. marsupialis* were exceptionally high (85.7%), and comparable with those observed at Querbradon del Pescado. Capture success of other species was too low to draw any meaningful conclusions.

4.2 Trypanosome genotyping results

In total 356 mammals were captured including at least 27 species spanning eight different orders (Table 24). Of these, 113 were shown to carry trypanosomes. 233 triatomines were captured including at least seven species and three genera (Table 25). Of the 189 examined, 90 were infected with trypanosomes. All 203 isolates were genotyped and, where possible, species and genetic lineage identified using the techniques described in Section 3.5. Due to contamination in the early stages of culture, as well as desiccation and mortality of field-caught triatomines, not all field samples generated enough genomic DNA to be subjected to high resolution analysis, as described in Chapters 5&6, or to be characterised beyond species level. Sample panels associated with later analyses are included in the relevant sections (See Chapters 5&6)

Genotypic characterisation of field isolates was made by comparison to reference strain DNA. For *T. cruzi* these were: X10 clone 1 – TCI, CANIII – TCIIa, Esmeraldo clone 3 – TCIIb, M5631 – TCIIc, SC43 clone 1 – TCIId, and Cl Brener – TCIIe. The *T. rangeli* isolate employed was R1271. A complete list of all parasite genotypes defined in this study is found in Table 24 (mammals) and Table 25 (triatomines).

4.2.1   Mini-exon PCR-FLP strain characterisation

Mini-exon PCR product sizes generated from reference strains were consistent with those described in the literature (Figure 37 & Figure 38)(184). TCI generated a strong band at 350bp; TCIIa a strong band at 400bp (although a weaker band at 350bp was also observed is some gels (Figure 37)); TCIIb, TCIId, and TCIIe a strong band at 300bp; and TCIIc a generally weaker band at 250bp. Interestingly, amplification of a mini-exon product from *T. rangeli* DNA was also achieved, and a distinctive ~550bp generated, along with, in some instances, a slightly weaker band at 350bp (Figure 38). In most cases some unexpected, weaker bands were also generated, indicative of a degree of non-specific amplification.

All isolates derived from both triatomines and mammals were subjected to analysis of the mini-exon gene (See Section 3.5.2.1). While the technique generated ample product to detect TCI, TCIIa and TCIIb/d/e isolates, identification of TCIIc was less efficient. The Brazilian reference strain (M5631) produced a clear 250bp band (Figure 37), as did TCIIc strains from Paraguay isolated in a previous study (29). By contrast, only one strain in this study, SAM6, isolated from a *D. novemcinctus* caught at San Antonio in the Chaco region, could be unequivocally classified as TCIIc (Data not shown). Other isolates classified as TCIIc using the *hsp60* RFLP (Section 4.2.3) showed little or no product using the mini-exon PCR-FLP.
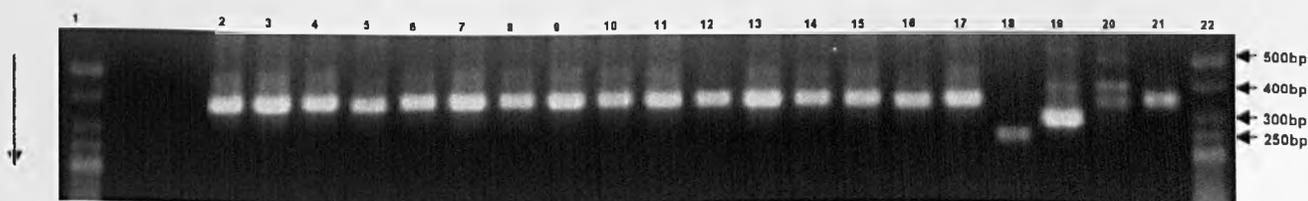
Figure 37 Mini-exon PCR-FLP products derived from a representative selection of silvatic mammals and triatomines caught during this study. Lanes: 1&22 – DNA ladder; 21 – X10 Clone 1 (TCI); 20 – CANIII (TCIIa); 19 – ESM Clone 3 (TCIIb); 18 -M5631 (TCIIc); 17 - M13; 16 – PARAMA39; 15 - SJM3; 14 – SJM41; 13 – SJMC7; 12 – SJM18; 11 – COTMA9; 10 – COTMA38; 9 – COTMA55; 8 – SMA3; 7 – SMA4; 6 - SRP2a; 5 – SRP2b; 4 - SRP2c; 3 – M12; 2 – M15.

Figure 37 shows mini-exon bands derived from *T. cruzi* isolates cultured from mammals and triatomines captured in Bolivia and Venezuela. All demonstrate the 350bp band typical of TCI. M13, M12, M15 and PARAMA39 correspond to samples derived from *D. marsupialis* caught in Pinolito and Querbradon del Pescado in Barinas state, Venezuela.

SJM3, SJM41 and SJMC7 are derived from *D. marsupialis, P. opossum* and *S. spadiceus,* respectively, all captured in San Juan de las Aguas Dulces, El Beni, Bolivia. SMA3 and SMA4 are derived from *P. opossum* and *D. marsipialis* respectively and were caught in Santa Maria de Apere, also in El Beni, Bolivia. COTMA9, COTMA38 and COTMA55 were isolated from murid rodents (*A. boliviensis* and *P. ocilae*) at the Cotopachi study site, Cochabamba, Bolivia. The remaining isolates in the figure, SRP2a, SRP2b, and SRP2c are derived from silvatic triatomines (*Rhodnius sp.* nymphs) in San Rafael de Catalina, Barinas State, Venezuela.

Mini-exon analysis of DNAzol®-extracted DNA from triatomine faeces in El Beni, Bolivia demonstrated the presence of double bands, indicative of mixed trypanosome species and genotype co-infections within these organisms (Figure 38). SJMOR23 and ABAC4 originate from silvatic *R. pictipes* adult and a domestic *Rhodnius sp.* nymph respectively. Both are infected with TCI, typified by a single 350bp product. LORI19, from a silvatic *Rhodnius sp.* nymph captured in an *Attalea spp.* palm shares the characteristic 350bp TCI band, but also shows amplification of a faint but distinctive band at ~550bp. This band is shared with the *T. rangeli* reference strain and suggests that both species of trypanosome are present in the sample. LORI12, LORI1 and NUAL1 all show amplification of two fragments at 300 and 350bp indicative of co-infection with TCI and TCIIb, d or e. All three were discovered at domestic or peridomestic foci. LORI12 and NUAL1 originate from *R. pictipes* while LORI1 was isolated from *Tr. sordida.*
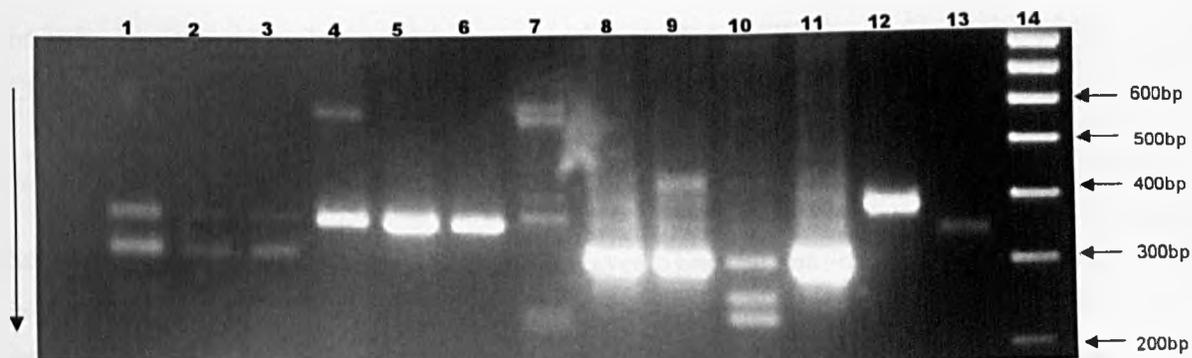
Figure 38 Mini-exon PCR-FLP analysis of triatomine gut homogenate from El Beni, Bolivia. Lanes: 14 – DNA ladder; 13 – X10 Clone 1 (TCI); 12 – CANIII (TCIIa); 11 – ESM Clone 3 (TCIIb); 10 – M5631 (TCIIc); 9 – SC43 Clone 1 (TCIId); 8 – ClBrener (TCIIe); 7 – M6421 (*T. rangeli*); 6 – SJMOR23; 5 – ABAC4; 4 – LORI19; 3 – LORI2; 2 – LORI1; 1 – NUAL1.

4.2.2   24Sα rRNA PCR-FLP strain characterisation

24Sα PCR-FLP was carried out only on strains that that could not be classified as either TCI or TCIIc (Sections 4.2.1 and 4.2.3). PCR amplification products generated were identical to those described in the literature (Figure 39) (278). Full identification of isolates classified as either TCIIb, d or e by mini-exon PCR-FLP from El Beni was not possible as the DNA concentration in DNAzol® - extracted samples was too low to permit multiple PCR-FLP analyses, a significant limitation of this technique.

Figure 39 24Sα rRNA PCR-FLP analysis of strains isolated from silvatic *Triatoma sp.* nymphs in the Bolivian Chaco. Lanes: 1&11 – DNA ladder; 10 – X10 Clone 1 (TCI); 9 – CANIII (TCIIa); 8 – ESM Clone 3 (TCIIb); 7 – M5631 (TCIIc); 6 – SC43 Clone 1 (TCIId); 5- ClBrener (TCIIe); 4 – N(5)a; 3 – N(5)c; 2 – Negative control.

136

In Figure 39, N5(a) and N5(c) are both samples derived from silvatic *Triatoma sp.* nymphs captured in an unidentified mammal nest in SAN in the Bolivian Chaco (Section 4.1.1.2.1). In both cases these samples share the same double banded 110/125bp profile as the SC43 Clone 1 reference strain, typical of hybrid *T. cruzi* lineage TCIId.

4.2.3 *Hsp60* PCR-RFLP strain characterisation

Strains for which no amplification could be achieved using the mini-exon PCR-FLP, or, as in the case of SAM6 (Section 4.2.1), that demonstrated a band at 250bp, were subjected to PCR-RFLP analysis of the *hsp60* gene as detailed in Section 3.5.3.1. Amplification fragment sizes obtained and restriction digest products were identical to those described in the literature (Figure 40)(191).
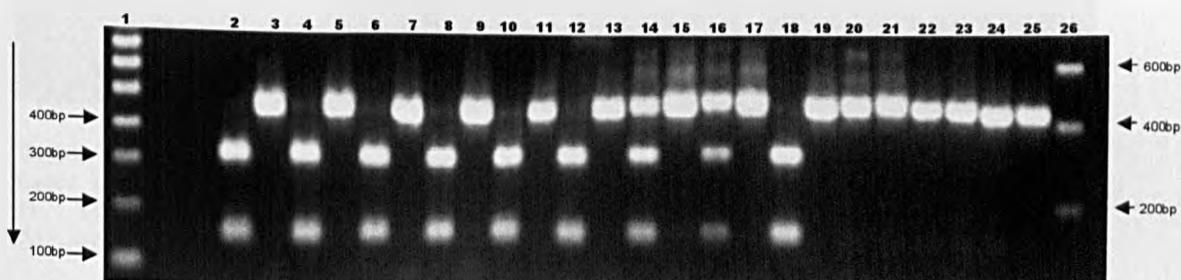
Figure 40 *Hsp60* **PCR-RFLP analysis of selected** *T. cruzi* **strains from silvatic mammals.** Lanes: 1&26 – DNA ladder; 25 – X10 Clone 1 (TCI) PCR product; 24 - X10 Clone 1 (TCI) restriction digest; 23 – CANIII (TCIIa) PCR product; 22 - CANIII (TCIIa) restriction digest; 21 – Esm Clone 3 (TCIIb) PCR product; 20 - Esm Clone 3 (TCIIb) restriction digest; 19 – M5631 (TCIIc) PCR product; 18 – M5631 (TCIIc) restriction digest; 17 – SC43 Clone 1 (TCIId) PCR product; 16 - SC43 Clone 1 (TCIId) restriction digest; 15 – ClBrener (TCIIe) PCR product; 14 - ClBrener (TCIIe) restriction digest; 13 – SMA9 PCR product; 12 – SMA9 restriction digest; 11 – PARAMA34 PCR product; 10 - PARAMA34 – restriction digest; 9 - MA204 PCR product; 8 – MA204 restriction digest; 7 – SAM6 PCR product; 6 – SAM6 restriction digest; 5 – SJMC10 PCR product; 4 – SJMC10 restriction digest; 3 – CAYMA3 PCR product; 2 - CAYMA3 restriction digest.

Figure 40 shows *hsp60* PCR products and their restriction digest profiles for seven reference strains and six representative strains from silvatic mammals. Host species and geographic localities are as follows: SMA9 was isolated from *D. novemcinctus* at Santa Maria de Apere in northern El Beni, Bolivia; PARAMA34 from *D. novemcinctus* at Querbradon del Pescado in Barinas state, Venezuela; MA204 from *D. novemcinctus* at Guiterrez in the Bolivian Chaco; SAM6 from *D. novemcinctus* at SAN in the Bolivian Chaco; SJMC10 from *D. novemcinctus* at San Juan de las Aguas Dulces in El Beni, Bolivia; and CAYMA3 from *E. sexinctus* at Quatro Canadas in Santa Cruz, Bolivia. All these strains share the same PCR and restriction profile: a band at 432-462bp which

generates only two bands after enzymatic digestion, one at ~315bp and one at ~120bp. This profile is identical to that of the TCIIc reference strain, M5631.

4.2.4 D7a divergent domain PCR-FLP characterisation

Strains identified as *T. rangeli* by a 550bp product using the mini-exon PCR-FLP, were confirmed by D7a divergent domain PCR-FLP analysis (Section 3.5.5). PCR products obtained were identical to those described in the literature (Figure 41)(185).
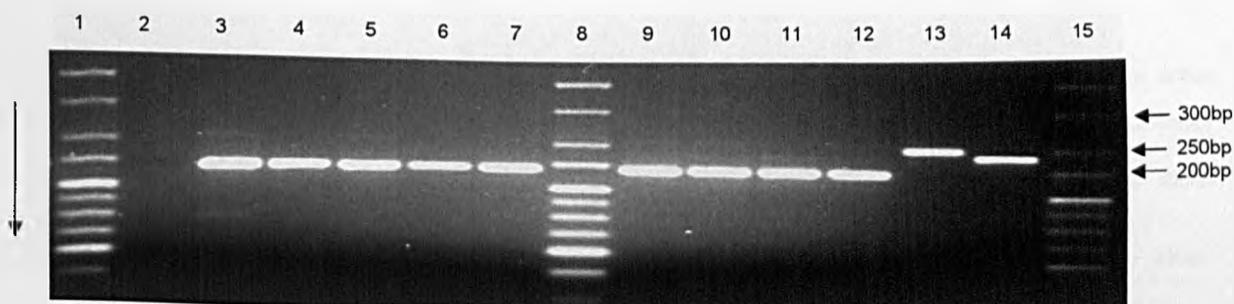
Figure 41 D7a divergent domain analysis of selected strains from silvatic triatomines and mammals. Lanes: **1, 8&15** – DNA ladder; **14** - X10 Clone 1 (TCI); **13** - Esm Clone 3 (TCIIb); **12** – R1271 (*T. rangeli*); **11** – SJM29; **10**-SJBUGAC2; **9** – SJM30; **7** – SJMC1; **6** – SJMC2; **5** – SJM8; **4** – PARAMA31; **3** – PARAMA13; **2** – Negative control.

Figure 41 shows the results from D7a divergent domain PCR-FLP analysis of selected strains. SJM30, SJMC1, and SJMC2 all originate from *Tam. tetradactyla* captured at San Juan de las Aguas Dulces, El Beni, Bolivia. PARAMA31 and PARAMA13 are from the same species captured at Querbradon del Pescado, Barinas state, Venezuela. SJM8 was isolated from a *P. opossum* at San Juan de las Aguas Dulces, SJM29 from a *S. sciureus boliviensis* at San Juan de las Aguas Dulces and SJBUGAC2 from a *Panstrongylus sp.* nymph captured from an armadillo burrow at the same locality. All share the same 210bp PCR product as the *T. rangeli* reference strain, R1271. All eight strains were also morphologically similar to *T. rangeli* on microscopic examination of the epimastigote form, as defined in Hoare, 1972 (140).

4.2.5 kDNA minicircle PCR-FLP characterisation

The same eight *T. rangeli* strains analysed in Section 4.2.4 were also subjected to analysis using a kDNA PCR-FLP (Section 3.5.5.1) to determine their genetic lineage (Figure 42). A

138

number of subtypes of *T. rangeli* have been defined in the Americas, coincident with major phylogenetic subdivisions within the tribe Rhodniini (339). A *T. rangeli* reference strain from only one of these subtypes was available for comparison. Three fragments were amplified from the kDNA of all eight strains. SJM30 (Lane 8) and SJMC2 (Lane 6) showed weak amplification of a third 165bp band which are only just visible in the figure. This three-banded amplification profile (760, 340-450, 165bp) is typical of the KP1(+) *T. rangeli* subgroup described in the literature (158).
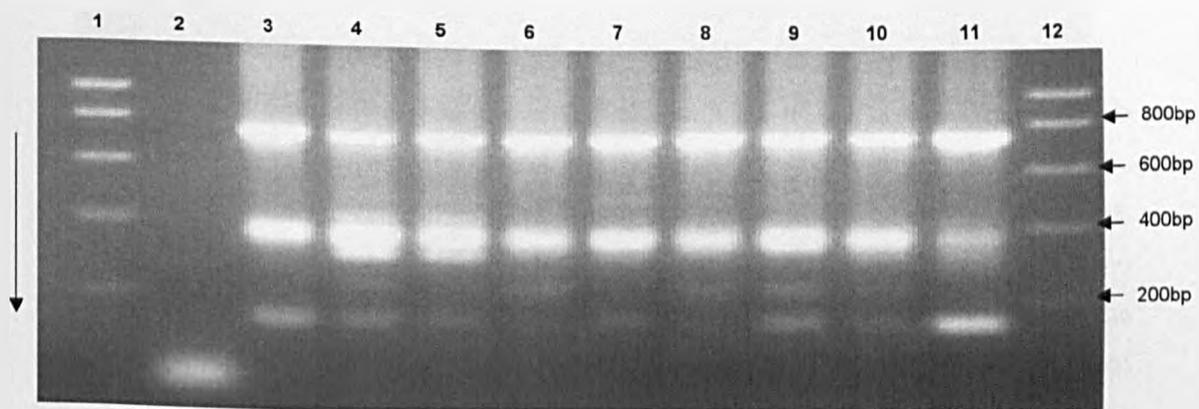
Figure 42 kDNA minicircle PCR-FLP characterisation of eight *T. rangeli* strains. Sample loading order is identical to lanes 12-2 in Figure 41. Lanes 1&12 contain DNA ladder.

4.2.6 Mini-exon gene intergenic region PCR-FLP characterisation

According to the literature, sequence length polymorphism in a fragment of the mini-exon gene intergenic region also supports a bipartite genetic subdivision within *T. rangeli* populations (158). PCR-FLP analysis was undertaken using the same eight strains as in Sections 4.2.4 & 4.2.5. Results are displayed in Figure 43.
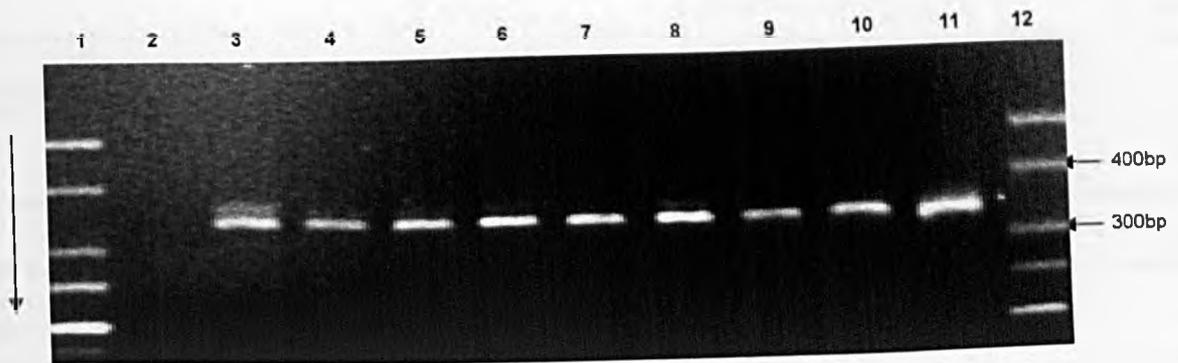
Figure 43 Mini-exon gene intergenic region PCR-FLP characterisation of eight *T. rangeli* strains. Sample loading order is identical to that in Figure 41.

A fragment of approximately 340bp in length was amplified from all eight strains, consistent with the KP1(+) genotype described in the literature. The lack of a KP1(-) reference strain, however, which is reported to produce a 380bp fragment, complicates the precise classification of the samples in this analysis. Nonetheless, the apparent concordance between this result and the previous figure strongly suggests that the these samples do belong to the KP1(+) genotype, and this was the reason that the two analyses were undertaken in parallel.

Order	Species	Captured	T. cruzi Infected	Prevalence %	Genotype TCI	Genotype TCIIc	Undefined (T. cruzi)[b]	T. rangeli[a] Infected	Prevalence %
Cingulata	D. novemcinctus	88	37	42.0	0	28	9	0	0
	E. sexcinctus	25	5	20	0	4	1	0	0
	C. vellerosus	21	1	4.7	0	1	0	0	0
	Tol. matacus	13	0	0	0	0	0	0	0
	Total	147	43	29.2	0	33	10	0	0
Artiodactyla	Ta. tajacu	2	0	0	0	0	0	0	0
	Total	2	0	0	0	0	0	0	0
Chiroptera	Eumops sp.	2	0	0	0	0	0	0	0
	Phyllostomus sp.*	2	0	0	0	0	0	0	0
	Total	4	0	0	0	0	0	0	0
Carnivora	N. nasua	1	0	0	0	0	0	0	0
	Conepatus sp.	2	0	0	0	0	0	0	0
	Total	3	0	0	0	0	0	0	0
Didelphimorpha	D. marsupialis	61	43	70.5	40	0	3	1	1.6
	P. opossum	22	12	54.5	7	0	5	1	4.5
	M. nudicaudatus	20	0	0	0	0	0	0	0
	D. albiventris	2	0	0	0	0	0	0	0
	Caluromys sp.	1	0	0	0	0	0	0	0
	Total	106	54	51	47	0	8	1	0.9
Primata	S. sciureus boliviensis	1	0	0	0	0	0	1	100
	Aotus sp.	3	0	0	0	0	0	0	0
	A. seniculus	1	0	0	0	0	0	0	0
	Total	5	0	0	0	0	0	1	20
Rodentia	D. agouti	8	0	0	0	0	0	0	0
	S. spadiceus	3	1	33.3	1	0	0	0	0
	Unidentified rodent	9	n/a	n/a	0	0	0	n/a	n/a
	Coendou sp.	1	0	0	0	0	0	0	0
	P. ocilae	23	3	13.0	3	0	0	0	0
	A. boliviensis	33	2	6.1	2	0	0	0	0
	G. musteloides	3	0	0	0	0	0	0	0
	P. semispinosus	2	0	0	0	0	0	0	0
	R. norvegicus	1	1	100	1	0	0	0	0
	Total	83	7	8.4	7	0	0	0	0
Pilosa	Tam. tetradactyla	5	0	0	0	0	0	5	100
	M. tridactyla	1	0	0	0	0	0	0	0
	Total	6	0	0	0	0	0	5	83.1
	Grand Total	356	105	29.2	54	33	18	8	2.2

*Infected with unidentified non-T. cruzi trypanosomes.
[a] All T. rangeli belonged to the KPI(+) genotype.
[b] Generated a T. cruzi – type hsp60 product but digestion failed.

Table 25 Total triatomine captures, associated trypanosomes and genotypes.

Species	Total captured	Domestic	Prevalence %	TCI	TCIIb, d, or e	Peri-domestic	Prevalence %	TCI	TCIIb, d or e	Undefined (T. cruzi)*	Silvatic	Prevalence %	TCI	TCIIc	TCIId	T. rangeli	Undefined (T. cruzi)*
P. geniculatus	2	0	n/a	0	0	0	n/a	0	0	0	2	0	0	0	0	0	0
Panstrongylus nymph	9	0	n/a	n/a	n/a	0	n/a	0	0	0	9(2)	22.2	0	1	0	1	
Psammolestes sp.	13	0	n/a	n/a	n/a	11	n/a	n/a	n/a	n/a	2	n/a	n/a	n/a	n/a	n/a	n/a
R. pictipes	9	0	n/a	0	0	5(4)	80	2	2ᵇ	2	4(4)	100	3	0	0	0	1
R. robustus	9	1(1)	100	1	0	2(2)	100	0	0	2	6(5)	83.3	4	0	0	0	1
R. prolixus	6	0	0	0	0	0	n/a	0	0	0	6(6)	100	6	0	0	0	0
Rhodnius nymph	147	3(3)	100	3	0	25 (6/23)	26.1	5	0	1	119 (45/77)	58.4	40	0	0	2ᵇ	6
Triatoma sordida	10	4(2)	50	2	2ᵇ	6(3)	50	0	1	2	0	n/a	n/a	n/a	n/a	n/a	n/a
Triatoma infestans	3	3	0	0	0	0	0	0	0	0	0	n/a	n/a	n/a	n/a	n/a	n/a
Triatoma nymph	25	12	0	0	0	10(4)	40	0	0	4	3(3)	100	0	0	2	0	1
Total (infected)	233	23(6)	26	6	2	59 (19/57)	33.3	7	3	11	151 (65/109)	59.6	53	1	2	3	8

Numbers in brackets represent number infected or number infected / number examined.

* Generated a *T. cruzi* – type *Hsp60* product but digestion failed.

ᵇ Part of a mixed infection with TCI.

4.3 Historical data analysis

647 records of *T. cruzi* infected silvatic mammals and 391 records of infected silvatic triatomines with associated genotypic data were gathered from the literature dating back to the first such study conducted by Miles *et al.,* 1981(164). These data were then collated with field data generated in this study to provide, as far as possible, a full record of *T. cruzi* genotypic diversity in silvatic mammalian reservoir and triatomine vector species (Table 26 & Table 27).

4.3.1 Silvatic mammalian reservoir species and associated *T. cruzi* genotypes.

Genotypic data are available for isolates derived from 74 species of mammal spanning eight orders. Mammals had been captured throughout the entire geographical range of *T. cruzi*, from the southern states of the USA to southern Argentina. With a view to examining possible associations between parasite lineage, host identity, host ecology and geographical distribution, a number of analyses were undertaken. Initially the frequency of occurrence of genotypes was tested for dependence on host taxonomic order using Pearson's χ^2 statistic applied to a 6x6 contingency table of host order against parasite genotype. The null hypothesis of independence between these two variables was strongly rejected ($\chi^2 = 882$, df = 25, p<<0.001). Orders which contained only one representative species, as well as mammals for which TCII subgroup was not defined, were excluded from the analysis. Among those data excluded were also 38 undefined TCII isolates from a captive group of *Leontopithecus rosalia*, thought to be the result of an epizootic outbreak originating from triatomine infested enclosures close to human habitation (226).

Secondarily mammals were categorised by ecological niche as being predominantly arboreal or terrestrial (156). This division is potentially over-simplistic - few mammals are obligate tree dwellers – and many, including *N. nasua, P. lotor*, and occasionally *Didelphis spp.*, are eclectic, foraging in many ecological strata from the canopy to the forest floor. In light of this, all species classified as arboreal were those thought to generally nest in trees, regardless of their foraging activity. Mammals designated as terrestrial were those that nest exclusively in burrows, rock piles, or among fallen logs. Again a Pearson's χ^2 statistic was

applied to test for dependence between niche and genotype and strong support for such an association generated ($\chi^2 = 289$, df = 5 p<<0.001). Mammals for which TCII subgroup was not defined, including the 38 *Leontopithecus rosalia* isolates mentioned previously, were again excluded from the analysis.

Table 26 Combined table of *T. cruzi* genotypes derived from mammals collected in this and previous studies.

Order	Species	Common name	TCI	TCIIa	TCIIb	TCIIc	TCIId	TCIIe	TCII (undefined)
Cingulata	*Chaetophractus vellerosus*	Screaming hairy armadillo				1(1)			
	Chaetophractus spp.	Hairy armadillo				1			
	Dasypus novemcinctus	Nine-banded armadillo	3	6	.	47(28)	1		
	Euphractus sexcinctus	Six-banded armadillo			1	5(4)	1		
Artiodactyla	*Sus scrofa*	Wild boar	1						
Chiroptera	*Molossus molossus*	Common bat	4	1					
Carnivora	*Conepatus chinga*	Hog-nosed skunk				1	1		
	*Nasua nasua**	Ring-tailed coati		2					
	*Procyon lotor**	Racoon	1	14					
Didelphimorpha	*Caluromys lanatus*	Woolly opossum	1						
	Didelphis albiventris	White-eared opossum	85		2				2
	Didelphis aurita	Common opossum	1						
	Didelphis azarae	Common opossum	5						
	Didelphis marsupialis	Common opossum	350(40)	1	1			1	2
	Didelphis spp.	Common opossum	5						
	Didelphis virginiana	Common opossum	2						
	Gracilinanus agilis	Agile gracile opossum	1						
	Marmosa cinerea (= Micoureus demerarae)	Woolly mouse opossum	1						
	Metachirus nudicaudatus	Brown four-eyed oposum	1						
	Monodelphis brevicauda	Short-tailed opossum		3			1		
	Monodelphis domestica	Gray short-tailed opossum	1				1	2	
	Philander frenata	Grey four-eyed oposum	4						4
	Philander opossum	Grey four-eyed oposum	23(7)		1				
	Thylamys elegans	Elegant fat-tailed opossum	4		.				
Primata	*Aotus nigriceps*	Owl monkey	1	1					
	Callicebus moloch	Dusky titi	1						
	Callithrix chrysoleuca	Gold and white marmoset	1						
	Cebus albifrons	Capuchin monkey	1						
	Cebus apella	Tufted capuchin monkey	1						
	Chiropotes satanas	Black bearded saki	1						
	Leontopithecus chrysopygus	Black lion tamarin							1
	Leontipithicus rosalia	Golden lion tamarin							38[a]
	Macaca mulatta	Rhesus macaque	1						
	Saguinus bicolor bicolor	Pied tamarin	2						
	Saguinus midas niger	Black-handed tamarin	1						
	Saguinus ochraceus	Bare-faced tamarin	1						
	Saimiri scurieus	Squirrel monkey	3	1					

144

Table 26 Continued from over leaf

	Species	Common name								
Pilosa	*Bradypus torquatus*	Three-toed Sloth	1							
	Cyclopes didactylus	Silky anteater	1							
Rodentia (Caviomorpha)	*Coendou spp.*	Porcupine	5							
	Dasyprocta fuliginosa	Agouti				1				
	Dasyprocta spp.	Agouti	2							
	Echymis chyrsurus	Arboreal spiny rat	1							
	Echymis dasythrix	Arboreal spiny rat	1							
	Octodon degus	Degu	1							
	Octodon sp.	Degu						1		
	Trichomys aperoides	Rock rat	14	1						1
	Trichomys pachyurus	Rock rat	1							
Rodentia (Other)	*Akodon boliviensis*	Bolivian grass mouse	5(2)							
	Akodon sp.	Grass mouse	1							
	Bolomys lactens	Rufous-bellied bolo mouse	5							
	Clyomys laticeps	Brown-headed spiny Rat	1							
	Holochilus braziliensis＊	Marsh rat	1							
	Nectomys squamipes	Water rat	1							
	Oecomys mamorae	Mamore arboreal rice rat	1							
	Oryzomys capito	Rice rat	2							
	Oryzomys scotti	Rice rat	1							
	Peromyscus sp.＊	Deer mouse	1							
	Phyllotis ocilae	Leaf-eared mouse	14(3)							
	Proechimys semispinosis	Spiny rat	3	1						
	Rattus rattus	Black rat	7	1						
	Sciureus spp.	Squirrel	1	1						
	Sciureus spadiceus	Amazon squirrel	1(1)							
	Tylomys mirae	Climbing rat	3							

Arboreal species are bold and underlined.
Numbers in parentheses are field isolates collected in this study.
*Eclectic behaviour but generally nests arboreally.
＊Likely to be an epizootic outbreak of human origin. See text.
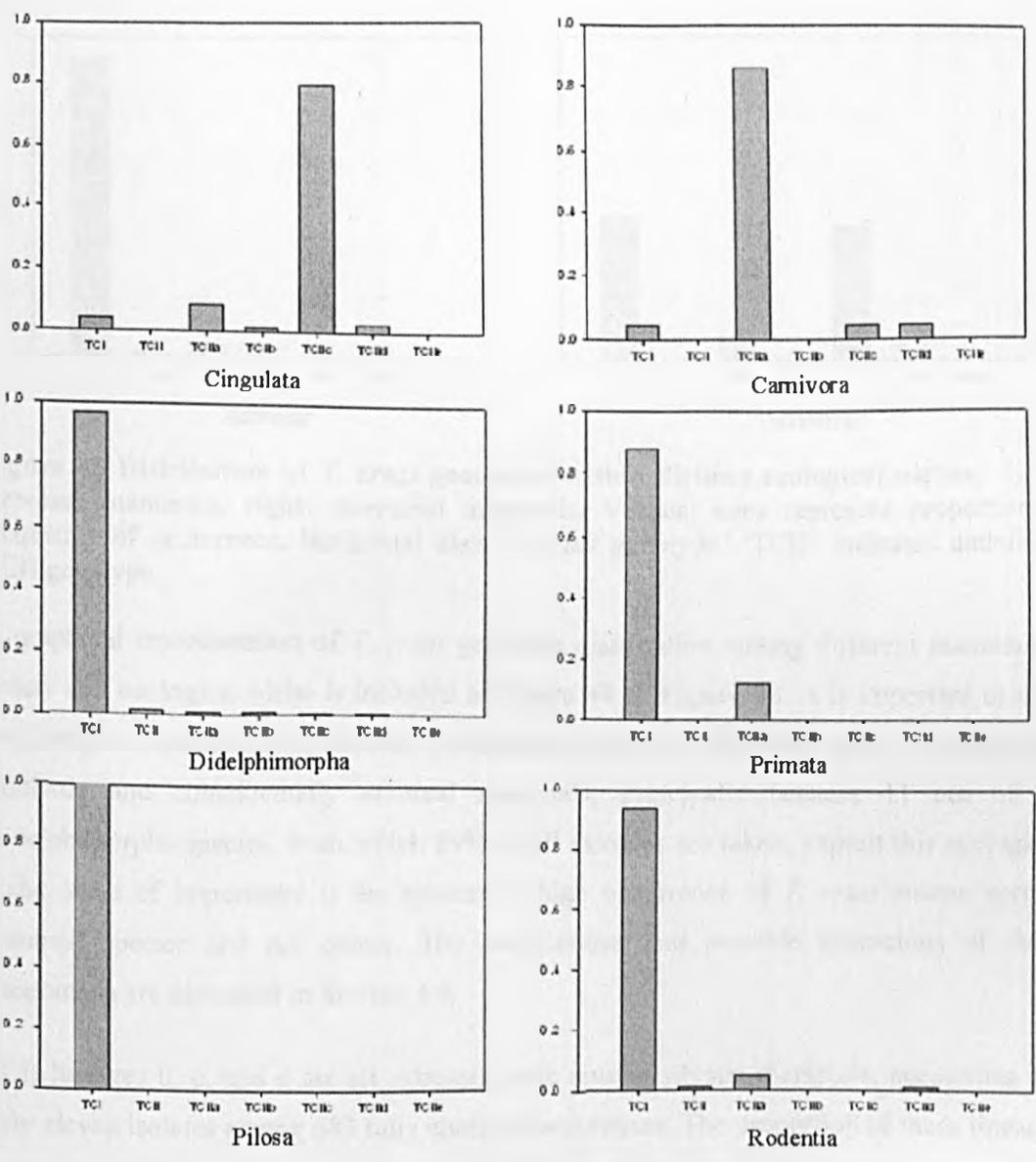
Figure 44 *T. cruzi* genotype frequencies among different six orders of neotropical mammal. Clockwise from bottom left: Y axes indicate proportionate frequency of each genotype. X axes indicate each individual genotype. 'TCII' indicates undefined TCII genotype.
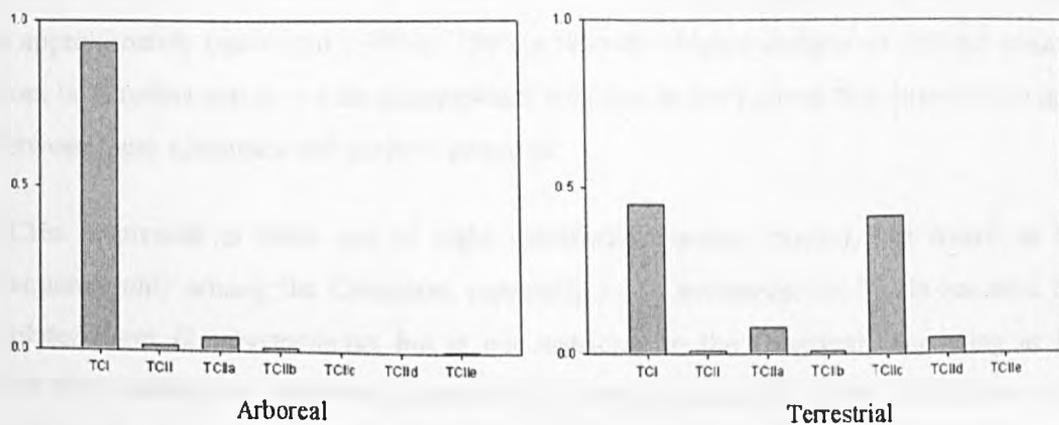
Figure 45 Distribution of *T. cruzi* genotypes within distinct ecological niches. Left: Arboreal mammals, **right:** terrestrial mammals. Vertical axes represent proportionate frequency of occurrence, horizontal axes *T. cruzi* genotype. 'TCII' indicates undefined TCII genotype.

A graphical representation of *T. cruzi* genotype distribution among different mammalian orders and ecological niche is included in Figure 44 & **Figure 45**. It is important to note that sample sizes are highly skewed towards three orders, Didelphimorpha, Cingulata and Rodentia and coincidentally arboreal mammals, principally because 11 out of 15 Didelphimorpha species, from which 69% of all samples are taken, exploit this ecological niche. Also of importance is the apparently high occurrence of *T. cruzi* among certain mammal species and not others. The implications and possible limitations of these phenomena are discussed in Section 4.4.

TCII lineages b, d, and e are all extremely rare among silvatic mammals, accounting for only eleven isolates among 685 fully characterised strains. The proportion of these lineages is marginally highest among Cingulata (~5.6%) and even greater when all terrestrial mammals are grouped together (~5.0% in terrestrial mammals compared with ~0.9% in arboreal mammals).

By far the most common lineage among all silvatic mammals is TCI, which accounts for ~80% of all isolates. TCIIc accounts for a further ~8% and TCIIa ~5%. TCI is distributed throughout all mammalian orders but occurs proportionately at its lowest among the Cingulata, followed closely by the Carnivora. The highest relative abundance of TCI occurs among Pilosa and Primata (100% excluding *Leontopithecus spp.*), although both orders are

147

under-represented in the data. The frequency of TCI among Rodentia and Didelphimorpha is approximately equivalent (~90%). The considerably higher number of isolates obtained from both orders across a wide geographical area (see below) points to a strong association between these mammals and parasite genotype.

TCIIc is present in three out of eight mammalian orders studied, but found at high frequency only among the Cingulata, especially in *D. novemcinctus*. TCIIa has also been isolated from *D. novemcinctus* but is not restricted to the Cingulata, occurring at high frequency among the Carnivora, especially in North America (27, 340). TCIIa also occurs in Rodentia, Chiroptera, Primata and two species of Didelphimorpha.

When mammals are divided by ecological niche (**Figure 45**) the vast majority of arboreal species are infected with TCI, with a smaller number infected with TCIIa. TCI and TCIIa also occur in terrestrial mammals. Interestingly ~80% of the terrestrial TCI isolates belong to rodents. Perhaps the most important observation of all, however, is the complete absence of TCIIc from arboreal mammals and its apparently exclusive association with terrestrial species.
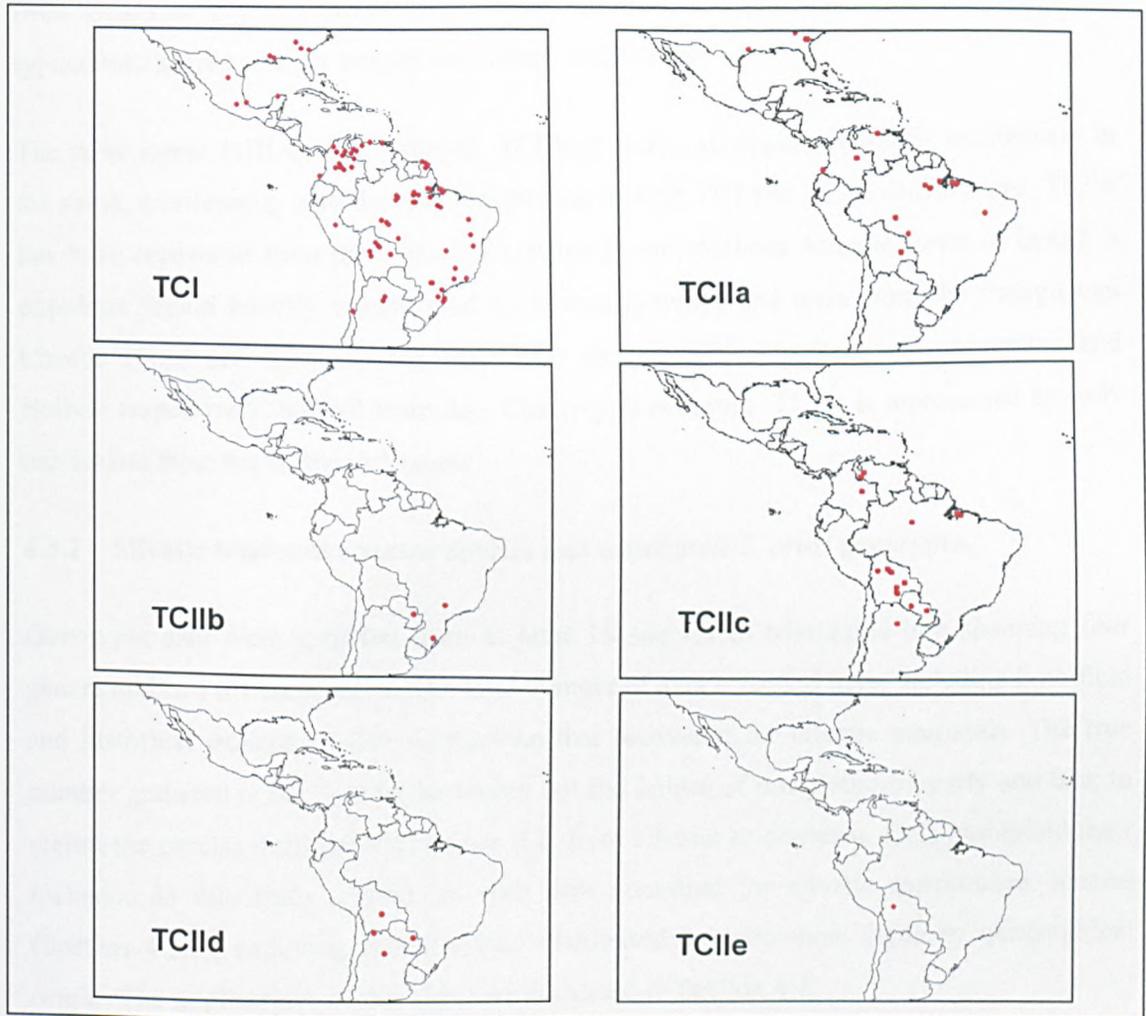
Figure 46 Geographical distribution of major *T. cruzi* lineages isolated from silvatic mammals in the Americas. Red points indicate sample sites.

The geographical distribution of TCI and TCII *T. cruzi* genotypes from silvatic mammals demonstrates a high degree of sympatry (Figure 46). The distribution of TCII subgroups, while broadly overlapping in a number of cases, is apparently non-random, and a degree of allopatry may exist. TCIIa occurs in the northern portion of the area sampled, from the southern states of the USA as far south as the Santa Cruz Dept. Bolivia. By contrast, TCIIc is absent from the USA, but occurs from Venezuela further south than TCIIa, with the southernmost recorded isolate taken from the Argentine Chaco (27°12´S 63°02´W). Large gaps exist in the possible total distribution of TCIIa and TCIIc, as predicted by outlying samples. This is a likely phenomenon of poor sampling of these genotypes and/or their relative rarity by comparison to silvatic TCI isolates. The co-occurrence of both lineages

149

from localities in north-eastern Brazil and Colombia is a key observation, and may be typical throughout a larger proportion of their distribution.

The three rarest TCII silvatic lineages, TCIIb, d and e, all occur apparently exclusively in the south, overlapping with the southern portion of both TCI and TCIIc distributions. TCIIb has been recovered from three localities, twice in the southern Atlantic forest in Brazil, a populous region heavily transformed by human activity, and once from the Paraguayan Chaco. TCIId has been isolated from only four localities in Paraguay, Argentina and Bolivia respectively, and all from dry, Chaco-type ecotopes. TCIIe is represented by only one isolate from the Bolivian Yungas.

4.3.2 Silvatic triatomine vector species and associated *T. cruzi* genotypes.

Genotypic data were gathered from at least 15 species of triatomine bug spanning four genera and two different tribes. The total number of data recorded here, including both field and historical isolates, is 440, lower than that recovered for silvatic mammals. The true number gathered is likely to be far higher but the failure of many studies, early and late, to record the precise origin of triatomines (i.e. from silvatic or domestic foci) prohibited their inclusion in this study. Again, as with data compiled for silvatic mammalian species (Section 4.3.1), sampling is not evenly distributed by triatomine clade or geographical origin. The implications of this skew are discussed in Section 4.4.

Table 27 Occurrence of major *T. cruzi* lineages among different species of silvatic triatomine bug

Tribe	Genus	Species	T. cruzi genotype						
			TCI	TCIIa	TCIIb	TCIIc	TCIId	TCIIe	TCII (undefined)
Rhodniini	*Rhodnius*	*Rhodnius spp.*	40(40)						
		R. pictipes	17(3)						
		R. pallescens	67						
		R. robustus	7(3)						
		R. paraensis	3						
		R. prolixus	48	2					
Triatomini	*Panstrongylus*	*Panstrongylus sp.*				1(1)			
		P. geniculatus	8	1		1			2
		P. lignarius	14	1		1			
		P. megistus	33		5				
		P. rufotuberculatus	1						
	Triatoma	*Triatoma sp.*					2(2)		
		Tr. infestans	120						
		Tr. sanguisuga	3						
		Tr. sordida	3						
		Tr. spinolai	51		1				
		Tr. tibiamaculata	3						2
	Eratyrus	*E. mucronatus*	3						

Numbers is parentheses correspond to samples collected in this study.

As in Section 4.3.1, the data were subjected to a Pearson's χ^2 contingency test. In this case triatomines were grouped by genus and *T. cruzi* genotype to either TCI or TCII, in view of the low sample size. Significant association was identified between triatomine genus and *T. cruzi* genotype ($\chi^2 = 34.7$, df = 2, p<<0.001). *E. mucronatus* was excluded from the analysis as it was the only representative of its genus for which data were available. A graphical representation of these data is shown in Figure 47.
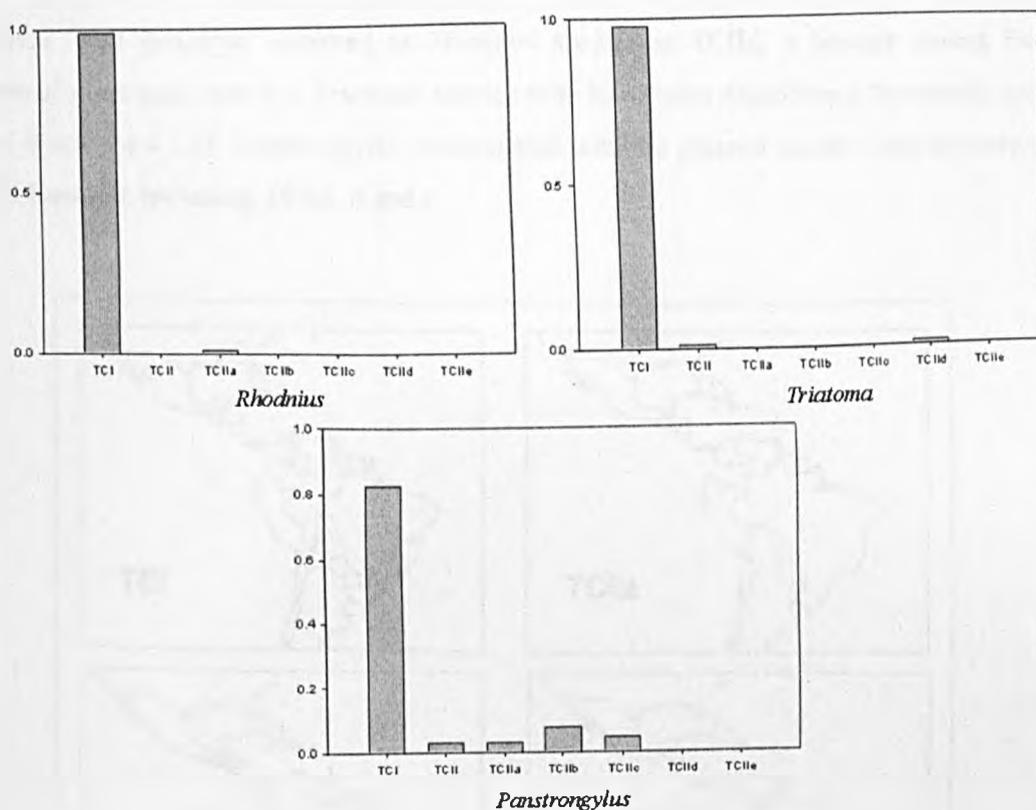
Figure 47 The distribution of TCI and TCII *T. cruzi* genotypes among different silvatic triatomine genera. Vertical axes represent proportionate frequency of occurrence, horizontal axes *T. cruzi* genotype.

Triatomine species were not subjected to analysis by ecological niche. Although a degree of niche specificity is thought to exist between different triatomine genera, a number of species, particularly within the Triatomini, behave eclectically, infesting both arboreal and terrestrial sites (Section 1.6.1).

Concordant with observations made in silvatic mammal populations (Section 4.3.1), TCI is by far the most common parasite lineage recovered from silvatic triatomines, accounting for ~95% of all samples. TCII isolates comprise the remaining ~5%, and are distributed evenly among the TCII genotypes, including undefined TCII lineages.

TCI is present in all three genera of triatomine but occurs at greatest frequency among *Rhodnius* species (98.9%), then *Triatoma* (97.3%), and lowest among *Panstrongylus* (82.4%). The only TCII clade represented among the exclusively arboreal *Rhodnius* is

TCIIa, perhaps correlating with the presence of TCIIa in arboreal mammals. The only fully defined TCII genotype observed in *Triatoma* species is TCIId, a lineage absent from arboreal mammals, and this *Triatoma* species may have been exploiting a terrestrial niche (See Section 4.4.1.1). *Panstrongylus* is associated with the greatest number and diversity of TCII lineages, including TCIIa, b and c.
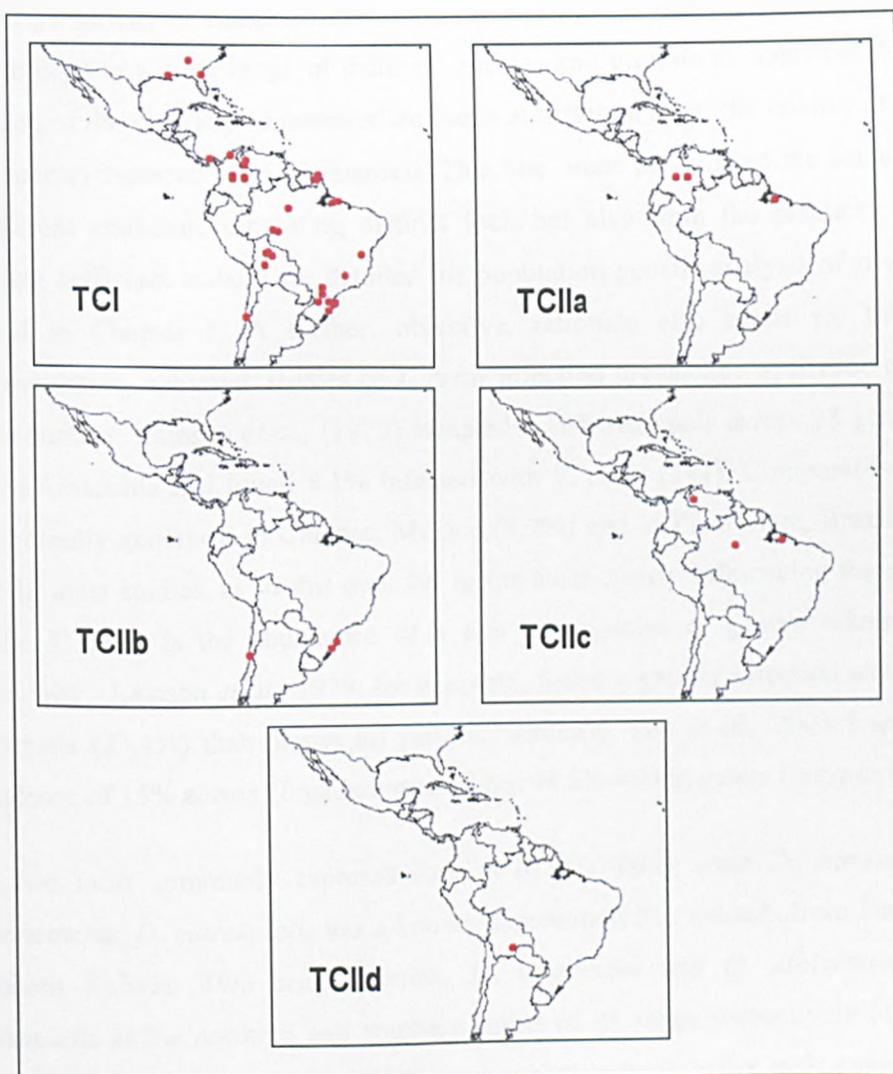
Figure 48 Geographical distribution of major *T. cruzi* lineages isolated from silvatic triatomines in the Americas. Red points indicate sample sites.

The geographical distributions of *T. cruzi* genotypes isolated from silvatic triatomines all broadly coincident with those from silvatic mammals. The principal exception to this

observation is attributed to a TCIIb isolate from northern Chile (27°12'S 71°00'W). The possible silvatic reservoir host of this genotype in this region has yet to be identified (28).

4.4 Discussion
4.4.1 Field data
4.4.1.1 Silvatic mammal distributions, associated trypanosomes and genotypes.

Al least 27 species of mammal were captured during this study. Although attempts were made to capture a wide range of different species and provide an approximately accurate reflection of the diversity of mammalian fauna at different sites, the relative proportions of those species captured were unbalanced. This bias arose partly from the actual abundance of different mammals inhabiting distinct foci, but also from the practical necessity of obtaining sufficient isolates for detailed the population genetic analysis of silvatic *T. cruzi* detailed in Chapter 5. A further, objective, rationale also exists for this approach. Comprehensive, unbiased, studies of *T. cruzi* infection prevalence in silvatic mammals are few in number. Lainson *et al.,* (1979) sampled 1,197 mammals across 28 genera in north-eastern Amazonia and found 8.1% infected with *T. cruzi* (341). Comparable results were more recently generated in Chiapas, Mexico (9.2%) and in Piaui State, Brazil (12%) (223, 342). In most studies, as in this one, the factor most clearly influencing the prevalence of silvatic *T. cruzi* is the abundance of a few key species or genera within the silvatic community. Lainson *et al.,* 1979, for example, found a greater infection among didelphid marsupials (23.1%) than across all genera. Similarly Yeo *et al.,* 2005 found an overall prevalence of 15% across 10 genera studied but 44.8% within genus *Dasypus* (29).

The two most commonly captured species in this study were *D. marsupialis* and *D. novemcinctus*. *D. marsupialis* has a known distribution that extends from Panama south to Northern Bolivia. Two sister species, *D. virginiana* and *D. albiventris*, replace *D. marsupialis* at the northern and southern limits of its range respectively (156). All three species are important reservoirs of *T. cruzi* where they co-occur with competent vectors, and the entire distribution of genus *Didelphis* extends from southern Canada to Patagonia. *D. albiventris* is also thought to replace *D. marsupialis* at high altitude (>2000m), as was observed in Cotopachi in this study, and in drier ecotopes in eastern and south-eastern Brazil (343). A further species, *D. aurita*, occurs in the Atlantic Forest of Brazil (344).

D. marsupialis is frequently found in close proximity to human populations. Some studies suggest that the abundance of this species might even be directly influenced by human disturbance (345). *D. marsupialis* was also found in abundance around human habitation in this study. House-holders reported this species stealing chicken eggs as well as killing and eating adult birds. Unlike *Monodelphis domestica*, a smaller marsupial species associated with humans, *D. marsupialis* does not generally nest in domestic or peridomestic structures (156). This species is known to nest among palms and in tree holes, consistent with spool-and-line tracking in this study, although few actual nests were located. A detailed study of a number of radio-tagged *D. marsupialis* in the Llanos region of Venezuela showed that nest location displayed a degree of seasonality (346). Although palms and tree holes were the predominant day time refuge in the dry season, a significant number favoured underground nesting localities in the wet season. Similarly, one individual at Montequilla in the Bolivian Yungas was tracked to a terrestrial locality. Thus, while it is clear that *D. marsupialis* has a strong affinity for an arboreal lifestyle, evidenced by numerous morphological adaptations including a prehensile tail, it is a consummate generalist, capable of adaptively modifying its foraging and nesting behaviour as circumstances require. Additionally, the mean range of one individual in a single night is thought to be over a kilometre (346). The widespread distribution of *D. marsupialis* throughout tropical lowland ecotopes in South America, as well as its relative uniformity as a species, likely stem from this capacity for adaptation and dispersal.

D. novemcinctus is present throughout an even greater proportion if the Americas than *D. marsupialis*, occurring from the southern states of the USA to Southern Argentina (156). This species is sympatric with a number of other armadillos throughout its distribution. In tropical clines these include *Dasypus kappleri*, *E sexcinctus*, and *Priodontes maximus*, and in dryer southern regions *C. vellerosus* andf *Tol. matacus*, among others. The relative abundance of *D. novemcinctus* was notably lower in the dryer Chaco region, especially in San Antonio, as compared to the moist forests of Venezuela and Northern Bolivia, for which it has a strong affinity. *D. novemcinctus* is hunted for food by the inhabitants of rural communities throughout South America. While there is increasing concern that these activities may be having an impact on endemic populations, *D. novemcinctus* is not yet

classified as endangered, and this species appears to be able to withstand moderate levels of human predation (347).

D. novemcinctus is an exclusively terrestrial mammal. By comparison to the burrows of other armadillo species discovered in this study, notably those of *C. vellerosus* an *E. sexicintus* in the Chaco, *D. novemcinctus* burrows in moist forests were relatively shallow. One nest found in Venezuela, and infested with *Panstrongylus* nymphs, was little more than a ball of leaves on the forest floor. Like *D. marsupialis, D. novemcinctus* it is an opportunistic omnivore and feeds on invertebrates, small amphibians, carrion and some plant matter (156). This generalist mode of nutrition, combined with rapid dispersal and a well documented ability to displace native fauna, probably accounts for the status of *D. novemcinctus* as a pioneer species and explains its rapid recent expansion into North America (348).

Three smaller marsupial species, *Caluromys sp., P. opossum,* and *M. nudicaudatus* were also found sympatrically with *D. marsupialis.* Although *P. opossum* and *M. nudicaudatus* look extremely similar, and share a contiguous range, the former is highly arboreal and the latter highly terrestrial in its ecology (156, 273). Despite occupying distinct niches, the two species did not co-occur in this study. *Caluromys sp.* was only encountered once. This specimen was captured by local hunters, and is the most arboreal of all marsupial species encountered (156).

No marsupials were encountered in Cotopachi or in the Chaco, although *Thylamys elegans* has been reported from the former locality (222). In both regions considerable trap effort was employed, using a range of trap sizes and bait types. This indicates that local marsupial population densities were probably very low. A degree of overlap apparently exists in the species compositions encountered through the tropical study sites from northern Amazonia into the drier Chaco, with tropical species like *D. marsupialis, Tayassu tajacu, D. agouti, Tam. tetradactyla, S. spadiceus* and *S. sciureus boliviensis* gradually giving way to dry-adapted armadillo genera. The rodent-dominated faunal composition of the Andean highland at Cotopachi is apparently discontinuous from those of the more tropical climes encountered at other study sites.

156

The overall prevalence of *T. cruzi* in silvatic mammals encountered in this study was 29.2%. *D. maruspialis* (70.5%), *P. opossum* (54.5%), and *D. novemcinctus* (42%) all showed high levels of infection, as predicted by previous studies (341, 349). Three further armadillo genera from which a greater number than ten individuals were sampled - *E. sexcinctus* (20%), *C. vellerosus* (4.7%) and *Tol. matacus* (0%) - showed proportionately lower levels of infection. The only species of marsupial found in significant quantities and not infected was *M. nudicaudatus*. It has been suggested that that the temporary nests this species make on the forest floor are unsuitable for infestation by triatomines (273). However, *D. marsupialis* found locally were also free of infection, a first in this study, and possibly an indicator of low silvatic vector abundance.

Among rodent genera, no infection was identified in terrestrial species inhabiting moist forest (*D. agouti* and *P. semispinosus*), despite a possible association with *Panstrongylus* species triatomines (Section 4.1.1.1.1). This is consistent with observations made in north-eastern Brazil (341). One arboreal rodent found in El Beni was infected, *S. spadiceus*, possibly correlating with a locally high density of *Atallea* species palms. Two species of terrestrial rodents inhabiting rocky ecotopes at Cotopachi were infected with *T. cruzi*, *A. boliviensis* (6.1%) and *P. ocilae* (13%), albeit at a substantially lower rate than *D. marsupialis*, *P. opossum* and *D. novemcinctus*, but corroborating findings from an earlier study (222).

The disparity of *T. cruzi* infection rate between different silvatic mammal species observed in this study, and in others, has many possible explanations. *T. cruzi* infectivity at a cellular level may be more effective in ancient endemic South American mammals, which includes the *Didelphidae* and *Cingulata* (Section 1.6.2). Similarly these mammals may be able to sustain longer term, asymptomatic, high parasitemic infections than others (350). Congenital infection, infrequent in humans (Section 1.1.2), could also be common in certain species. Finally, studies have demonstrated that the anal scent gland secretions of *D. marsupialis* can harbour metacyclic *T. cruzi*, and suggest that the parasite can be transmitted in the absence of a vector during territorial or defensive behaviour (351). The simplest explanation, however, is perhaps the most attractive. *T. cruzi* transmission by the faecal route is highly inefficient, especially in mammals with thick fur. All three key reservoir species identified in this study are omnivores with tendency towards insectivory.

All three tend to build their nests in sites susceptible to triatomine infestation – opossums in *Atallea* palms and *D. novemcinctus* in shallow burrows filled with leaf litter. These species may actively predate insects present in their refuges – an engorged triatomine represents a protein-rich food source – allowing *T. cruzi* to exploit a far more efficient transmission route.

Two *T. cruzi* genotypes were encountered among silvatic mammal species, TCI and TCIIc. In the non-Andean ecotopes studied, the affinity of these lineages with distinct ecological niches and host cliques was striking. *D. marsupialis, P. opossum,* and *S. spadiceus* – all predominantly arboreal species – exclusively harboured TCI, concordant with local silvatic vector species, where identified (Section 4.4.1.1). TCIIc was apparently limited to terrestrial species in this study, specifically the Pilosa (*D. novemcinctus, E. sexcinctus,* and *C. vellerosus*), and both sympatric and allopatric to TCI.

TCI was also identified from Andean rodents at Cotopachi, as was the case in a previous study (222). Interestingly, *T. cruzi* was not isolated from *G. musteloides* in this study, or in the one undertaken three years earlier at the same site. It has been postulated that *G. musteloides,* a caviomorph rodent and an ancestor of the modern Guinea pig, could be the reservoir host of *T. cruzi* Z2 (TCIIb) (100). *T. cruzi* has been isolated from this species before at silvatic foci near Cochabamba but never typed (352, 353). Furthermore, no evidence of any TCII lineage was detected in mammals or triatomines at this study site in this study or any other.

The association between TCI and terrestrial rodents is not limited to Cochabamba. A large number of the caviomorph rodent *Trichomys aperoides* were found to be infected with this lineage in rocky ecotopes in Piaui state, north-eastern Brazil (223). Terrestrial rodents captured in the Pantanal region, including *Clyomys laticeps,, Oryzomys scotti* and *Trichomys pachyurus* also showed evidence of infection with TCI (354). The precise ecotope from which these species were derived is not clear, as the Pantanal region comprises large wetlands as well as arid cerrado-type ecosystems similar to those in Piaui.

T. rangeli was isolated from eight mammals, *D. marsupialis, P. opossum, S. sciureus boliviensis* and *Tam. tetradactyla.* All four species of mammals are arboreal or semi-

arboreal, including *Tam. tetradactyla* which is thought to nest in hollow trees (156). All five *Tam. tetradactyla* encountered demonstrated infection with *T. rangeli*, despite being caught ~2500km apart. *Tam. tetradactyla* is a known host of *T. rangeli* and infected individuals have been identified from the Brazilian Amazon and French Guiana (355, 356). Natural infection of *Tam. tetradactyla* with *T. cruzi* has also been reported (140, 357). Efficient salivary transmission of *T. rangeli* is invariably associated with *Rhodnius* species triatomines. While the high rate of *T. rangeli* infection identified in *Tam. tetradactyla* is surprising, the low rate of infection identified among arboreal Didelphimorpha is perhaps more so. Infection of sympatric arboreal triatomines with *T. rangeli* was demonstrated (Section 4.4.1.1), yet only two isolates were derived from opossums. This observation highlights the potential difficulties of accurately sampling trypanosome populations from wild mammals. The growth kinetics of *T. cruzi* and *T. rangeli* in biphasic culture are distinct, and *T. rangeli* fails to grow effectively in standard liquid RPMI (data not shown). Additionally, the susceptibility of different triatomine species to intestinal *T. rangeli* infection is known to be different (358). None of the eight *T. rangeli* isolates identified were derived from xenodiagnostic triatomines, and the majority of mammals were captured in Bolivia, where *Tr. infestans* was the only available laboratory colony. The only mixed infection identified (TCI/*T. rangeli*) from an opossum (*D. marsupialis*) was derived from a DNAzol® extraction of a sample of bacterially-contaminated biphasic culture within ten days of field inoculation with whole blood. *T. rangeli* was observed to grow much more rapidly in culture than *T. cruzi*, reaching stationary phase and then dying off within a ten day period. It is possible that a number of *T. cuzi*/*T. rangeli* co-infections were missed, while only pure *T. rangeli* infections, where trypanosomes failed to survive in RPMI liquid culture, were noted, as DNA was consequentially extracted from biphasic culture.

The high rate of *T. rangeli* infection observed in *Tam. tetradactyla* suggests frequent exposure to *Rhodnius* species. Sympatric *Rhodnius* species demonstrated widespread infection with *T. cruzi*, and it is logical to assume that *Tam. tetradactyla* was also exposed to this trypanosome. However, *Tam. tetradactyla* is thick-furred, exclusively myrmecophagous, and, in contrast to key *T. cruzi* reservoir hosts, is unlikely to consume triatomines it encounters, providing circumstantial evidence for the importance of the oral route in maintaining silvatic *T. cruzi* transmission.

All the *T. rangeli* isolates genotyped from mammals belonged to the KP1(+) genotype. KP1(+) is reportedly associated with *R. prolixus* group triatomines, which includes *R. robustus*, a species that occurred sympatrically with infected mammals (Section 4.4.1.1) (339). It could not be confirmed whether a second *T. rangeli* lineage, possibly in association with *R. pictipes*, was also in circulation (see Section 4.4.1.2). Sympatric infection of mammal species with different *T. rangeli* lineages has been reported where associated vector distributions overlap.

Two trypanosome isolates from leaf-nosed bats (*Phyllostomus sp.*) were also discovered. No amplification of gene regions with any of the primers employed for genotyping *T. cruzi* or *T. rangeli* was possible. Amplification and sequencing of more highly conserved genetic markers, e.g. the SSU rRNA gene region, as described in pan-species studies of trypanosomatid evolution, is a logical approach to shedding light on the evolutionary affinities of these samples (160). Both isolates are held in cryogenic storage at the LSHTM, pending further analysis.

4.4.1.2 Triatomine distributions, associated trypanosomes and genotypes.

At least seven species of triatomine were recovered during this study, only three of which were captured from truly domestic transmission cycles: *Tr. sordida* and *R. robustus* from El Beni, Bolivia, and *Tr. infestans* from the Bolivian Chaco. *Tr. sordida* and *Tr. infestans* are widely reported as important vector species in Bolivia but the occurrence of a *R. robustus* specimen inside a house in the country is as yet unrecorded (96, 188, 333). *R. robustus* is more commonly understood to be a widespread silvatic vector, infesting palm species throughout Colombia, Venezuela, Bolivia, Ecuador, French Guiana and northern Brazil, although sporadic invasion, but not colonisation, of houses is reported (118, 359). Without molecular identification, it is impossible to ascertain to what species the three further domestic *Rhodnius* nymphs belong, but it is of epidemiological importance that all four triatomines were infected with *Tr. cruzi*.

Domestic *Tr. sordida* infestation was limited to the southern portion of the study area in El Beni, principally in San Juan de Mocovi and the surrounding communities. This species was unique in its exploitation of cracks in adobe walling - no other, captured domestically

or peridomestically, behaved in a similar manner. The reported silvatic habitats of *Tr. sordida* include bird nests, tree holes, bromeliads, and occasionally palm crowns (29, 97, 101). Peridomestic triatomines were recovered, but a silvatic focus could not be identified. *Tr. infestans* is well known to disperse passively with humans and it is entirely possible that *Tr. sordida* owes its presence in southern El Beni to a similar phenomenon (95). The lack of *Tr. sordida* infestation observed in the remote north of El Beni around San Cristobal and Santa Maria de Apere probably correlates with lower human immigration. Domestic and peridomestic *Tr. sordida* infestation in Santa Cruz Department is widespread – the nearest reported focus is ~400km south-east of San Juan de Mocovi in Velasco province – but others, unreported, could be even closer (97, 333).

Peridomestic searches, all in El Beni Department, yielded four different species of triatomine: *Tr. sordida, R. pictipes, R. robustus,* and *Psammolestes sp.* With the exception of two *Rhodnius* nymphs captured in a dog kennel, all were found in association with chickens nesting near houses. All four species are known to predate birds in the silvatic environment (100, 101). *Psammolestes spp.* are either unknown or extremely rare in a peridomestic ecotope - this is usually an exclusively silvatic genus strictly associated with birds - and some individuals were also found in bird nests in nearby thorn scrub. Experimental infection of some *Psammolestes sp.* with *T. cruzi* has been demonstrated, but natural infection is thought to be very rare, and these individuals were not examined for trypanosomes (101, 360).

The occurrence of *R. robustus* at peridomestic foci in El Beni is further evidence that this species could be of epidemiological importance. The discovery of *R. pictipes* co-infesting the same foci is also significant. One previous suspected discovery of *R. pictipes* in the Bolivia was made in Alto Beni, where this species was found infesting domestic and peridomestic foci, and unconfirmed reports of silvatic specimens in Bolivia from the 1940s, 1950s and 1970s do exist (359, 361, 362). The Alto Beni specimens were later attributed to *R. stali,* a morphologically similar species, and it was generally assumed that this species replaced *R. pictepes* in Bolivia (40, 102). *R. pictipes* adults from the current study were morphologically distinct from examples of *R. stali* held in La Paz by Dr. Mirko Rojas Cortez, as well as to examples in the literature, and indicate that both species are actually present Bolivia (102).

The majority of domestic and peridomestic triatomine infestation, in those communities systematically searched, was found in and around San Juan de Mocovi (12.9% of houses). The overall infestation rates of the Santa Maria de Apere and San Cristobal were lower (~3%), a likely result of lower population density in these regions, and a correspondingly smaller number of houses searched. The rate of *T. cruzi* infection among these triatomines was 26% in domestic samples, and 33% in peridomestic samples. This is significantly higher than the rate of infection found among domestic and peridomestic *Tr. sordida* in nearby Velasco, Dept. Santa Cruz (16.2%) and far higher than that observed in *R. stali* from foci in Alto Beni, where none were infected (40, 333). Most peridomestic triatomines discovered, including those infected, were found in association with chickens. These vectors must have picked up infection from another source, either from domestic/peridomestic mammals (including humans), or from the silvatic environment. Moreover, a number of these bugs were nymphs, and are thus extremely unlikely to have encountered the parasite from anything other than a local (domestic or peridomestic) source.

T. cruzi genotype data from domestic and peridomestic triatomines shed some light on the nature of local *T. cruzi* transmission, especially in the context of local occurring silvatic *T. cruzi* genotypes. Aside from isolates only definable to species level, two lineages, TCI and TCIIb, d or e, were present, as defined by miniexon PCR-FLP direct from bug faeces. Among the domestic isolates, four, one from an adult *R. robustus* and three derived from *Rhodnius* nymphs, belonged to the TCI genotype, and two, derived from adult *Tr. sordida*, contained mixed infections of TCI and TCIIb, d or e. Peridomestic isolates comprised five TCI, derived from *Rhodnius* nymphs, two TCI/TCIIb, d or e mixes from *R. pictipes*, and one TCIIb, d or e from a *Triatoma* nymph. Lineages belonging to either TCIIb, d or e were not recovered from the silvatic environment anywhere in El Beni (see below and Section 4.3.2). Moreover, analysis of historical isolates suggests that all three of these lineages are extremely rare in the silvatic cycle (Section 4.3). TCIIb, d and e lineages, especially TCIId, are common in Bolivian human populations and, while it remains a possibility that an important silvatic reservoir or vector was missed, at least some domestic and/or peridomestic triatomines captured in El Beni could have fed on humans (248, 363). TCI also occurs in human populations in Bolivia; thus it is hard to deduce whether these strains

were introduced from the immediate silvatic environment without higher resolution genetic analysis (227, 248). It is perhaps unsurprising that *Tr. sordida*, a known secondary vector in Bolivia, behaves anthrophilically. But the occurrence of TCIIb, d, or e in *R. pictipes*, albeit mixed with TCI infection, is an interesting observation. This species is primarily a silvatic vector, though some instances of adults within human habitations have been reported (101).

The triatomine capture effort employed outside El Beni department was significantly lower than that elsewhere. Nevertheless, effort was made to sample from as many distinct silvatic ecotopes as possible, particularly with a view to establishing reservoir host – vector relationships. *Rhodnius* species were by far the most commonly encountered vector, all in association with arboreal ecotopes and *Attalea* species palms. The association between *Rhodnius* and *Attalea* species is well established (100, 101). Three species: *R. prolixus* from Venezuela, and *R. pictipes* and *R. robustus* from northern Bolivia, were recovered. The status of silvatic *R. prolixus* populations in Venezuela is the subject of recent controversy, particularly with reference to its highly similar sister species, *R. robustus*, with which it shares much of its silvatic distribution in Venezuela and Colombia (42, 118). Nevertheless, other *Rhodnius* specimens captured from nearby localities in Barinas State were identified as *R. prolixus*, providing a degree of justification for the classification used here (364).

As mentioned in earlier in this section, the discovery of *R. robustus* infesting *Attalea* species palms in El Beni is a new but not unexpected observation, as they lie within the expected distribution of this species i.e. moist ecotopes in association with *Attalea* palms from north-western Brazil to Colombia and Venezuela (118). *R. pictipes* was found co-infesting the same sites. It has been suggested these two species prefer subtly distinct ecological niches, including different palm species and even different microhabitats within the same palm(100, 135) *R. robustus* is marginally lighter in coloration than *R. pictipes*, which could represent an affinity for conditions higher up in the palm, above the dense crown. As the vast majority of triatomines were caught with Noireau traps, and palm dissections yielded no adults, no specific observations concerning the micro-distribution of these two species could be made. However, no evidence was found for differential infestation of palm tree species here, and previous observations relate to a study undertaken in Amazonian Brazil in non-*Attalea* species (365).

Efforts were made to unequivocally establish the reservoir host of *Rhodnius* species at most localities in Bolivia and Venezuela by spool-and-line tracking, with only limited success. However, circumstantial evidence points to arboreal marsupials. Cage traps set among groves of *Attalea* palms were more successful in trapping marsupials than those set elsewhere, and Noireau traps placed in the same groves demonstrated the presence of *Rhodnius* species. In San Juan de Mocovi a number of engorged *R. pictipes* were found among a ball of cloth and plastic in the crown of an *Attalea* palm. All these triatomines were heavily infected with *T. cruzi* as well as some *T. rangeli*, suggesting that the nest belonged to a mammal. Additionally, during spool-and-line tracking of *D. marsupialis* and *P. opossum*, the trail was invariably lost among the upper fronds of palms. Marsupials were not the only species observed inhabiting *Attalea* species, however, whose abundant fruit attract a variety of silvatic fauna. *A. seniculus* in Venezuela and *S. sciureus boliviensis* in Bolivia were also noted, and could be important secondary hosts.

More success was had in establishing the silvatic host of *Panstrongylus* species captured in moist forest in both Venezuela and Bolivia. In Venezuela two *Panstrongylus* 2nd instar nymphs, identifiable by the morphology of the ante-ocular region, were found infesting the burrows of *D. novemcinctus* (101). *P. geniculatus* has been reported from armadillo burrows at numerous localities including Colombia, Venezuela, Brazil, and Trinidad in the West Indies (138, 224, 366, 367). Neither of the specimens captured could be identified to a species level. An attempt was made to grow these triatomines on to adult stage, but *P. geniculatus* is notoriously difficult to rear in the laboratory, and both perished (272). The association discovered between *P. geniculatus* and *D. novemcinctus* in El Beni, Bolivia, has never been previously recorded and is both testament to the widespread dispersal of the two species and the strength of their association. Nonetheless *P. geniculatus* was found to behave fairly eclectically within the terrestrial niche. Specimens were also recovered from the burrow of a second fossorial mammal, *D. agouti*, which it shared with a juvenile caiman (*Caiman crocodilus*).

The only silvatic triatomines discovered outside tropical moist forest biomes were three 5th instar *Triatoma* nymphs from the Bolivian Chaco at San Antonio. Three *Triatoma* species are recorded from this region, *Tr. guasayana*, *Tr. sordida*, and *Tr. infestans* (97, 115). The

size of the specimens (>20mm in length) likely precludes their classification as *Tr. sordida*, but further identification is not possible in the absence of molecular data (101). The Bolivian Chaco is thought to be an important focus for silvatic *Tr. infestans*, and additional foci have been isolated in the Paraguayan Chaco further south (115, 277). The extent of silvatic-domestic invasion of this species is this region is debatable and co-occuring domestic and silvatic populations are morphologically distinct. *Tr. guasayana* is recorded throughout the Chaco region, primarily in the silvatic ecotope, although peridomestic infestation has been reported (97, 131, 277). The silvatic hosts of both *Tr. infestans* and *Tr. guasayana* in the Chaco are not well defined but could include rodents, birds, and, in the case of *Tr. guasayana*, even toads (101). All three triatomines discovered were infected with trypanosomes, two definitely *T. cruzi*, and it seems that these must have fed on a mammal. However, as the individual who collected these triatomines was not able to confirm their precise origin, the identity of the host remains unknown.

Attempts to capture silvatic triatomines from another important focus of wild *Tr. infestans* at Cotopachi in Bolivia were unsuccessful (43). This failure probably arose from the lower capture effort employed in this study. Moreover, the study area was sampled in October, at the end of the cold season. Previous studies found a 50% decrease in triatomine abundance in this period, as compared with the summer months (November – March) (43, 222).

The majority of *T. cruzi* strains isolated from silvatic *Rhodnius* species in both Bolivia and Venezuela belong to lineage TCI. This is concordant with the high prevalence of TCI found in arboreal mammals in this and other studies, especially among the arboreal Didelphimorpha (Section 4.3.1) (29). The overall rate of infection among *Rhodnius* species in Venezuela (100%) was twice that noted in Northern Bolivia (~55%), although the rate deduced in Bolivia may be a more accurate estimate of overall infection because the sample size was a factor of ten higher. Nonetheless, both results suggest an extremely active silvatic transmission cycle. In line with observations made in the Amazonian region as a whole, where similarly active cycles are thought to occur, these vectors present a risk of adventitious infection to human populations through oral contamination and/or attraction to light, even if significant domestic colonization is infrequent (77). While no specific data are available for comparison from Bolivia, a study in Merida, an adjacent state to Barinas, identified a much lower prevalence of *T. cruzi* infection than observed here (18%) (368).

165

T. rangeli was another trypanosome found among silvatic *Rhodnius* species in El Beni, Bolivia, identified by a ~550bp product from mini-exon PCR-FLP, and present as a mixed infection with TCI in two nymphs. *T. rangeli* is usually transmitted to the vertebrate host via the proboscis during feeding, a mode of transmission that is closely linked with *Rhodnius* species. Parasites can occur intestinally in a variety of different triatomine genera, as evidenced by an isolate obtained from a *Panstrongylus* nymph in this study, but this is not thought to be the primary mode of transmission (140). Recent molecular studies have demonstrated population genetic structure in *T. rangeli*, apparently governed by strict association with two and possibly three species complexes within *Rhodnius* and their respective geography (158, 339). Only a small quantity of parasite DNA was obtained from the two *Rhodnius* specimens that harboured *T. rangeli*, however, and intra-species genotyping not possible.

A *T. rangeli* isolate obtained from the gut of a *Panstrongylus* nymph at San Juan de Mocovi was analysed and belonged to the KP1(+) genotype, one that corresponds with lineages isolated from the *R. prolixus* complex (*R. prolixus, R. robustus, and R. neglectus*) in previous studies, as well as those from locally caught mammals (158). *R. robustus* did occur sympatrically, though it can only be speculated what route the parasite could have taken to the hind-gut of a *Panstrongylus* nymph in a terrestrial niche. Recent maximum likelihood sequence analysis of two *T. rangeli* nuclear genes identified a further, divergent, *T. rangeli* lineage associated with *R. brethesi* and possibly *R. pictipes*, on the basis of circumstantial evidence (339). *R. pictipes* also occurred locally, but no means of identifying the putatively corresponding *T. rangeli* lineage is possible using the methodology described in the current study.

T. cruzi was also isolated from a *Panstrongylus* nymph, in this case in Venezuela. Strikingly this strain, identified using the *hsp60* PCR-RFLP, belonged to TCIIc, despite being found in an armadillo burrow within 100 metres of *Atallea* species palms in which TCI-infected *Rhodnius* species triatomines were abundant. *D. novemcinctus* captured in this area were also exclusively infected with TCIIc, suggesting a strong local relationship between parasite, niche, vector and host.

T. cruzi from silvatic *Triatoma* nymphs in the Bolivia Chaco were typed as hybrid lineages TCIId. On initial observation, this result is suggestive of an epizootic outbreak, as TCIId is primarily associated with humans in Bolivia (248). However, a recent study of silvatic reservoir hosts in the Paraguayan Chaco did identify the presence of TCIId in *E. sexcinctus* and *D. novemcinctus* (29). Again, without formal confirmation of the source of the infected triatomines, no concrete conclusions can be drawn.

4.4.2   Host, ecological and geographical correlates with parasite genotypic diversity.

The data presented in Section 4.3 are the most complete review of silvatic *T. cruzi* genotypic diversity ever assembled. Although it is over 30 years since molecular approaches were first applied to the study genetic variation in *T. cruzi*, the dataset is woefully incomplete (165). Sampling is geographically fragmented, skewed towards certain mammal and triatomine species and subject to confusion arising from epizootic outbreaks (226). Additionally genotyping techniques are inadequately standardised and TCII subgroups frequently left undefined, even in recent publications, despite clear phylogenetic evidence that two subgroups within TCII, TCIIa and TCIIc, have stronger affinity with TCI than they do with TCIIb, the most divergent clade (199, 216, 369).

Not all the apparent shortcomings of the last 30 years of sampling can be laid at the door of researchers. Many aspects of the currently incomplete dataset probably reflect the true biology of the system. TCI, the most frequently sampled and geographically dispersed genotype, probably is the most common genotype in silvatic transmission cycles. Silvatic TCII lineages are apparently much rarer. Nonetheless, TCIIc was identified in 42% *D. novemcinctus* in this study across a wide geographic range and 42.1% by Yeo *et al.*, (2005) in Paraguay (29). By comparison, the rate of infection indentified from *D. marsupialis* in this study was 70.5% across a wide geographic range, 55% by Steindel *et al.*, (1995) in south-eastern Brazil, and 72% by Travi *et al.*, (1994) in Colombia (370, 371). The number of genotypes identified in the historical data should approximately reflect the *T. cruzi* prevalence observed in these two mammals in the field. In fact, including those from this study, 350 TCI isolates have been recorded from *D. marsupialis*, and a meagre 46 TCIIc from *D. novemcinctus*. This represents clear and unacceptable historical sampling bias towards one species. It is perhaps no surprise that, although it is also an abundant silvatic

167

mammal in the Americas, *D. novemcinctus* requires considerably more time and effort to catch.

Comprehensive surveys like that of Lainson *et al.,* (1979) do offer some biological basis for this skew (341). Infection prevalence is certainly higher among some mammal species than others, even when a more balanced approach is taken, the possible reasons for which were discussed earlier (Section 4.4.1.1). In light of sampling bias, the results generated from the statistical analysis of the entire *T. cruzi* dataset, as undertaken in Sections 4.3.1 & 4.3.2, must be approached with considerable caution. Nonetheless, there are clear patterns within the data that must be addressed, even if the relative abundance of certain genotypes may not reflect the true picture.

TCI is apparently present throughout the known geographical distribution of *T. cruzi* and two distinct silvatic niches for this genotype emerge from the data. The first, and most frequently sampled, is predominantly arboreal and from tropical or semi-tropical biomes, involving *Rhodnius* species and possibly some representatives of the Triatomini, as well as arboreal mammals. Among these, the highest number of infections have been reported among the arboreal *Didelphimorpha,* but sporadic infection has been reported from arboreal species within a number of other orders of mammal, including the *Primata, Rodentia, Pilosa*, and *Chiroptera*.

The second TCI silvatic niche affinity is that with terrestrial rodents from arid rocky ecotopes, and possibly triatomines of genus *Triatoma,* indeed, over one hundred TCI-infected silvatic *Tr. infestans* were recovered from around Cochabamba (222). Only three examples of non-rodent terrestrial TCI lineages have been reported, one from *Sus scrofa,* and two from *D. novemcinctus.*

TCII sub-lineages may be rarer in silvatic cycles than TCI. TCIIa does not apparently occur much further south than the Amazon region, and is found in arboreal and terrestrial species of both mammal and vector. TCIIc is found exclusively in terrestrial transmission cycles, most often with the *Cingulata,* but with sporadic infections in terrestrial species from two other orders, the *Carnivora* and the *Didelphimorpha.* Corresponding isolates have been identified from terrestrial *Panstrongylus* species. TCIIb, d, and e have rarely been sampled

from the silvatic environment, despite being abundant in domestic transmission cycles, and there are grounds to suspect that many may be epizootic in origin (e.g Lisboa *et al.*, 2004 (226)). All seem to cluster in the densely populated Southern Cone region, with TCIIb occurring in the Atlantic Forest in arboreal *Didelphimorpha* and *P. megistus*, in terrestial *Cingulata* in the Chaco, as well as once in *Tr. spinolai* in Chile. TCIId has been isolated, sporadically, from terrestrial mammals, including members of four orders, *Carnivora, Didelphimorpha, Rodentia* and *Cingulata*, and also from terrestrial *Triatoma* species nymphs in this study. TCIIe is represented by a single isolate in silvatic transmission cycles, a *D. marsupialis* in the Bolivian Yungas.

At this stage it is important to highlight a recent study of silvatic *T. cruzi* genotypes undertaken in Chile (372). Across a sample of 90 mammals taken from arid, rocky ecotopes in central Chile, including *Phylotis darwini, Arbothrix olivaceus, Octodon degus,* and *Thylamis elegans*, the authors identified a high frequency of mixed infections of genotypes TCI, TCIIb, TCIId, and TCIIe. There is considerable doubt concerning the validity of these data. Two major criticisms arise. Firstly the panel of reference strains employed entirely lacked any representatives of TCIIa or TCIIc. Secondly the authors used kDNA minicircle probes to detect parasite sequences within DNA extracted from whole mammal blood. Minicircle DNA is highly variable, and a recent analysis of 170 minicircle sequences from a representative panel of 19 *T. cruzi* strains demonstrated an extremely limited capacity to resolve major *T. cruzi* lineages (3). Pending resolution of these inconsistencies, these data have not been included in this study.

The overall pattern that emerges from *T. cruzi* genotype, host and vector associations is far from clear. It is hard to ascertain whether mammal or vector plays the dominant role in driving *T. cruzi* diversification. In general, it seems that trypanosomes tend not to cospeciate with their host or vector species (221). Even *T. rangeli*, within which genetic variation is closely associated with different *Rhodnius* clades, is unlikely to have cospeciated with these vectors. Incongruent parasite and vector phylogenies instead point to widespread host switching (339). Silvatic genetic variation within *T. brucei brucei* and *T. b. rhodesiense* may be largely independent of mammalian host and dipteran vector species, especially because genetic exchange is likely be acting to break up population substructure (235, 261). Inter-lineage recombination in *T. cruzi* is very rare, an observation that can be

explained, especially in this study, by the lack of contact between different lineages inhabiting distinct niches at sympatric sites (191). Some lineages do occur within the same silvatic niche, however, as is clear from the historical data. Some other mechanism, other than physical separation, may be preventing recombination. Nonetheless, the limited recombination that has thought to have occurred, between TCIIb and TCIIc, as well as possibly between TCIIb and TCI, does fit with the geographical and niche distribution of the genotypes involved(191, 199).

Pan-species trypanosome phylogenies, in combination with the substantial genetic diversity within *T. cruzi*, point to an ancient association with South America (Section 1.8). The extant diversity of *T. cruzi* may, in part, find its origins in the evolutionary history of its mammalian reservoirs on the continent and corresponding paeleoclimatological events. The rise of the western Andes 40 million years ago, followed by the eastern Andes, diverted the original north-south drainage pattern of South America to the current west-east drainage of the Amazon and Orinoco river systems, creating a vast area of moist tropical lowlands in north-eastern South America (152). The apparently simultaneous initial divergence of the modern Didelphimorpha and armadillo lineages appears to correlate with this event (Section 1.6.2). The timing of the divergence of the major triatomine bug tribes is less clear. However, their ancestors are likely to have been present in South America before its separation from Gondwanaland, and estimates place the divergence between the Rhodniini and the Triatomini at 48.9-64.4 MYA(373). The widespread occurrence of TCI from tropical and semi tropical biomes east of the Andes, especially in arboreal didelphids and associated vectors, suggests a possible correlation with the formation of the Amazon and Orinoco basins. TCIIc, and to a lesser extent TCIIa, could have radiated simultaneously or later in the same region, possibly as the result of fusion between TCI and TCIIb, to exploit a terrestrial niche, vector clique, and the ancestors of modern armadillos. The occurrence of TCI in terrestrial rodents is anomalous with respect to this conjecture, but could fit with a hypothesis of glacial refugia or even marine transgression (374). According to these two, sometimes competing, theories, the Amazon forest system has expanded and contracted over the last ~2 million years in response to dryer climatological conditions and/or through fragmentation due to marine flooding of low-lying areas. TCI in rodents from arid

terrestrial ecotopes could have originated from moist tropical ecotopes that have since receded and/or been cut off from Amazonia as a whole.

TCIIb is perhaps the most enigmatic of all *T. cruzi* lineages. As the most genetically divergent sublineage, one might expect it to be widespread in silvatic transmission cycles, which it clearly is not. This has lead to the erroneous assumption that it must have diverged before the separation of South America, entering in association with mammals during the GAI (Section 1.6.2). This theory may be refuted if any of the small number of isolates found in association with ancient mammalian lineages are truly silvatic. Additionally, if the theory of an early fusion event between TCI and TCIIb is correct, then TCIIb is highly likely be more than 5 million years old in South America (Section 1.8.2). Finally, the theory is based around the misidentification of a number of North American TCIIa isolates as TCIIb (27, 196).

4.4.3 Findings and Conclusions: a summary

Field study

- TCI is a common genotype widespread throughout Northern Bolivia and Western Venezuela, especially in didelphid marsupial mammals and *Rhodnius* species tratomines captured in ecotopes abundant with *Attalea* species palms. A further association between TCI and arid, terrestrial ecotopes was also identified in the Western Andes.

- TCIIc is not limited to a Southern Amazonian distribution but rather occurs sympatrically with TCI in association with *Panstrongylus* species triatomines and armadillos (especially *D. novemcinctus*) in both Northern Bolivia and Venezuela. Niche specificity is identified as a possible factor driving TCI and TCIIc diversification.

- *T. rangeli* was associated with tropical arboreal ecotopes in both Bolivia and Venezuela and was especially common in *Tam. tetradactyla*. All those analysed belonged to the KP1 (+) genotype, thought to be associated with *R. robustus* group triatomines.

171

- El Beni in Northern Bolivia cannot be considered an area hyper-endemic for domestic Chagas disease transmission, but limited infestation of domestic and peridomestic structures by four different triatomine species was identified. Three species were shown to harbour *T. cruzi* infection, including some strains not present in the silvatic cycle, indicative of some contact with humans.

Historical and Field silvatic genotypic data combined

- TCI is associated primarily with *Rhodnius*, arboreal mammals, especially didelphid marsupials and is found throughout the silvatic distribution if *T. cruzi*. A secondary, infrequently sampled but widespread (three dispersed sites) association also seems to exist with terrestrial rodents and possibly *Triatoma* species triatomines.
- TCIIc is exclusively associated with terrestrial mammals, especially the *Cingulata*, and has only ever been isolated from *Panstrongylus* species triatomines.
- TCIIa has eclectic affinities between different hosts and niches, but appears to be limited to Northern Amazonia.
- TIIb has been infrequently isolated from the silvatic environment, but does occur with ancient indigenous mammalian lineages in Southern South America, refuting a recent introduction via the Isthmus of Panama.

5   Results: Population genetic structure of *T. cruzi* evaluated using sequence and microsatellite data.

5.1   Inter-lineage results

5.1.1   Inter-lineage population structure

5.1.1.1   Samples analysed.

In collaboration with Michael Lewis (MLEW, LSHTM), a panel of 48 strains was assembled across the six major lineages of *T. cruzi* drawn from strains held at the LSHTM as well as some donated by other individuals and institutions (see Table 28). All strains were biologically cloned by MLEW using the technique described in Section 3.4.2. Genotypes described for the lineages in Table 28 were defined using standard typing methodologies (See Section 3.5) (375).

Table 28 Representative panel of strains across the six *T. cruzi* gentoptypes strains assembled for microsatellite analysis and glucose phosphate isomerase (*gpi*) sequence analysis.

Genotype	Strain	Origin	Host Species	Origin[a]
I	X10/1	Belem, Brazil	*Homo sapiens*	LSHTM
I	C8 CL1	La Paz, Bolivia	*Triatoma infestans*	LSHTM
I	P I (CJ007)	Carajas, Brazil	*Didelphis marsupialis*	HC
I	P II (CJ005)	Carajas, Brazil	Unidentified triatomine	HC
I	B187 CL10	Para state, Brazil	*Didelphis marsupialis*	LSHTM
I	CHILE C22	Flor de Valle, Chile	*Triatoma spinolai*	LSHTM
I	SAXP18 CL5	Majes, Peru	*Homo sapiens*	LSHTM
I	92101601P CL1	Georgia, USA	*Didelphis marsupialis*	IRD
I	JR CL4	Venezuela	*Homo sapiens*	HC
IIa	CAN III CL1	Belem, Brazil	*Homo sapiens*	LSHTM
IIa	10 R26	Santa Cruz, Bolivia	*Aotus sp*	IRD
IIa	92122102R	Georgia, USA	*Procyon lotor*	IRD
IIa	SAIMIRI 3 CL1	Venezuela	*Saimiri sciureus*	IRD
IIa	STC10R CL1	Georgia, USA	*Procyon lotor*	IRD
IIa	ERA CL2	Venezuela	Unknown	HC
IIa	X10610 CL5	Venezuela	*Homo sapiens*	HC
IIb	ESM CL3	Sao Felipe, Brazil	*Homo sapiens*	LSHTM
IIb	POT7A CL1	San Martin, Paraguay	*Triatoma infestans*	MY
IIb	POT7B CL5	San Martin, Paraguay	*Triatoma infestans*	MY
IIb	CHACO 23 COL4	Chaco, Paraguay	*Triatoma infestans*	MY
IIb	RITA CL5	Brazil	*Homo sapiens*	LSHTM
IIb	TU18 CL2	Tupiza, Bolivia	*Triatoma infestans*	IRD
IIb	CBB CL2	Tulahuen, Chile	*Homo sapiens*	IRD
IIb	IVV CL4	Cuncumen, Chile	*Homo sapiens*	LSHTM

Table 28 Continued from overleaf

IIc	M5631 CL5	Marajo, Brazil	*Dasypus novemcinctus*	LSHTM
IIc	M6421 CL6	Belem, Brazil	*Homo sapiens*	LSHTM
IIc	JA2 CL2.2	Amazonas, Brazil	Unknown	LSHTM
IIc	ARMA13 CL1	Campo Lorro, Paraguay	*Dasypus novemcinctus*	MY
IIc	ARMA18 CL3	Campo Lorro, Paraguay	*Dasypus novemcinctus*	MY
IIc	85/847 CL2	Alto Beni, Bolivia	*Dasypus novemcinctus*	IRD
IIc	CM25 CL2	Carimaga, Colombia	*Dasyprocta fugilinosa*	IRD
IIc	SABP19 CL1	Peru	*Triatoma infestans*	LSHTM
IId	SC43CL1	Santa Cruz, Bolivia	*Triatoma infestans*	LSHTM
IId	92.80 CL2	Santa Cruz, Bolivia	*Homo sapiens*	IRD
IId	PARA4 CL3	Paraguari, Paraguay	*Triatoma infestans*	MY
IId	PARA6 CL3	Paraguari, Paraguay	*Triatoma infestans*	MY
IId	CHACO2 CL3	Chaco, Paraguay	*Triatoma infestans*	MY
IId	VINCH101 CL1	Limari, Chile	*Triatoma infestans*	LSHTM
IId	BUG 2148 CL1	Rio Grande do Sul, Brazil	*Triatoma infestans*	LSHTM
IId	PAH179 CL5	Chaco, Argentina	*Homo sapiens*	IRD
IIe	CL BRENER	Rio Grande do Sul, Brazil	*Triatoma infestans*	LSHTM
IIe	CHACO9 COL15	Chaco, Paraguay	*Triatoma infestans*	MY
IIe	CHACO17 COL1	Chaco, Paraguay	*Triatoma infestans*	MY
IIe	TULA CL2	Tulahuen, Chile	*Homo sapiens*	IRD
IIe	P251 CL7	Cochabamba, Bolivia	*Homo sapiens*	LSHTM
IIe	EPV20-1 CL1	Chaco, Argentina	*Triatoma infestans*	PD
IIe	LHVA CL4	Chaco, Argentina	*Triatoma infestans*	PD
IIe	VFRA1 CL1	Francia, Chile	*Triatoma infestans*	PD

[a] IRD – Institut de Recherche pour le Développement, Montpellier, France, courtesy of M. Tibayrenc. LSHTM – held at LSHTM, London. HC – Hernan Carrasco, Universidad Central de Venezuela, Caracas. PD – Patricio Diosque, Universidad de Salta, Argentina. MY – Matthew Yeo, LSHTM.

5.1.1.2 Sequence analysis

All molecular cloning, sequencing and sequence analysis of the 48 strain panel was undertaken by MLEW, using the methodologies described in Section 3.6. After assembly and alignment, a 1038bp fragment of the glucose phosphate isomerase (*gpi*) gene was used to construct a maximum likelihood tree in DNAml, part of the PHYLIP v3.66 package (282). The best-fit (highest -ln likelihood) model of nucleotide substitution was tested using BaseMl, part of the Paml v4 package, and the General Reversible Model selected, allowing for rate heterogeneity among sites (283). Bootstraps were generated using SEQBOOT in PHYLIP v3.66, also as in Section 3.6.
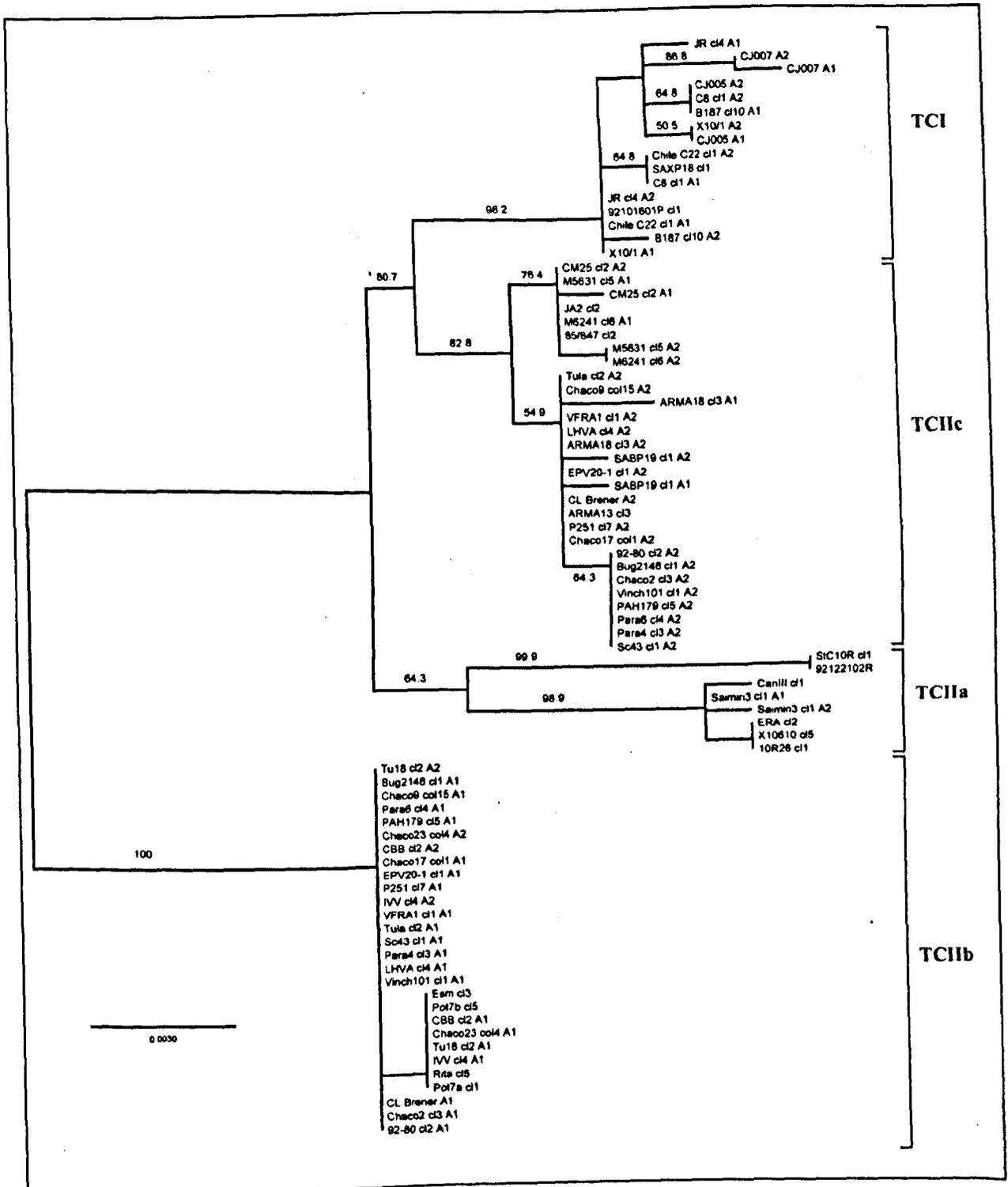
Figure 49 Un-rooted maximum likelihood tree showing the phylogenetic relationships between 48 isolates drawn from the six major *T. cruzi* genotypes. Bootstraps, generated using SEQBOOT, from the PHYLIP v3.66 package, ≥50% are shown (282). Reproduced with the permission of the author (MLEW) (375).

175

Apart from of TCIId and TCIIe, all major *T. cruzi* groups are resolved as monophyletic clades within the tree shown in Figure 49. TCIId and TCIIe haplotypes are spread equally across TCIIb and TCIIc lineages. All major clades receive strong bootstrap support (>80%), with the exception of TCIIa (64.3%). TCIIa is also apparently the most diverse clade, equating to 15 single nucleotide polymorphisms (SNPs) within the group (375). The next most diverse groups are TCIIc and TCI, where eight SNPs were identified and finally TCIIb, where only a single SNP was present among all isolates. No variation was detected within either TCIId or TCIIe, although the two populations are distinguishable by a single SNP. For a detailed discussion of within clade clustering see Lewis (2008) and Section 5.3 (375).

5.1.2 Microsatellite analysis
5.1.2.1.1 Microstatellite loci employed

Ten microsatellite markers were selected by the author for analysis of the 48 strain panel, and the laboratory-based typing effort shared between the author and MLEW. The selection criteria were based upon the robustness of fragment amplification across a representative subset of the panel. Monomorphic microsatellite loci were excluded. Among those selected, one originated from an earlier study, while the remaining nine were identified using the data-mining technique described in Section 3.7.1 (215). Reaction quantities, conditions and allelic size scoring techniques were as described in Section 3.7.

Table 29 Ten microsatellite markers selected for analysis of reference strain panel.

Primer code	Contig ID[a]	Position[b]	Repeat[c]	Forward/Reverse primer
famp1a	7098	64001-64129	$(TG)_n$	GAGCAGATCTTCCTTGTGCC
				TGGTGAAATGCACGCATC
famp2a	7118	76285-76410	$(TC)_n$	GACATGTATGCTTGAAACCTCC
				TCCATCTCCCTTCACACTCC
famp6a	7143	119499-119520	$(TG)_n$	TCGTTCTCTTTACGCTTGCA
				TAGCAGCACCAAACAAAACG
vicp2a	7143	88658-88832	$(CT)_n$	CATCAAGGAAAAACGGAGGA
				CGGTACCACCTCAAGGAAAG
nedp5a	7206	69979-70016	$(TC)_n$	CCAACATTCAACAAGGGAAA
				GCATGAATATTGCCGGATCT
vicp4a	7206	25751-25779	$(TCC)_n$	AGACGTTCATATTCGCAGCC
				AGCCACATCCACATTTCCTC
tetp3a	8305	49478-49495	$(TC)_n$	CGTACGACGTGGACACAAAC
				ACAAGTGGGTGAGCCAAAAG
vicp1b	8305	46598-46638	$(TA)_n$	AACCCGCGCAGATACATTAG
				TTCATTTGCAGCAACACACA
famp8a	8646	88680-88708	$(TCG)_n$	ACCACCAGGAGGACATGAAG
				TGTACACGGAACAGCGAAG
famp1b*	Unknown	Unknown	$(CA)_n A(C)_n$	GCGTAGCGATTCATTTCC
				ATCCGCTACCACTATCCAC

[a] Refers to a syntenous fragment of sequence identified by the *T. cruzi* genome project (376). Colours represent primers on the same contiguous fragment (contig). Uncoloured contig codes indicate a unique contig. Microsatellite loci on unidentified contigs are marked 'unknown'.

[b] Refers to the position of the of the primer binding site, flanking region and repeat motif 5'-3' along the sense strand. Unidentified positions are marked 'unknown'.

[c] Refers to the repeat type (e.g. di-nucleotide, tri-nucleotide). Unidentified repeat types are marked 'unknown'.

*Alternative code MLcf10 (215).

5.1.2.1.2 Distance measures at the individual level.

Genetic distances were assessed at an individual level using two different measures, one under an infinite allele model – D_{AS}; and one under a stepwise model of mutation, $\delta\mu^2$ (See Section 1.9.4). The resulting matrices are not shown here. Instead a single tree based upon values for D_{AS} and constructed in PHYLIP v3.66 and is displayed in Figure 50. Bootstrap values are given on major branches for both D_{AS} and $\delta\mu^2$, where branching topology was consistent between the two datasets.
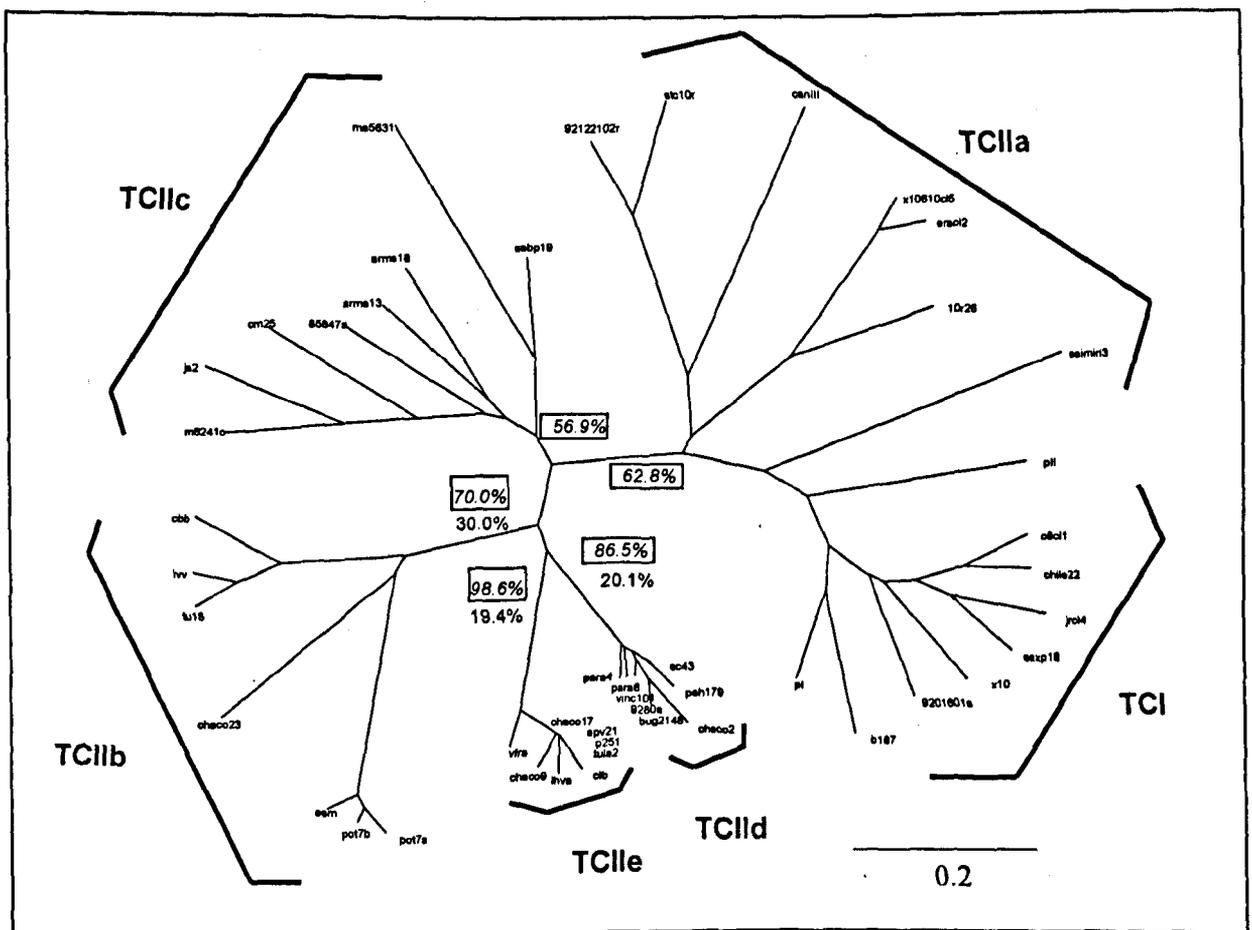
Figure 50 Un-rooted neighbour-joining (NJ) tree constructed using D_{AS} values between 48 *T. cruzi* strains across the six major *T. cruzi* lineages. D_{AS} matrix was generated using MICROSAT v1.2 and tree assembled in the PHYLIP package v3.66 (282, 290). Bootstrap values over 1000 trees for D_{AS} (>50% - italics in boxes) and $\delta\mu^2$ (for comparison), calculated in MICROSAT v1.2 and analyzed under extended rule majority consensus, also in the PHYLIP package v3.66, are displayed for major branches.

178

Bootstrap values varied considerably between the two distance measures employed. Only major clades that received ≥70% bootstrap support in the D_{AS} tree occurred as monophyletic groups in that based on $\delta\mu^2$. Even in these cases, $\delta\mu^2$ bootstraps were far from robust (≤30%). Interestingly, no D_{AS} bootstrap support >50% was found for the division of TCIIa and TCI.

There are major differences between the tree topology based on microsatellite-derived D_{AS} distances compared with that constructed from ML sequence analysis by MLEW (375). While TCIIc, TCIIb and, to a lesser extent, TCIIa and TCI, are supported as monophyletic clades, the branching order is distinct. Sequence analysis defines TCI and TCIIc as sister groups, while microsatellite analysis defines TCIIa as the sister group of TCI. Furthermore, because it is not possible to accurately define multilocus microsatellite haplotypes for loci of unknown multi-allelic linakge empirically or through simulation, TCIId and TCIIe do not cluster among TCIIb and TCIIc isolates, as in Figure 49, but form monophyletic clades, albeit between TCIIb and TCIIc (281).

5.1.2.1.3 Pairwise population gene flow

Population comparisons were made using two statistics: F_{ST} and R_{ST}. F_{ST} provides a measure of the level of subdivision within a population by comparing the average expected heterozgosity across loci within subpopulations with the average expected heterozygosity across the total population (Section 3.7.6.1.5). Pairwise comparisons can be made by sequentially calculating F_{ST} for every possible grouping of two populations in the dataset. R_{ST} was developed specifically for microsatellites to incorporate a stepwise mutational model into a F_{ST} – like statistic (Section 3.7.6.1.5) (302).

Table 30 Pairwise matrix of F_{ST} values between six *T. cruzi* lineages based on 10 microsatellite loci. Italics indicate p-values generated from 1000 random permutations leading to a value larger than or equal to that observed. Constructed in ARLEQUIN v3.1 (295).

Population	TCI	TCIIa	TCIIb	TCIIc	TCIId	TCIIe
TCI	*	*0.00000*	*0.00000*	*0.00000*	*0.00010*	*0.00000*
TCIIa	0.35368	*	*0.00020*	*0.00020*	*0.00010*	*0.00020*
TCIIb	0.52053	0.38901	*	*0.00000*	*0.00020*	*0.00010*
TCIIc	0.4994	0.37486	0.46691	*	*0.00020*	*0.00010*
TCIId	0.53821	0.4346	0.31332	0.30269	*	*0.00000*
TCIIe	0.57332	0.46121	0.40435	0.41932	0.29647	*

Significant inter-population subdivision was detected after Bonferroni correction between all populations, indicating either limited gene flow ($0.2 < F_{ST} > 0.33$) or, in effect, complete isolation ($F_{ST} > 0.33$), as predicted by Porter's (1990) general guidelines (see Section 3.7.6.1.5) (301). An un-weighted pair group method with arithmetic mean (UPGMA) tree was constructed using the F_{ST} matrix is shown in Figure 51. Lowest gene flow was apparent between TCIId&e, which, in turn, were more closely related to TCIIb and TCIIc than they were to TCIIa and TCI, which form a genetically isolated, monophyletic clade.

```
                                        ─── TCIId
                                        ─── TCIIe
                                        ─── TCIIb
                                        ─── TCIIc
                                        ─── TCI
                                        ─── TCIIa

        0.20      0.15      0.10      0.05      0.00
```
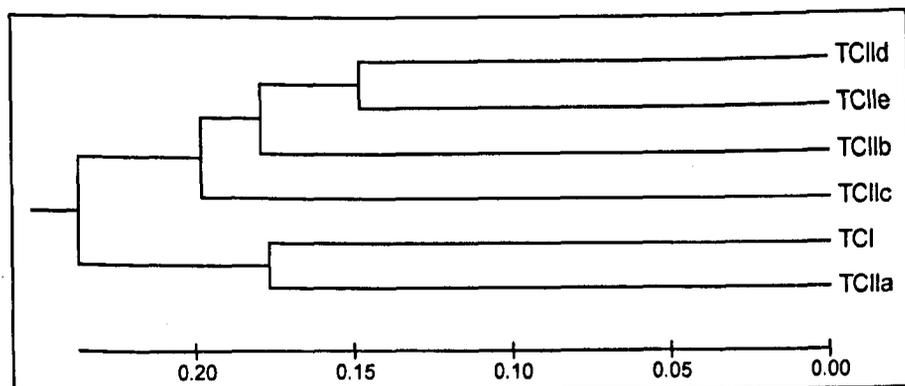
Figure 51 Unrooted UPGMA tree of pairwise F_{ST} values between six major *T. cruzi* lineages constructed in MEGA v4.0 (377).

A matrix of pairwise R_{ST} values is displayed in Table 28. Considerable gene flow was again indicated between TCIId and TCIIe (0.202), as well as between TCIIb and TCIIe (0.137). Concordant with estimates of F_{ST}, the R_{ST} estimate between TCIIb and TCI was among the highest observed (0.538). In contrast to F_{ST} estimates, however, among other anomalies, a degree of gene flow was detected TCI - TCIId and TCI - TCIIe (0.293 and 0.239 respectively). The correlation between R_{ST} and F_{ST} was examined in the form of a Mantel's test, as in Section 3.7.6.1.3, and no significant relationship observed (p>>0.05).

Table 31 Pairwise matrix of Slatkin's R$_{ST}$ values between six *T. cruzi* lineages based on 10 microsatellite loci. Values in italics indicate the standard error of the mean over 1000 bootstraps. Constructed in MICROSAT v1.2 (290).

Population	TCI	TCIIa	TCIIb	TCIIc	TCIId	TCIIe
TCI	*	*0.06*	*0.13*	*0.156*	*0.115*	*0.133*
TCIIa	0.431	*	*0.128*	*0.148*	*0.078*	*0.108*
TCIIb	0.538	0.437	*	*0.186*	*0.064*	*0.056*
TCIIc	0.284	0.336	0.459	*	*0.089*	*0.151*
TCIId	0.293	0.341	0.323	0.185	*	*0.093*
TCIIe	0.239	0.276	0.137	0.294	0.202	*

5.2 Intra-lineage results

5.2.1 *Gpi* sequence data

5.2.1.1 Samples analysed

A ~1100bp fragment of the *gpi* gene was sequenced across 45 *T. cruzi* strains, including 24 from TCI and 21 from TCIIc according to the methodology detailed in Section 3.6. Additional sequences from a further five TCI and TCIIc isolates, as well as two from TCIIb (to act as an out-group) were donated by MLEW. Strains were chosen to act as a representative subset of those subjected to microsatellite analysis (Sections 5.2.2.1 and 5.2.2.2), and to span a variety of major geographic groups, different mammalian hosts and triatomine vectors (Table 32).

Table 32 52 *T. cruzi* strains drawn from lineages TCI, TCIIc and TCIIb for sequencing of the glucose phosphate isomerase (*gpi*) gene region.

Lineage	Sample code	Tree code[b]	Host/Vector	Date	Region, Country	Latitude[c]	Longitude[c]	Source[b]
TCI	458	PF_COL1	*Potus flavus*	1987	Bajo Calima, Colombia	3.983	-76.96	IRD
TCI	93070103P	DM_NA1	*Didelphis marsupialis*	1993	Georgia, USA	32.43	-83.31	IRD
TCI	92101601P	DM_NA2	*Didelphis marsupialis*	1992	Georgia, USA	32.43	-83.31	IRD
TCI	361TA	DM_COL1	*Didelphis marsupialis*	1986	Inguapi del Guadual, Colombia	1.71	-77.91	LSHTM
TCI	CHILE C22*	TS_CH1	*Triatoma spinolai*	c.1985	Flor de Valle, Chile	-30.31	-71.22	LSHTM
TCI	COTMA55	POC_COT1	*Phyllotis ocilae*	14.10.04	Cochambamba, Bolivia	-17.438	-66.279	TS
TCI	COTMA9	POC_COT2	*Phyllotis ocilae*	10.10.04	Cochambamba, Bolivia	-17.438	-66.279	TS
TCI	DAVIS 9.90	TD_HO1	*Triatoma dimidiata*	1983	Tegucigalpa, Honduras	14.083	-87.2	IRD
TCI	JR*	HS_V1	*Homo sapiens*	Unknown	Anzoategui, Venezuela	9.01	-64.34	HC
TCI	M13	DM_V1	*Didelphis marsupialis*	12.6.04	Barinas, Venezuela	7.5	-71.23	TS
TCI	M18	DM_V2	*Didelphis marsupialis*	13.6.04	Barinas, Venezuela	7.5	-71.23	TS
TCI	M7	DM_V3	*Didelphis marsupialis*	14.5.04	Barinas, Venezuela	7.5	-71.23	TS
TCI	P268	HS_COT1	*Human*	Unknown	Cochabamba	-17.381	-66.167	IRD
TCI	PALDA22	DA_AR1	*Didelphis albiventris*	23.3.2001	Chaco, Argentina	-27.133	-61.46	PD

181

Table 32 Continued from overleaf.

TCI	PALDA3	DA_AR2	*Didelphis albiventris*	23.3.2001	Chaco, Argentina	-27.133	-61.46	PD
TCI	PALDA5	DA_AR3	*Didelphis albiventris*	23.3.2001	Chaco, Argentina	-27.133	-61.46	PD
TCI	SJM32	PO_BO1	*Philander opossum*	7.9.04	Beni, Bolivia	-14.81	-64.6	TS
TCI	SJM34	DM_BO1	*Didelphis marsupialis*	7.9.04	Beni, Bolivia	-14.81	-64.6	TS
TCI	TEDA	DA_AR4	*Didelphis albiventris*	23.3.2001	Chaco, Argentina	-26.933	-61.583	PD
TCI	XE1313	PO_BR1	*Philander opossum*	23.11.83	Carajas, Brazil	-5.983	-51.333	EC
TCI	XE2913	PO_BR2	*Philander opossum*	05.05.88	Para, Brazil	-1.36	-48.366	EC
TCI	XE3776	DM_BR1	*Didelphis marsupialis*	23.10.98	Para, Brazil	-1.51	-49.21	EC
TCI	XE3981	DM_BR2	*Didelphis marsupialis*	09.01.95	Para, Brazil	-0.15	-50.383	EC
TCI	XE5012	DM_BR3	*Didelphis marsupialis*	13.01.99	Braganca, Brazil	-1.05	-46.76	EC
TCI	XE5495	PO_BR3	*Philander opossum*	21.02.01	Bagre, Brazil	-1.866	-52.2	EC
TCI	SJMC7	SS_BO1	*Scuireus spadiceus*	11.9.04	Beni, Bolivia	-14.81	-64.6	TS
TCIIc	M5	DN_V1	*Dasypus novemcinctus*	9.5.04	Barinas, Venezuela	8.483	-70.733	TS
TCIIc	85/847^	DN_BO1	*Dasypus novemcinctus*	Unknown	Alto Beni, Bolivia	-15.5	-67.5	IRD
TCIIc	ARMA12	DN_PA1	*Dasypus novemcinctus*	2001	Campo Lorro, Paraguay	-22.33	-58.93	MY
TCIIc	ARMA13	DN_PA2	*Dasypus novemcinctus*	2001	Campo Lorro, Paraguay	-22.33	-58.93	MY
TCIIc	ARMA18	DN_PA3	*Dasypus novemcinctus*	2001	Campo Lorro, Paraguay	-22.33	-58.93	MY
TCIIc	ARMA25	DN_PA4	*Dasypus novemcinctus*	2001	San Pedro, Paraguay	-24	-57	MY
TCIIc	ARMA9	DN_PA5	*Dasypus novemcinctus*	2001	Campo Lorro, Paraguay	-22.33	-58.93	MY
TCIIc	CAYMA14	DN_BO2	*Dasypus novemcinctus*	14.01.05	Santa Cruz, Bolivia	-17.5	-61.5	TS
TCIIc	CAYMA18	DN_BO3	*Dasypus novemcinctus*	17.01.05	Santa Cruz, Bolivia	-17.5	-61.5	TS
TCIIc	CAYMA19	DN_BO4	*Dasypus novemcinctus*	17.01.05	Santa Cruz, Bolivia	-17.5	-61.5	TS
TCIIc	CAYMA25	ES_BO1	*Euphractus sexcinctus*	17.01.05	Santa Cruz, Bolivia	-17.5	-61.5	TS
TCIIc	CM25^	DF_COL1	*Dasyprocta fugilinosa*	1982	Carimaga, Colombia	3.3	-73	IRD
TCIIc	M10	DN_V2	*Dasypus novemcinctus*	17.5.04	Barinas, Venezuela	8.483	-70.733	TS
TCIIc	M5631^	DN_BR1	*Dasypus novemcinctus*	c.1981	Marajo, Brazil	-1	-49.5	LSHTM
TCIIc	MA194	CV_PA1	*Chaetophractus vellerosus*	27.01.05	Campo Lorro, Paraguay	-22.33	-58.93	TS
TCIIc	MA25	DN_PA9	*Dasypus novemcinctus*	c.2003	Campo Lorro, Paraguay	-22.33	-58.93	MY
TCIIc	SAM6	DN_BO5	*Dasypus novemcinctus*	21.11.04	Santa Cruz, Bolivia	-20.016	-63.015	TS
TCIIc	SJMC10	DN_BO6	*Dasypus novemcinctus*	13.9.04	Beni, Bolivia	-14.81	-64.6	TS
TCIIc	SJMC19	DN_BO7	*Dasypus novemcinctus*	9.9.04	Beni, Bolivia	-14.81	-64.6	TS
TCIIc	SJMC4	DN_BO8	*Dasypus novemcinctus*	9.9.04	Beni, Bolivia	-14.81	-64.6	TS
TCIIc	SP13	DN_PA6	*Dasypus novemcinctus*	c.2003	nr San Pedro, Paraguay	-24	-57	MY
TCIIc	SP15	DN_PA7	*Dasypus novemcinctus*	c.2003	nr San Pedro, Paraguay	-24	-57	MY
TCIIc	SP16	DN_PA8	*Dasypus novemcinctus*	c.2003	nr San Pedro, Paraguay	-24	-57	MY
TCIIc	SP4	MD_PA1	*Monodelphis domestica*	c.2003	nr San Pedro, Paraguay	-24	-57	MY
TCIIb	CBB^	CBB	*Homo sapiens*	Unknown	São Felipe, Brazil	-14.392	-41.352	LSHTM
TCIIb	ESM^	ESM	*Homo sapiens*	Unknown	Tulahuen, Chile	-31.052	-70.795	LSHTM

[a] Sequenced by MLEW.
[b] See Figure 53.
[c] Decimal degrees.
[d] **IRD** – Institut de Recherche pour le Developpement, Montpellier, France, courtesy of M. Tibayrenc. **LSHTM** – held at LSHTM, London. **TS** – Isolated during this study. **HC** – Hernan Carrasco, Universidad Central de Venezuela, Caracas. **PD** – Patricio Diosque, Universidad de Salta, Argentina. **MY** – Matthew Yeo, LSHTM. **EC** - Aldo Valente, Institute Evando Chagas, Belem, Brazil.

5.2.1.2 *Gpi* sequence diversity and SNPs by population.

After assembly, a 1004bp fragment of the *gpi* gene region was aligned across all 52 strains (See Section 3.6.3). Of those 1004bp, 976 sites were invariable across TCIIc and TCI, equating to a total of 28 variable sites, including those within and between the groups (~2.7% sequence diversity). A full table of variable sites within and between TCI and TCIIc is displayed in Table 33. Transition: transversion ratio was approximately 6.25:1. Within subgroup diversity accounted for ~1.4% (TCI - 14 variable sites) and ~0.9% (TCIIc - nine variable sites) of total sequence diversity. Fixed sequence polymorphism between the two groups (five variable sites) accounted for the remaining ~0.5%. Multiple SNPs at a single site were only detected once, in TCI.

Table 33 Within and between subgroup (TCI&TCIIc) SNPs in a 1004bp fragment of the *gpi* gene region. SNPs in bold indicate transversions.

Position	101	128	150	200	212	225	285	308	387	398	440	468	488	494
TCI	C/T	-	-	-	C/T	-	-	C/T	A/C	-	-	C/T	C/T	-
TCI/TCIIc	-	G/A	-	-	-	-	-	-	-	-	T/C	-	-	-
TCIIc	-	-	G/A	G/A	-	C/T	G/A	-	-	C/T	-	-	-	G/A
Position	563	629	653	665	679	752	769	779	785	812	896	948	963	1004
TCI	-	G/A T/G	C/T	C/T	-	-	C/G	-	G/T	G/A	C/T	-	-	C/T
TCI/TCIIc	C/T	-	-	-	G/A	-	-	-	-	-	-	C/T	-	-
TCIIc	-	-	-	-	-	C/A	-	G/A	-	-	-	-	C/T	-

5.2.1.3 Heterozygosity and haplotype diversity after molecular cloning and phase simulation.

Of the 50 TCI and TCIIc strains analysed, 24 (48%) showed ambiguities ('split peaks' – Figure 52) between base calls at one site or more. Of these, seven were found in TCIIc, and 17 in TCI.
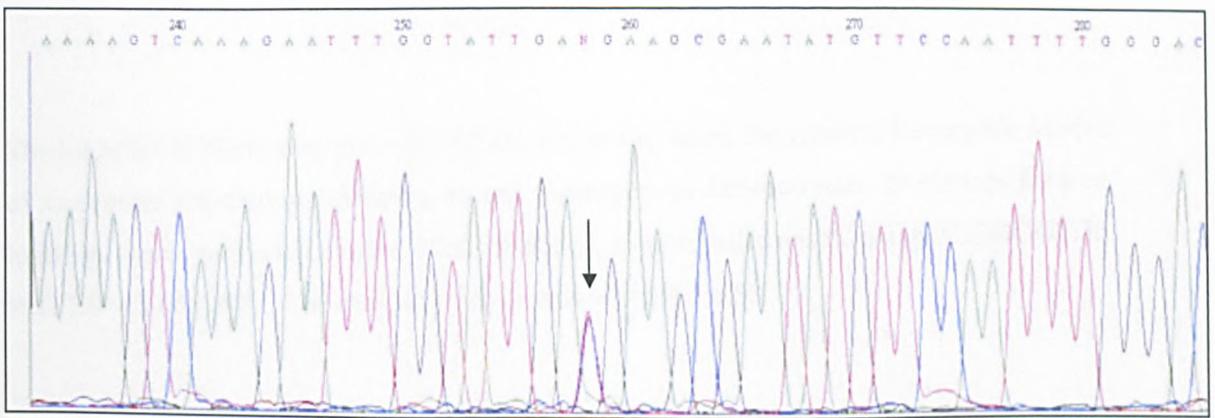
Figure 52 An ambiguous base call, or 'split peak' in a sequence trace derived from the *gpi* gene indicating a heterozygote **(C/T)**. The variable site is marked with an arrow. The sequence trace is derived from COTMA55 (TCI)

Haplotypes for sequences heterozygous at only one site (N=14) where defined by default. The remaining ten sequences that possessed two or more heterozygous sites were cloned into a plasmid vector, as described in Section 3.6.2. The entire un-phased dataset was also subjected to *in silico* phase simulation (see Section 3.6.4.1). Direct cloning identified 31 haplotypes within the dataset, not including those samples provided by MLEW, once the possibility of recombination through PCR strand-switching has been discounted (See Section 3.6.2). Phase simulation was undertaken *in silico* using PHASE v2.1 as in Section 3.6.4.1 (281). Five separate runs were undertaken, consisting of 1000 iterations, with a different random number seed per run. A high degree of consistency was observed between runs, suggesting that the appropriate parameters (number of iterations, length of burn-in) had been set. Nonetheless, PHASE recovered only 93.5% (29/31) of haplotypes to a probability of over 0.5. Similarly, simulation failed to correctly indentify all haplotype pairs (90% recovery, 45/50) to a probability of over 0.5, which accounted for five out of the ten isolates heterozygous at two or more sites.

As discussed in Section 3.6.4.1, PHASE, may fail to recover rare or unique haplotypes/haplotype pairs in the population, a feature common to all those misidentified. In light of these inconsistencies, only empirically derived haplotypes were used in later phylogenetic analyses.

5.2.1.4   Phylogenetic analysis

Maximum likelihood phylogenetic analysis of TCI and TCIIc empirically derived haplotype sequences was undertaken as in Section 3.6.4.2, including two TCIIb isolates as an outgroup. The overall strategy employed was similar to that described in Section 5.1.1.2.

The highest –ln likelihood tree (-1678.243) was found under the General Reversible Model of nucleotide substitution, allowing for rate heterogeneity between sites. Bootstraps for tree topology were generated from a 1000 randomly re-sampled datasets using SEQBOOT in PHYLIP v3.66 (282). The resulting tree is shown in Figure 53.
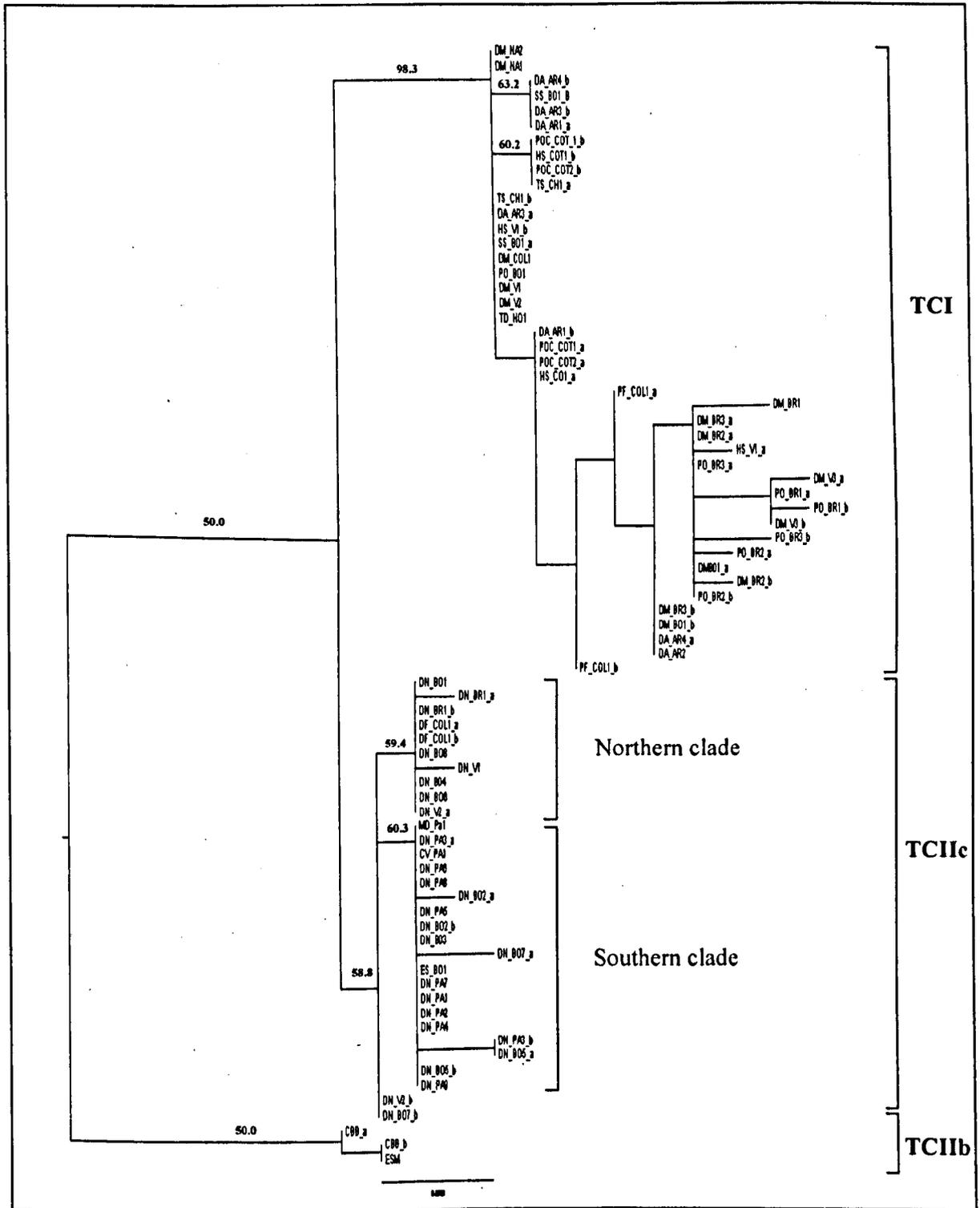
Figure 53 Un-rooted maximum likelihood tree showing the phylogenetic relationships among 50 representative TCI and TCIIc isolates with two TCIIb isolates as an outgroup. Bootstraps, generated using SEQBOOT from the PHYLIP v3.66 package, ≥50% are shown(282). Sample codes correspond to those in Table 32.

186

Both TCI and TCIIc are resolved as monophyletic clades in Figure 53, with TCIIc apparently more closely related to TCIIb than TCI, as is also the case in Figure 49. The divergent status of TCI with respect to the other two groups is strongly supported by bootstrap values (98.3%). TCIIc as a monophyletic clade receives only weak bootstrap support (58.8%), when members of the other major *T. cruzi* lineages are not included (see Figure 49). While the analysis successfully defines inter-genotype subdivision, the within-subgroup branching topology does not strongly correspond to patterns of host, vector or geography, and few internal groups receive >50% bootstrap support. Minor spatial genetic variation is apparent within TCIIc, where a general division, with a number of exceptions, can be made between isolates originating from Northern of Southern sample sites. This observation is approximately mirrored in Figure 49, where a division between Paraguayan TCIIc and other TCIIc isolates from further north is also evident.

5.2.1.4.1 Breakpoint analysis

All 50 TCI and TCIIc sequences were screened for possible mosaics using the RDP package (see Section 3.6.4.3) (284). The three tests implemented, MaximumChi (285), Bootscan (287), and Chimaera (286), all gave negative results, indicating a lack of recombination within the 1004bp sequence analysed.

5.2.2 Microsatellite analysis
5.2.2.1 TCI sub-lineage
5.2.2.1.1 Samples analysed

An expanded panel of 135 *T. cruzi* TCI genotype strains was assembled for microsatellite analysis (Table 34). Isolates were derived from 11 different mammal species, including humans, and seven different species of triatomine. The geographic distribution of these strains is shown in Figure 54. Samples were chosen to represent as fully as possible the entire geographic range of TCI.

Table 34 Panel of *T. cruzi* TCI genotype isolates assembled for microsatellite analysis.

Host/Vector	Original Code[a]	Tree code[b]	Date	Locality	Latitude[c]	Longitude[c]	Population 1[d] /2[e]	Source[f]
Homo sapiens	7570[g] (67/M)	H1	Unknown	Tachira, Venezuela	7.76	-72.25	VD/VD	HC
Homo sapiens	8104[g] (51/F)	H2	Unknown	Miranda, Venezuela	10.23	-66.66	VD/VD	HC
Homo sapiens	9010[g] (56/M)	H3	Unknown	Trujillo, Venezuela	9.55	-70.51	VD/VS	HC
Didelphis marsupialis	92090802P	DM1	1992	Georgia. USA	32.43	-83.31	AM/AM	IRD
Didelphis marsupialis	93070103P	DM2	1993	Georgia. USA	32.43	-83.31	AM/AM	IRD
Homo sapiens	9354[g] (70/M)	H4	Unknown	Sucre, Venezuela	10.46	-63.61	VD/VD	HC
Homo sapiens	10775[g] (65/F)	H5	Unknown	Barinas, Venezuela	8.37	-70.51	VD/VD	HC
Homo sapiens	10801[g] (65/M)	H6	Unknown	Guarico, Venezuela	8.91	-65.38	VD/VD	HC
Homo sapiens	11006[g] (61/F)	H7	Unknown	Cojedes, Venezuela	9.66	-68.20	VD/VD	HC
Homo sapiens	11042[g] (55/F)	H8	Unknown	Portuguesa, Venezuela	8.66	-69.5	VD/VD	HC
Homo sapiens	11124[g] (57/F)	H9	Unknown	Dtto Fed. Venezuela	10.61	-67.04	VD/VD	HC
Homo sapiens	11398	H10	Unknown	Merida, Venezuela	8.59	-71.23	VD/VD	HC
Homo sapiens	11541[g] (58/M)	H11	Unknown	Merida, Venezuela	8.59	-71.23	VD/VD	HC
Homo sapiens	11713[g] (53/M)	H12	Unknown	Lara, Venezuela	10.04	-69.32	VD/VD	HC
Homo sapiens	11804[g] (49/M)	H13	Unknown	Portuguesa, Venezuela	9.01	-69.29	VD/VD	HC
Homo sapiens	11838	H14	Unknown	Guarico, Venezuela	8.71	-66.62	VD/VD	HC
Homo sapiens	11881[g] (51/M)	H15	Unknown	Anzoat, Venezuela	9.36	-65.12	VD/VD	HC
Didelphis marsupialis	92101601P	DM3	1992	Georgia, USA	32.43	-83.31	AM/AM	IRD
Didelphis marsupialis	361TA	DM4	1986	Guadual, Colombia	1.71	-77.91	COL(VS)/COL(VS)	IRD
Potus flavus	458	PF	1987	Bajo Cal. Colombia	3.98	-76.96	COL(VS)/COL(NEB)	IRD
Didelphis marsupialis	AM3	DM5	Unknown	Sucre, Venezuela	10.11	-64.55	VS/VS	HC
Didelphis marsupialis	AM6	DM6	Unknown	Sucre, Venezuela	10.11	-64.55	VS/VS	HC
Rhodnius prolixus	AMRPA34	RP1	Unknown	Anzoat. Venezuela	9.01	-64.34	VS/VS	HC
Philander opossum	B1947	PO1	28.01.92	Barcarena, Brazil	-1.5	-48.18	NEB/NEB	EC
Philander opossum	B2026	PO2	09.05.89	Camatá, Brazil	-2.61	-49.53	NEB/NEB	EC
Philander opossum	B2077	PO3	13.12.90	Belem, Brazil	-1.36	-48.36	NEB/NEB	EC
Didelphis marsupialis	B2310	DM7	9.12.94	Afua, Brazil	-0.15	-50.38	NEB/NEB	EC
Didelphis marsupialis	B2311	DM8	9.12.94	Afua, Brazil	-0.15	-50.38	NEB/NEB	EC
Didelphis marsupialis	XE5740	DM9	18.6.02	Para, Brazil	-1.38	-48.86	NEB/NEB	EC
Triatoma maculata	BARM*	TM1	Unknown	Barinas, Venezuela	8.37	-70.51	VD/VS	HC
Rhodnius prolixus	CALC104	RP2	Unknown	Carabobo, Venezuela	10.19	-68.00	VS/VS	HC
Rhodnius prolixus	CASABONIF*	RP3	Unknown	Portuguesa, Venezuela	9.01	-69.29	VD/VD	HC
Didelphis marsupialis	CD45	DM10	Unknown	Dtto. Fed. Venezuela	10.35	-67.03	VS/VS	LSHTM
Mepraia spinolai	CHILEC22	TSP1	Unknown	Flor de Valle, Chile	-30.31	-71.22	CH(ARG)/CH(COT)	LSHTM
Mepraia spinolai	CHILEWALL	TSP2	Unknown	Flor de Valle, Chile	-30.31	-71.22	CH(ARG)/CH(COT)	LSHTM
Didelphis marsupialis	CJO20	DM11	Unknown	Carajas, Brazil	-5.98	-51.33	NEB/NEB	HC
Didelphis marsupialis	CJO33	DM12	Unknown	Carajas, Brazil	-5.98	-51.33	NEB/NEB	HC
Rattus rattus	C022	RR1	Unknown	Dtto Fed. Venezuela	10.54	-67.80	VS/VS	HC
Didelphis marsupialis	CO57	DM13	Unknown	Dtto Fed. Venezuela	10.54	-67.80	VS/VS	HC
Didelphis marsupialis	CO75	DM14	Unknown	Dtto Fed. Venezuela	10.54	-67.80	VS/VS	HC
Rattus rattus	CO84	RR2	Unknown	Miranda, Venezuela	10.31	-66.39	VS/VS	HC
Akodon boliviensis	COTMA22	AB1	11.10.04	Cotopachi, Bolivia	-17.43	-66.27	COT/COT	TS
Akodon boliviensis	COTMA38	AB2	13.10.04	Cotopachi, Bolivia	-17.43	-66.27	COT/COT	TS
Phyllotis ocilae	COTMA47	POC1	13.10.04	Cotopachi, Bolivia	-17.43	-66.27	COT/COT	TS
Phyllotis ocilae	COTMA55	POC2	14.10.04	Cotopachi, Bolivia	-17.43	-66.27	COT/COT	TS
Phyllotis ocilae	COTMA9	POC3	10.10.04	Cotopachi, Bolivia	-17.43	-66.27	COT/COT	TS
Triatoma dimidiata	DAVIS 9.90	TD	1983	Tegucigal. Honduras	14.08	-87.20	AM/AM	IRD
Didelphis marsupialis	DM1	DM15	Unknown	Trujillo, Venezuela	9.55	-70.51	VS/VS	HC
Didelphis marsupialis	DM4	DM16	Unknown	Trujillo, Venezuela	9.55	-70.51	VS/VS	HC

Table 34 Continued from overleaf.

Didelphis marsupialis	DMSU8	DM17	Unknown	Sucre, Venezuela	10.46	-63.61	VS/VS	HC
Didelphis marsupialis	DMSUC	DM18	Unknown	Sucre, Venezuela	10.46	-63.61	VS/VS	HC
Triatoma sanguisuga	FLORIDAC1D12	TS	Unknown	Florida, USA	30.50	-81.66	AM/AM	IRD
Didelphis marsupialis	IM4810	DM19	23.4.02	Manaus, Brazil	-3.07	-60.16	NEB/NEB	EC
Homo sapiens	**JR**	H16	Unknown	Anzoat. Venezuela	9.01	-64.34	VD/VS	HC
Rhodnius prolixus	LL2LA	RP4	15.5.04	Barinas, Venezuela	7.50	-71.23	VS/VS	TS
Didelphis marsupialis	M12	DM20	10.6.04	Barinas, Venezuela	7.50	-71.23	VS/VS	TS
Didelphis marsupialis	**M13**	DM21	12.6.04	Barinas, Venezuela	7.50	-71.23	VS/VS	TS
Didelphis marsupialis	M15	DM22	13.6.04	Barinas, Venezuela	7.50	-71.23	VS/VS	TS
Didelphis marsupialis	**M16**	DM23	13.6.04	Barinas, Venezuela	7.50	-71.23	VS/VS	TS
Didelphis marsupialis	**M18**	DM24	13.6.04	Barinas, Venezuela	7.50	-71.23	VS/VS	TS
Didelphis marsupialis	**M7**	DM25	14.5.04	Barinas, Venezuela	7.50	-71.23	VS/VS	TS
Homo sapiens	P234	H17	Unknown	Cochabamba, Bolivia	-17.38	-66.16	COT/COT	IRD
Homo sapiens	P238	H18	Unknown	Cochabamba, Bolivia	-17.38	-66.16	COT/COT	IRD
Homo sapiens	P234	H19	Unknown	Cochabamba, Bolivia	-17.38	-66.16	COT/COT	IRD
Didelphis albiventris	**PALDA1**	DA1	23.3.2001	Chaco, Argentina	-27.13	-61.46	AR/AR	PD
Didelphis albiventris	**PALDA20**	DA2	23.3.2001	Chaco, Argentina	-27.13	-61.46	AR/AR	PD
Didelphis albiventris	PALDA21	DA3	23.3.2001	Chaco, Argentina	-27.133	-61.46	AR/AR	PD
Didelphis albiventris	**PALDA22**	DA4	23.3.2001	Chaco, Argentina	-27.133	-61.46	AR/AR	PD
Didelphis albiventris	**PALDA3**	DA5	23.3.2001	Chaco, Argentina	-27.133	-61.46	AR/AR	PD
Didelphis albiventris	**PALDA4**	DA6	23.3.2001	Chaco, Argentina	-27.133	-61.46	AR/AR	PD
Didelphis albiventris	**PALDA5**	DA7	23.3.2001	Chaco, Argentina	-27.133	-61.46	AR/AR	PD
Triatoma infestans	PALDAV2^3*	TI1	23.3.2001	Chaco, Argentina	-27.133	-61.46	AR/AR	PD
Didelphis marsupialis	**PARAMA39**	DM26	6.08.2005	Barinas, Venezuela	8.43	-70.55	VS/VS	TS
Didelphis marsupialis	PARAMA40	DM27	6.08.2005	Barinas, Venezuela	8.43	-70.55	VS/VS	TS
Didelphis marsupialis	PARAMA41	DM28	6.08.2005	Barinas, Venezuela	8.43	-70.55	VS/VS	TS
Triatoma maculata	PGN900*	TM2	Unknown	Miranda, Venezuela	10.36	-66.75	VD/VS	HC
Homo sapiens	PII(BOL)	H20	Unknown	Cochabamba, Bolivia	-17.38	-66.16	COT/COT	IRD
Rattus rattus	RR5	RR3	Unknown	Trujillo, Venezuela	9.55	-70.51	VS/VS	HC
Didelphis marsupialis	SJM18	DM29	5.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Didelphis marsupialis	**SJM22**	DM30	6.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Didelphis marsupialis	SJM23	DM31	6.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Didelphis marsupialis	SJM26	DM32	6.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Didelphis marsupialis	SJM3	DM33	2.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Philander opossum	SJM32	PO4	7.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Didelphis marsupialis	SJM33	DM34	7.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Didelphis marsupialis	**SJM34**	DM35	7.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Didelphis marsupialis	SJM35	DM36	7.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Didelphis marsupialis	**SJM37**	DM37	9.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Didelphis marsupialis	**SJM39**	DM38	9.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Didelphis marsupialis	SJM40	DM39	9.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Philander opossum	**SJM41**	PO5	9.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Philander opossum	SJMC12	PO6	13.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Didelphis marsupialis	**SJMC3**	DM40	6.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Scuireus spadiceus	SJMC7	SS	11.9.04	Beni, Bolivia	-14.81	-64.6	LB/LB	TS
Rhodnius prolixus	SANRAFLB	RP5	16.5.04	Barinas, Venezuela	8.48	-70.73	VS/VS	TS
Rhodnius prolixus	SANRAFP2A	RP6	4.5.04	Barinas, Venezuela	8.48	-70.73	VS/VS	TS
Rhodnius prolixus	SANRAFP2B	RP7	4.5.04	Barinas, Venezuela	8.48	-70.73	VS/VS	TS
Rhodnius prolixus	TCSCII	RP8	Unknown	Cojedes, Venezuela	9.828	-68.43	VS/VS	HC
Didelphis albiventris	**TEDA**	DA8	23.3.2001	Chaco, Argentina	-26.93	-61.58	AR/AR	PD

189

Table 34 Continued from overleaf.

Rhodnius prolixus	TERF-1	RP9	Unknown	Portuguesa, Venezuela	8.34	-68.68	VS/VS	HC
Triatoma infestans	TEV55*	TI2	23.3.2001	Chaco, Argentina	-26.93	-61.58	AR/AR	PD
Didelphis marsupialis	USAOPOSSUM	DM41	Unknown	Lousiana, USA	30.5	-91.00	AM/AM	IRD
Dasypus novemcinctus	USAARMA	DN1	Unknown	Lousiana, USA	30.5	-91.00	AM/AM	IRD
Panstrongylus	V1*	PG	Unknown	Anzoat. Venezuela	10.04	-64.32	VD/VS	HC
Triatoma maculata	V2*	TM3	Unknown	Anzoat. Venezuela	10.04	-64.32	VD/VS	HC
Rhodnius prolixus	V4*	RP10	Unknown	Anzoat. Venezuela	10.04	-64.32	VD/VS	HC
Philander opossum	XE1313	PO7	23.11.83	Carajas, Brazil	-5.98	-51.33	NEB/NEB	EC
Didelphis marsupialis	XE1342	DM42	29.11.83	Carajas, Brazil	-5.98	-51.33	NEB/NEB	EC
Didelphis marsupialis	XE1381	DM43	16.12.83	Carajas, Brazil	-5.98	-51.33	NEB/NEB	EC
Didelphis marsupialis	XE1383	DM44	16.12.83	Carajas, Brazil	-5.98	-51.33	NEB/NEB	EC
Didelphis marsupialis	B2084	DM45	13.12.90	Belem, Brazil	-1.36	-48.36	NEB/NEB	EC
Didelphis marsupialis	B2085	DM46	03.01.91	Belem, Brazil	-1.36	-48.36	NEB/NEB	EC
Philander opossum	XE2913	DM47	05.05.88	Belem, Brazil	-1.36	-48.36	NEB/NEB	EC
Didelphis marsupialis	XE2929	DM48	10.08.88	Pará, Brazil	-5.83	-48.03	NEB/NEB	EC
Philander opossum	B3159	DM49	16.6.89	Pará, Brazil	-2.32	-49.52	NEB/NEB	EC
Didelphis marsupialis	XE3309	DM50	13.12.90	Belem, Brazil	-1.36	-48.366	NEB/NEB	EC
Marmosa cinerea	B3726	MC	10.6.98	Amapa, Brazil	0.03	-51.05	NEB/NEB	EC
Didelphis marsupialis	XE3776	DM51	23.10.98	Para, Brazil	-1.51	-49.21	NEB/NEB	EC
Didelphis marsupialis	XE3981	DM52	09.01.95	Para, Brazil	-0.15	-50.383	NEB/NEB	EC
Didelphis marsupialis	XE4389	DM53	16.05.96	Carajás, Brazil	-5.98	-51.33	NEB/NEB	EC
Didelphis marsupialis	XE4682	DM54	Unknown	Carajás, Brazil	-1.51	-49.21	NEB/NEB	EC
Philander opossum	XE4993	PO8	20.8.01	Para, Brazil	-3.66	-49.63	NEB/NEB	EC
Didelphis marsupialis	XE5011	DM55	13.01.99	Para, Brazil	-1.05	-46.76	NEB/NEB	EC
Didelphis marsupialis	XE5012	DM56	13.01.99	Para, Brazil	-1.05	-46.76	NEB/NEB	EC
Didelphis marsupialis	XE5017	DM57	13.01.99	Para, Brazil	-1.05	-46.76	NEB/NEB	EC
Didelphis marsupialis	XE5164	DM58	14.09.99	Para, Brazil	-1.71	-48.88	NEB/NEB	EC
Didelphis marsupialis	**XE5165**	DM59	14.09.99	Para, Brazil	-1.71	-48.88	NEB/NEB	EC
Didelphis marsupialis	**XE5167**	DM60	14.09.99	Para, Brazil	-1.71	-48.88	NEB/NEB	EC
Didelphis marsupialis	B5770	DM61	4.13.99	Para, Brazil	-1.05	-46.76	NEB/NEB	EC
Didelphis marsupialis	B5781	DM62	16.9.99	Para, Brazil	-1.71	-48.88	NEB/NEB	EC
Didelphis marsupialis	B5787	DM63	16.9.99	Para, Brazil	-1.71	-48.88	NEB/NEB	EC
Didelphis marsupialis	XE5809	DM64	23.5.02	Tocantins, Brazil	-9.18	-48.183	NEB/NEB	EC
Didelphis marsupialis	XE5847	DM65	12.11.01	Para, Brazil	-1.38	-48.86	NEB/NEB	EC
Philander opossum	XE5495	PO9	21.02.01	Para, Brazil	-1.86	-52.2	NEB/NEB	EC
Didelphis marsupialis	XE6004	DM66	17.10.03	Roraima, Brazil	2.76	-60.53	NEB/NEB	EC

[a] Sample codes in bold represent strains cloned in this study.

[b] See Figure 55.

[c] Decimal degrees

[d] Assigned by geography and transmission cycle (Section 5.2.2.1.6). VD – Venezuela domestic; AM – North America; NEB – North Eastern Brazil; CH – Chile; COT – Cochabamba; COL-Colombia; AR – Argentina; LB – Lowland Bolivia.

[e] Assigned by geography and population reassignment (Section 5.2.2.1.7).

[f] Codes as in Table 32.

[g] Known to be in the chronic, symptomatic phase of Chagas disease.

*Originates from a domestically caught vector.

Number in parentheses: age/sex

190
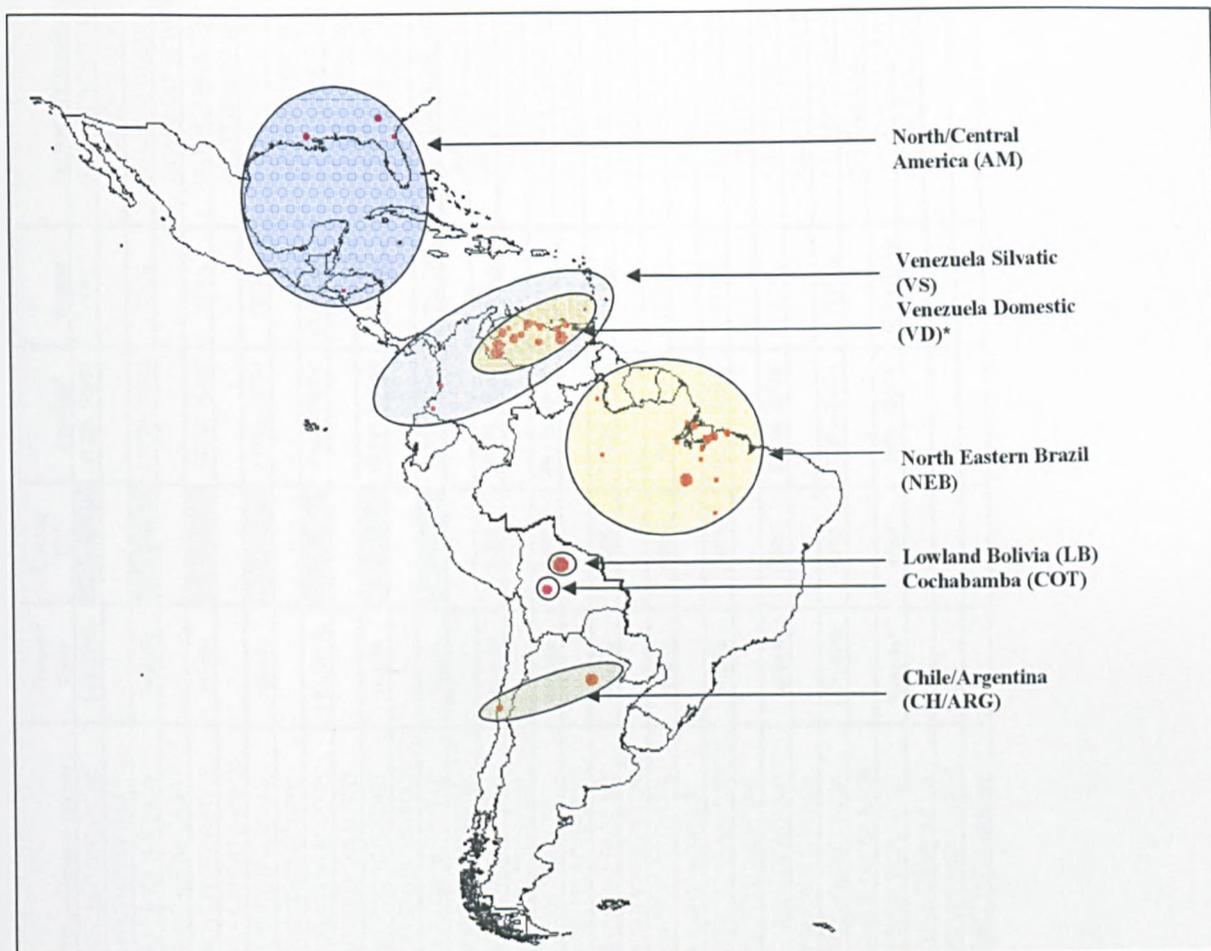
Figure 54 Geographic distribution of TCI strains subjected to microsatellite analysis in this study. Red points indicate sample origins and point size is proportionate to sampling density. Coloured ellipses indicate geographic population subdivision before statistical assignment (Population 1 in Table 34) *A priori* population assignment rationale is detailed in Section 5.2.2.1.6. *Venezuelan domestic isolates fall within the same approximate geographic distribution as Venezuelan silvatic isolates.

5.2.2.1.2 Microsatellite loci employed in analysis

The panel of microsatellites employed to analyse the TCI strains is shown in Table 35. Microsatellites repeat sequences were identified and selected for analysis as described in Section 3.7. Loci monomorphic within TCI were excluded, as were those that amplified poorly across all samples, to reduce the chance of non-co-dominance due to null alleles, also as in Section 3.7. In total 48 microsatellites worked robustly across the dataset, and are listed in Table 35. Among these, three had been employed in an earlier study (215).

Table 35 Microsatellite loci employed to study TC1 intra-lineage diversity (continued overleaf).

Primer Code	Contig ID[a]	Position[b]	Repeat[c]	Forward Primer/Reverse primer
1TETP2c	6986	84365..84500	$(CA)_n$	ACCAATACCACCACCACCAC / TCAGAAATCTGCTTTCCC
1FAMP1b	6986	120977..121012	$(TA)_nCAGT(GA)_n$	CTAAGAAGCAGGCAACCCAG / CACCGCCAGCAGTAACAAAT
1FAMP3c	6986	99171..99204	$(TA)_n$	CGGCACACACTTCGAGATAA / CGCCTAGTCAAATGATGATACG
1TET14b	6986	75330..75360	$(TCC)_n$	AITCCTTGTCGGTGAACACC / GCCCTCCGTTTATTCCTCTC
1TETP13a	6986	92287..92216	$(CA)_n$	TGTTGAAATGATTTGACCCGA / AGAGTCACGCCGCAAAGTAT
1TETP2b	6986	75669..75701	$(TA)_n$	TGAAGGAGATTCTCTGCGGT / CTCTCATCTTTGTTGTGTCCG
1TETP5a	6996	42683..42715	$(TC)_n$	CGCTCTCAAAGGCACCTTAC / ATATGGACGCGTAGGAGTGC
1TETP6a	6996	28781-28980	$(TG)_n$	TTGCTTGCTTTCCACGTGTA / TATTCCCTTTGCCTTTGCTG
1TETP10b	6996	94129-94191	$(CAA)_n$	AITCGGCCTGTTTGTATTCG / TTGGTTGTGTTTACTGCTGC
1TETP11a	6996	16332-16526	$(GA)_n$	GCAAAAGCAACAAAAACAGC / GAGCAACACGAAGGGGAATA
famp1a	7098	64001-64129	$(TG)_n$	GAGCAGATCTTCCTTGTGCC / TGGTTGAAATGCACGCATC
famp2a	7218	76285-76410	$(TC)_n$	GACATGTATGCTTGAAACCTCC / TCCATCTCCCTTCACACTCC
1TETP12a	7218	5369-5576	$(TA/GA)_n$	TGTTGATCAACGCGCATAAAT / TTCCATTGCCTCGTTTTAGA
famp6a	7143	119499..119520	$(TG)_n$	TCGTTCTCTTTACGCTTGCA / TAGCAGCACCAAACAAAACG
nedp8a	7143	48024..48053	$(TG)_n$	GAAACGCACTCACCCACAC / GGTAGCAACGCCAAACTTTC
vicp2a	7143	88658-88832	$(CT)_n$	CATCAAGGAAAAACGGAGGA / CGGTACCACCTCAAGGAAAG

Primer Code	Contig ID[a]	Position[b]	Repeat[c]	Forward Primer/Reverse primer
1TETP6b	7206	51225-51373	$(TA)_n$	GGAAACACATCACGCAAAGA / AGTGACAAAGGGGGACATTG
nedp5a	7206	69979..70016	$(TC)_n$	CCAACATTCAACAAGGGAAA / GCATGAATATTGCCGGATCT
tetp8a	7206	16233..16257	$(TA)_n$	CGTGTGCACAGGAGAGAAAA / CGTTTGGAGGAGGATTGAGA
vicp4a	7206	25751..25779	$(TCC)_n$	AGACGTTCATATTCGCAGCC / AGCCACATCCACATTTCCTC
1FAMP9a	7206	39675-39924	$(TAA)_n(A)_n$	TCAAAGGGTGCTAATCCAAA / AACCATGAGGGAAGAAGAAA
vicp2b	7853	94443..94463	$(TA)_n$	TGTAACGGTAGGTCTCAATTCG / TTGCACTTGTGTATCTCGCC
vicp5a	7853	72728-72864	$(CT)_n(TA)_n$	GCAGAAGCGCACAGACACAT / AAAGTGCCATCCCACCCTC
petp6a	8305	88534..88549	$(CA)_n$	GTCGCCATCATGTACAAACG / CTGTTGGCGAATGGTCATAA
tetp3a	8305	49478..49495	$(TC)_n$	CGTACGACGTGGACACAAAC / ACAAGTGGGTGAGCCAAAAG
tetp7a	8305	15729..15742	$(CA)_n$	ACCCAGAGGGGAAAAAAAGA / TTTACGGTTGGTTCGTGTGA
vicP1a	8305	987..1016	$(TAA)_n$	CCGGGTAGAAGAACCATAA / TGGGTATTCACGACGAGAAG
vicp1b	8305	46598-46638	$(TA)_n$	AACCCGGCAGATACATTAG / TTCATTTGCAGCAACACACA
vicp6b	8305	87584..87606	$(CA)_n$	GTGTCGTTGCTCCCAAACTC / AAACTTGCCAAATGTGAGGG
1FAMP9c	8370	24890-25024	$(TTG)_n$	AIGGGTGCGGAGAGGTATGTC / TGTCAAAACAGCGGAAAGAA
famp4a*	8589	1001-10154	$(CA)_n(GA)_n$	ACGACCAAAGCCATCATT / GATGCTAACTGCTCAAGTGA

Primer Code	Contig ID[a]	Position[b]	Repeat[c]	Forward Primer/Reverse primer
1TETP3a	8328	44002..44057	(TA)b	AGAAAAGGTTTACAACGAGCG / CGATGGAGAACGTGAAACAA
1TETP4a	8328	71097-71226	(GA)a	GTCACACCACTAGCGATGACA / ACTGCACAATACCCCCTTTG
1TETP4b	8328	32430-32629	(TTA)b	GAGAGAGATTCGGAAACTAATAGC / CATGTCCCTTCCTCCGTAA
1TETP8a	8328	39618-39874	(CA)b(TA)b	CATGTCATTAAGTGGCCACG / GCACCATGTTGGTTGTTGGAA
1FAMP12a	8328	124985-125219	(TG)b	GCG°GTTATTAAACACTCGCT / GCCCGGTATCATTGAAAAGA
1FAMP7b	8328	83338-83417	(CA)b	CTACCTTCTTTCCTCCCTAACC / TTTGCTCTGGACTGCATGC
1TETP1b	8328	63326-63566	(TG)a	AAGAGAGGCACTTCCCTGTGA / GAGGAAGAGGAAGTACAGTTGAGC
famp3a	8646	89251-89397	(TG)b	TGCTCTTGTTGTTGGAGTGC / CAGCAGGACAATCCAACCAAC
famp5a	8646	58996-59147	(A)b(CA)b	GCGGTACACCAACATGTACG / GTGTGTTTGTGTGTGAGAGGC
famp8a	8646	886680..88708	(TCG)a	ACCACCAGGAGGACATGAAG / TGTACACGGAACAGCGAAG
famp8b	8646	12639-12819	(TA)b	AACATCCTCCACCTCACAGG / TTTGAATGCGAGGTGGTACA
vicp6a	8812	7458..7500	(CA)b	AGTTGACATCCCCAAGCAAG / CCCTGATGCTGCAGACTCTT
famp1b*		Unknown	(CA)aA(CA)b	GCGTAGCGATTCATTTCC / ATCCGCTACCACTATCCAC
famp2c*		Unknown	(CT)b(TG)b	GATCCCGCAATAGGAAAC / GTGCATGTTCCATGGCTT
petp4a		Unknown	Unknown	CTTAAAGAGATACAAGAGGGAAGG / CTGTTATTTCAATAAACACGGGG
1FAMP2a		Unknown	Unknown	TGAGAGTGACGGGAGAGAGC / GGGGCACATTATGTGTGTGCT
1TETP15		Unknown	Unknown	AACAAAATCTAGCGTCTACCATCC / GGTGTTGGCGTGTATGATTG

[a] Refers to a syntenous fragment of sequence identified by the *T. cruzi* genome project (376). Colours represent primers on the same contiguous fragment (Contig). Uncoloured contig codes indicate a unique contig. Microsatellite loci on unidentified contigs are marked 'unknown'.

[b] Refers to the position of the of the primer binding site, flanking region and repeat motif 5'-3' along the sense strand. Unidentified positions are marked 'unknown'.

[c] Refers to the repeat type (e.g. di-nucleotide, tri-nucleotide). Unidentified repeat types are marked 'unknown'.

* Alternative codes are as follows: famp1b = mclf10; famp2c = scle10; famp4a = scle11 (215).

Microsatellite loci were distributed across 13 known syntenous regions (contigs) within the online *T. cruzi* genome database (376). The largest number of syntenous loci was seven, the smallest one. The contig number and position of five microsatellite loci could not be defined, including two from a previous study. Microsatellite repeat types for the majority of loci were based on sequences from the online *T. cruzi* genome database (376). Microsatellite repeat types for those loci used in the previous study are based on those in the literature (215).

5.2.2.1.3 Missing data, clone correction and multiple alleles.

Even after the exclusion of null loci (See Sections 5.2.2.1.2 and 3.7.2), as well as modified annealing temperature re-runs of selected samples and loci (Section 3.7.4), a level of missing data remained within the dataset. The most common limiting factor influencing this residual level was the availability of parasite whole genomic DNA, particularly for those samples obtained from overseas for which no live organisms were available. In total, 5.3% of all loci failed to amplify, equating to 2.6 loci per sample. Two samples, 8104 and 11804 (Table 34), failed to amplify over more than 50% of all loci – again due to a shortage of DNA – and were dropped from all analyses except that described in Section 5.2.2.1.4. Excluding these two isolates, the level of missing data drops to 4.5%, or ~2.2 loci per sample.

Clone correction was attempted using GENALEX v6.1 to remove bias in the estimation certain of population parameters (292). Remarkably, no multilocus matching samples were found in any populations.

Multiple alleles (3+) at single loci, indicating the possibility if mixed infections and/or possible aneuploidy, were observed in the dataset. Samples were screened by linkage group for linked multi-allelic loci as a means of distinguishing aneuploidy from multi-clonality. This phenomenon was only observed once, where four linked possibly trisomic loci were present in isolate AM6 along contig 8486. Samples demonstrating multiple alleles were biologically cloned (see Section 3.4.2) where live organisms were available, and the relevance of multiclonality and/or aneuploidy in generating multi-allelic loci is investigated fully in Chapter 6. After cloning, 27 samples with one or more multi-allelic loci remained,

all of which were uncloned, and for which no live organisms were available. The frequency of multi-allelic loci per individual per population is summarised in Table 37, while the final frequency of multiple alleles as a proportion of the total number was ~0.51%. The approach developed to circumvent the population genetic problems posed by the occurrence of multi-allelic loci is described in Section 3.7.6.1.1.

5.2.2.1.4 Distance measures at the individual level
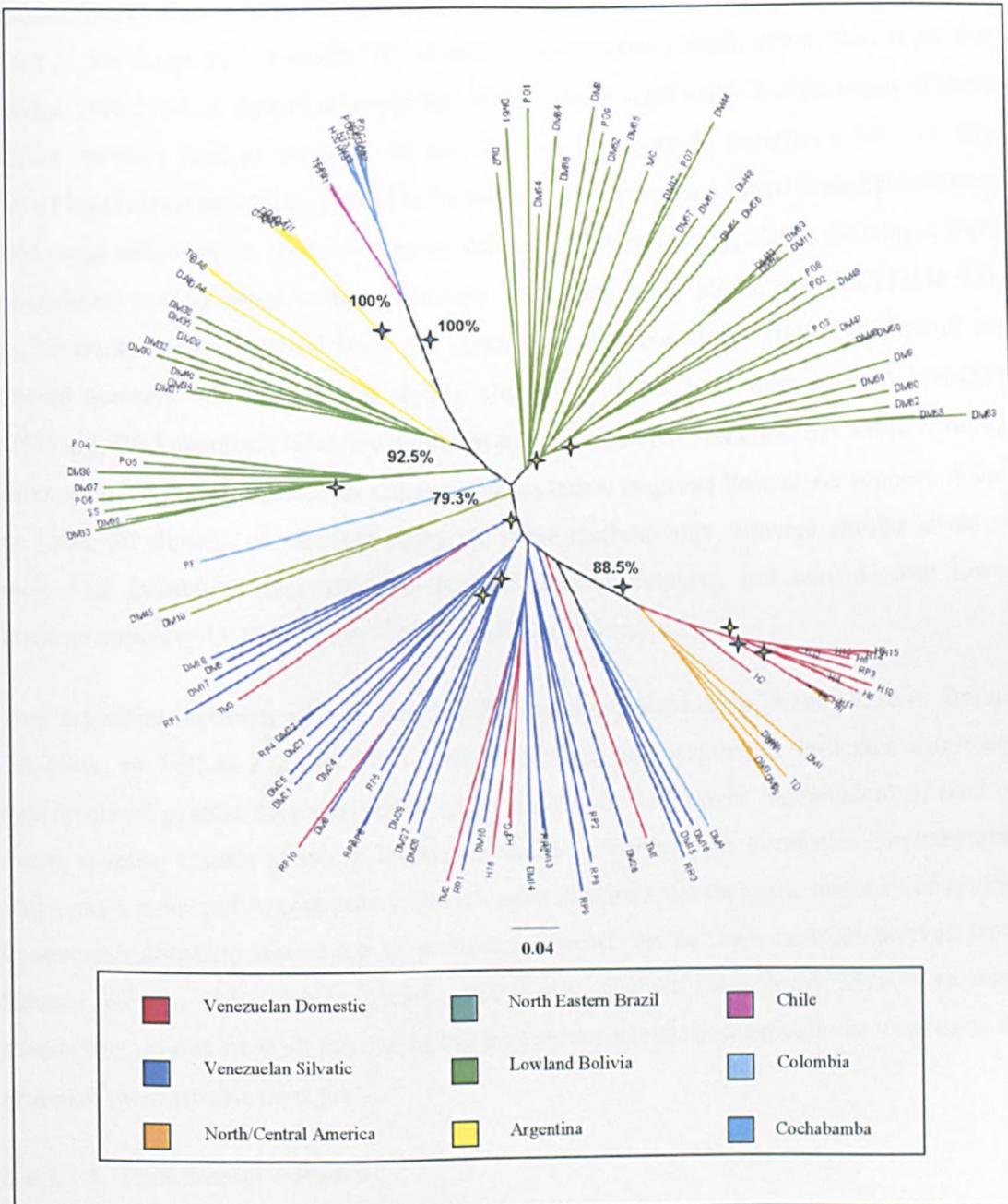
Figure 55 Unrooted neighbor-joining D_{AS} tree based on the multilocus microsatellite profiles of 135 TCI isolates. Values indicate D_{AS}-based bootstrap values over 10,000 trees on major clades >75%. Blue stars indicate $\delta\mu^2$ bootstrap values >50% over 10,000 trees. Yellow stars indicate multi-allelic topology instability over 1000 trees (<90% recovery, see Section 3.7.6.1.2 & 3.7.6.1.1) Tip labels indicate host/vector origin, as in Table 34.

Genetic distances between individual strains were assessed using two models of microsatellite evolution, D_{AS} (IAM) (257) and $\delta\mu^2$ (SMM) (212), as described in Section 3.7.6.1.2. To avoid the exclusion of samples demonstrating multi-allelic loci from the analysis, 1000 random diploid re-samplings of the dataset were made and the mean of these distance matrices used as the basis for the D_{AS} tree in Figure 55 (Section 3.7.6.1.1). The level of topological instability proved to be low, with only seven internal branches showing <90% congruence across 1000 re-sampled datasets. Unsurprisingly, clades showing a level of instability corresponded with populations containing multi-allelic samples (Table 37). D_{AS} bootstrap values, derived from the mean over 100 bootstraps from 100 diploid re-sampled datasets, were robust for tightly clustering clades with long internal branches (>75% e.g. Cochabamba/Chile). By comparison, short internal branches, like those defining Venezuelan and North Eastern Brazilian silvatic clades, received little or no support. A $\delta\mu^2$ tree (data not shown), constructed using the same methodology, showed similar levels of topological instability over 1000 re-sampled diploid datasets, but considerably lower bootstrap values (≤73.7%). Those >50% are shown in Figure 55.

Deep branching between silvatic isolates derived from Venezuela, North Eastern Brazil, Colombia, as well as a subset from lowland Bolivia and Argentina, indicates extremely high levels of genetic diversity within geographic foci, seemingly independent of host or vector species. Clades showing limited diversity - Venezuelan domestic, Cochabamba, Chile, and a subset of Argentinean isolates - exist in stark contrast to the majority of strains. Venezuelan domestic strains are of particular interest: all but two samples derived from humans form a tight cluster, clearly genetically distinct from local silvatic isolates. Meanwhile, strains from all but one of the triatomines caught domestically in Venezuela do resemble local silvatic lineages.

5.2.2.1.5 Isolation by distance

Correlation between genetic distance and geographic distance were tested using a Mantel's test, implemented in GENALEX v6, as described in Section 3.7.6.1.3 (292). Briefly, two matrices, one a pairwise measure of D_{AS} between all individuals identical to that used to construct the tree in Figure 55, and one a pairwise measure of physical distance (constructed using ARCVIEW v3.3 as in Section 3.7.6.1.3), were correlated. A random

permutation procedure involving 999 permutations was implemented to generate a p-value for the correlation statistic R_{XY}. The results of this analysis are shown in Figure 56.
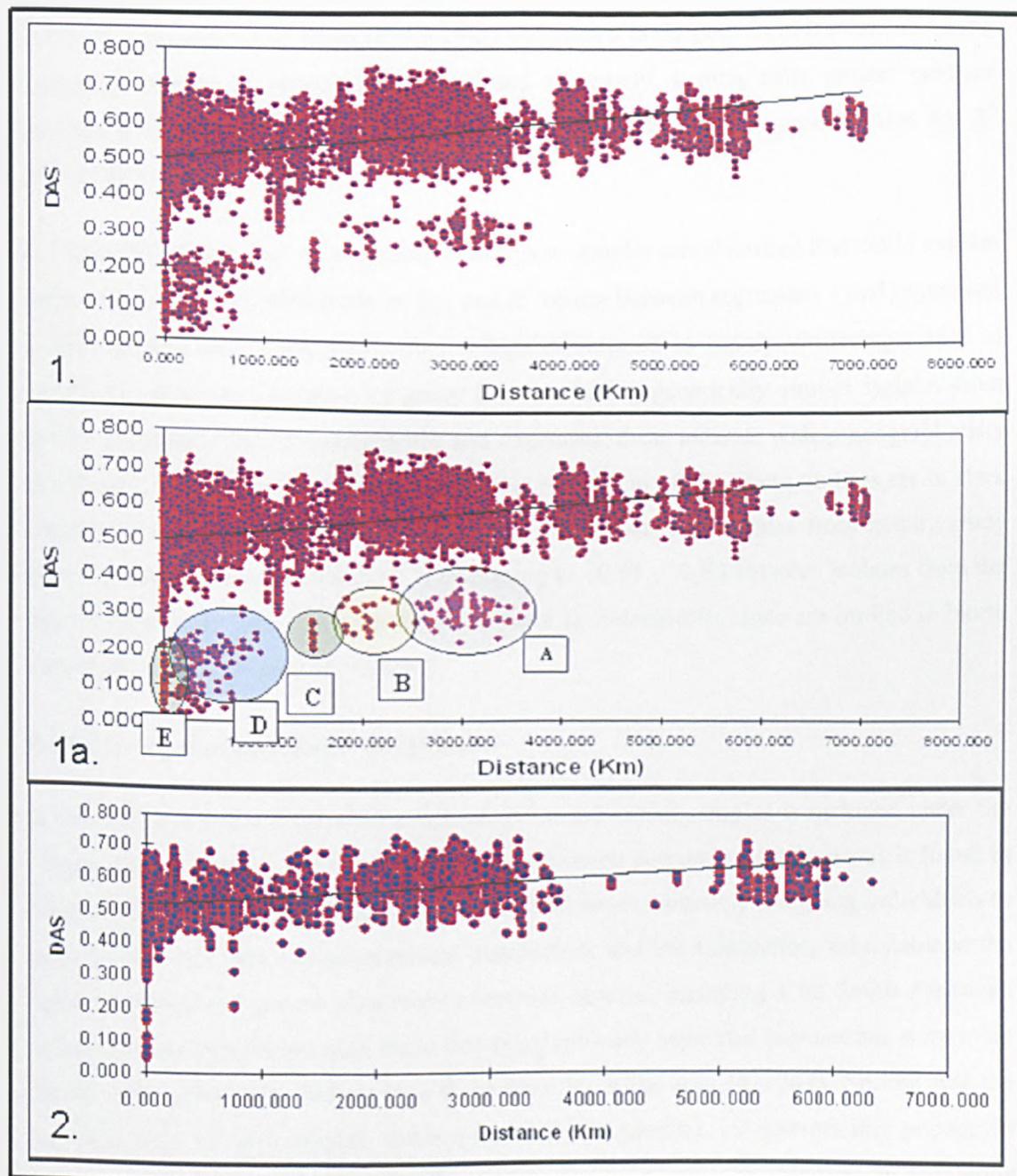
Figure 56 Linear regression (Mantel's test) of microsatellite D_{AS} pairwise genetic distance against geographic distance (Km) for TCI isolates. Regression lines plotted in black. 1. – Entire TCI dataset, non-annotated. 1a. – Entire TCI dataset, outliers annotated (A- Venezuela domestic vs North/Central America; B – North/Central America vs North/Central America; C- Cochabamba vs Chile; D – Venezuela domestic vs Venezuela domestic; E – Argentina vs Argentina (subset of samples) + Cochabamba vs Cochabamba. 2. - All isolates from *D. marsuplialis*. See text for correlation statistics.

A further regression, involving only isolates derived from *D. marsupialis* was also undertaken (Figure 56). A weak but significant positive correlation between genetic and geographic distance was observed when all TCI isolates were considered (R_{XY} = 0.394, p<<0.05), although the R^2 value (R^2= 0.1543) indicates a fairly poor fit of the data about the regression line. If *D. marsupialis* is analysed separately, a marginally greater positive correlation is observed (R_{XY} = 0.429, p=<<0.05), coinciding with a greater value for R^2 (R^2= 0.1838).

In Figure 56, scatter plot 1a, a number of outlying samples are identified that could explain some of the observed differences in R_{XY} and R^2 values between regression 1 and regression 2. Five groups have been highlighted, which correspond to tightly clustering clades in Figure 55. With the exception of group E, representing genetically similar isolates from nearby geographic foci in Cochabamba and Argentina, these indicate widely geographically distributed, but genetically similar groups (D_{AS} = ~0.01 to ~0.4). These outliers are in stark contrast to the majority of strains, where D_{AS} values between isolates from nearby study sites fluctuate between ~0.4 and ~0.7, increasing to ~0.55 - ~0.65 between isolates from the most distant study sites. Analogous outliers from *D. marsupialis* alone are limited to North American isolates (Figure 56, Graph 2)

5.2.2.1.6 *A priori* population assignment.

A full list of *a priori* population assignments for individual samples is included under the 'Population 1' column in Table 34. A map of all seven population subdivisions is found in Figure 54. Three principal factors were considered when arbitrarily assigning individuals to populations. The first was geographical distribution, and the assumption, reasonable in the light of population genetic data from numerous species, including wild South American mammals and triatomines, was made that geographically separated populations were more likely to be genetically distinct than those from the same area (42, 344). Second was the possible role of transmission cycle (domestic or silvatic) in determining population subdivision in *T. cruzi* (136). Finally sample size is a key consideration in generating reliable estimates of population genetic parameters, and, if possible, populations consisted of 10+ individuals. All populations conformed to these three criteria, except North/Central America (N=7), as well as Cochabamba and Argentina/Chile, where domestic and silvatic

populations were grouped together to provide a sample size of nine and twelve individuals respectively.

5.2.2.1.7 Statistical population assignment

Statistical allele frequency-based population assignment was carried out as in Section 3.7.6.1.6 using GENALEX v6 in order to test *a priori* population sub-division by geography and transmission cycle (292). A full table of –ln likelihood scores is found in Table 36 and a list of re-assigned individuals in the 'Population 2' column in Table 34.

Table 36 Allele frequency-based –ln likelihood values for population assignment of individuals within the TCI dataset. Values in bold indicate most likely population (least negative –ln likelihood) value for each individual.

Argentina +Chile	Argentina +Chile	Lowland Bolivia	North Eastern Brazil	Cochabamba	North/Central America	Venezuela Domestic	Venezuela Silvatic
CHILE22	-36.987	-50.800	-65.183	**-24.643**	-99.420	-93.405	-65.093
PALDAPDV^2V3	**-20.553**	-62.109	-64.358	-62.061	-103.001	-89.345	-64.769
TEDA	**-30.323**	-66.618	-69.497	-79.305	-112.721	-102.584	-70.385
TEV55	**-46.579**	-63.324	-72.089	-95.424	-109.512	-98.017	-73.621
CHILEWALL	**-38.321**	-57.610	-68.177	-34.957	-104.684	-97.270	-67.922
PALDA1	**-19.132**	-59.207	-61.553	-56.464	-106.943	-97.211	-64.194
PALDA20	**-16.685**	-56.238	-64.143	-51.987	-100.825	-94.085	-64.160
PALDA21	**-19.122**	-66.712	-67.555	-61.929	-114.040	-102.073	-70.557
PALDA22	**-28.083**	-65.096	-66.361	-75.907	-112.897	-100.186	-68.493
PALDA3	**-43.519**	-59.398	-65.797	-95.123	-101.804	-92.101	-69.630
PALDA4	**-18.727**	-62.171	-65.456	-58.816	-113.244	-100.622	-68.901
PALDA5	**-19.390**	-63.534	-64.042	-58.803	-110.709	-100.878	-68.722
Lowland Bolivia	Argentina +Chile	Lowland Bolivia	North Eastern Brazil	Cochabamba	North/ Central America	Venezuela Domestic	Venezuela Silvatic
SJM18	-60.631	**-50.799**	-60.927	-90.792	-97.453	-85.577	-65.543
SJM35	-78.531	**-36.874**	-67.983	-80.489	-102.305	-86.371	-71.239
SJM37	-84.168	**-47.178**	-75.858	-86.834	-107.359	-97.823	-77.372
SJM40	-90.883	**-57.719**	-76.411	-90.388	-107.266	-99.151	-79.039
SJM41	-87.656	**-44.590**	-70.956	-93.619	-91.895	-79.863	-69.171
SJMC12	-85.900	**-53.895**	-78.896	-92.724	-105.343	-95.660	-71.823
SJMC3	-55.525	**-35.211**	-57.571	-85.656	-97.944	-84.457	-62.116
SJMC7	-79.575	**-43.468**	-74.592	-76.834	-97.369	-85.986	-70.255
SJM22	-71.770	**-48.780**	-74.560	-90.754	-110.837	-96.661	-73.715
SJM23	-60.737	**-38.508**	-61.054	-74.096	-98.378	-85.300	-67.179
SJM26	-69.747	**-40.522**	-70.671	-86.147	-111.931	-100.733	-67.093
SJM39	-58.056	**-45.086**	-64.002	-77.219	-102.173	-88.035	-66.364
SJM3	-89.816	**-42.909**	-79.124	-89.258	-112.021	-93.386	-77.517
SJM32	-81.582	**-52.706**	-71.421	-83.902	-112.828	-102.350	-73.339
SJM33	-63.082	**-38.601**	-57.813	-81.141	-101.316	-85.355	-66.132
SJM34	-66.507	**-51.318**	-64.636	-80.944	-103.282	-91.481	-67.737

North Eastern Brazil	Argentina +Chile	Lowland Bolivia	North Eastern Brazil	Cochabamba	North/Central America	Venezuela Domestic	Venezuela Silvatic
B2310	-83.592	-72.966	-46.823	-103.006	-98.565	-90.554	-66.801
XE1342	-70.477	-63.469	-40.567	-96.157	-83.134	-84.350	-64.377
XE1381	-70.454	-64.631	-36.813	-98.491	-87.702	-87.397	-61.923
XE1383	-84.008	-78.091	-49.619	-106.951	-103.445	-96.540	-69.897
XE2929	-69.585	-66.165	-49.823	-99.795	-102.982	-93.584	-72.758
XE4389	-72.325	-65.240	-39.080	-102.203	-91.132	-92.709	-68.740
XE4993	-70.684	-56.701	-40.925	-88.912	-75.531	-73.813	-53.410
XE5809	-78.961	-73.836	-61.154	-103.870	-102.879	-80.824	-58.488
B1947	-84.553	-86.982	-57.660	-106.976	-99.140	-94.341	-76.353
B2026	-81.507	-70.881	-48.463	-111.509	-94.943	-89.879	-65.017
B2077	-71.189	-64.813	-41.924	-90.844	-84.785	-81.551	-60.768
B2311	-84.860	-82.394	-60.473	-102.272	-108.394	-99.512	-72.121
B5470	-96.260	-70.913	-51.405	-100.477	-109.043	-101.447	-79.357
XE2084	-91.876	-86.224	-72.953	-105.705	-97.249	-95.564	-76.999
XE2085	-74.810	-70.505	-45.648	-93.127	-94.126	-90.043	-69.203
XE2913	-78.896	-74.613	-43.080	-97.780	-96.864	-95.068	-69.864
XE3159	-79.547	-69.841	-54.157	-103.770	-102.270	-94.819	-69.647
XE3309	-81.685	-76.009	-48.951	-98.951	-96.962	-93.626	-69.844
XE3776	-86.335	-71.680	-59.188	-99.886	-103.985	-99.328	-67.571
XE4682	-48.949	-39.608	-29.320	-59.872	-62.753	-58.617	-41.090
XE5011	-65.782	-58.968	-40.862	-90.719	-86.256	-84.368	-63.612
XE5012	-72.028	-67.206	-46.064	-96.117	-87.270	-88.932	-68.115
IM4810	-78.777	-75.436	-63.046	-98.146	-95.069	-89.764	-71.743
XE5017	-75.006	-67.566	-43.971	-94.282	-96.641	-92.118	-65.271
XE5164	-91.532	-78.748	-48.756	-115.735	-111.536	-105.318	-78.414
XE5165	-77.698	-69.950	-46.648	-96.861	-99.963	-91.793	-70.417
XE5167	-85.620	-78.259	-50.831	-107.907	-99.434	-93.974	-72.044
XE5770	-76.104	-68.890	-42.014	-83.249	-92.266	-93.026	-69.421
XE5781	-84.293	-75.139	-44.294	-100.785	-95.486	-101.017	-72.969
XE5787	-98.825	-87.773	-54.152	-121.735	-115.802	-113.532	-85.852
XE5847	-77.728	-65.003	-51.372	-87.929	-107.237	-95.149	-71.287
XE6004	-61.148	-66.484	-48.838	-86.439	-91.044	-83.176	-52.828
XE3726	-83.633	-80.554	-52.586	-97.688	-90.807	-86.711	-61.447
XE3981	-81.681	-75.592	-50.169	-98.759	-99.940	-90.797	-64.006
XE5945	-72.668	-71.200	-53.676	-100.308	-94.971	-82.993	-65.180
CJO20	-75.790	-71.480	-41.573	-100.208	-84.896	-87.783	-67.235
CJO35	-71.669	-61.103	-36.861	-85.127	-88.182	-74.651	-52.080
XE1313	-85.281	-76.381	-44.469	-101.252	-103.146	-90.609	-63.282

Cochabamba	Argentina +Chile	Lowland Bolivia	North Eastern Brazil	Cochabamba	North/Central America	Venezuela Domestic	Venezuela Silvatic
COTMA22	-38.997	-49.025	-67.491	-6.504	-108.541	-99.678	-64.208
COTMA38	-40.271	-50.145	-66.621	-11.041	-105.541	-98.055	-64.111
COTMA47	-32.902	-46.011	-64.571	-6.696	-99.143	-95.573	-60.597
COTMA55	-47.321	-56.221	-72.275	-9.677	-111.766	-104.678	-70.326
COTMA9	-34.043	-49.006	-67.441	-7.235	-104.872	-99.671	-62.934
P234	-42.660	-57.026	-70.417	-15.706	-109.106	-103.769	-69.198
P238	-42.941	-53.875	-67.779	-7.221	-107.687	-101.538	-66.643
P268	-39.760	-51.470	-69.571	-11.232	-109.076	-100.864	-66.702
PII(BOL)	-39.966	-52.915	-63.299	-10.335	-101.152	-97.869	-64.990
North/Central America	Argentina +Chile	Lowland Bolivia	North Eastern Brazil	Cochabamba	North/Central America	Venezuela Domestic	Venezuela Silvatic
92090802P	-97.343	-79.279	-74.207	-104.062	-32.772	-42.124	-55.154
93070103P	-99.248	-82.449	-76.547	-111.363	-13.736	-48.117	-62.621
92101601P	-87.850	-73.008	-69.888	-100.806	-15.569	-41.189	-53.395
USAARMA	-98.389	-73.372	-65.670	-112.376	-17.230	-44.272	-54.120
USAOPOSSUM	-95.950	-71.532	-63.931	-107.233	-16.996	-44.779	-51.945
DAVIS9.90	-102.826	-77.187	-70.756	-120.319	-29.204	-47.291	-63.201
FLORIDAC1D12	-97.286	-82.170	-75.350	-107.062	-18.250	-44.167	-63.743
Venezuela Domestic	Argentina +Chile	Lowland Bolivia	North Eastern Brazil	Cochabamba	North/Central America	Venezuela Domestic	Venezuela Silvatic
BARM	-77.195	-79.519	-72.032	-92.240	-78.062	-60.975	-46.131
PGN900	-89.203	-79.874	-73.530	-101.066	-96.062	-75.765	-53.417
V1	-83.510	-70.644	-62.496	-102.053	-79.434	-43.179	-39.768
V2	-89.553	-84.836	-71.111	-104.268	-98.624	-74.952	-59.474
V4	-91.806	-85.740	-73.349	-111.634	-107.845	-77.630	-51.142
7570	-85.211	-66.287	-71.940	-97.795	-33.298	-23.797	-53.442
11541	-88.441	-69.828	-76.873	-100.306	-43.006	-26.009	-56.749
11713	-86.434	-68.376	-75.020	-101.247	-35.649	-17.850	-53.997
11838	-75.493	-62.748	-65.644	-92.310	-30.357	-17.341	-47.491
11881	-93.038	-69.679	-73.601	-104.787	-38.600	-19.653	-56.146
CASA BONIFICA	-89.078	-74.336	-75.611	-108.922	-39.148	-29.060	-58.727
JR	-88.443	-76.568	-69.214	-103.117	-95.112	-61.719	-48.880
9010	-82.855	-72.163	-65.562	-107.258	-79.763	-42.799	-39.477
9354	-93.259	-65.986	-74.820	-101.548	-39.646	-22.735	-55.309
10775	-84.691	-66.662	-70.723	-101.795	-32.562	-20.301	-50.801
10801	-86.450	-68.072	-73.378	-101.548	-39.295	-21.759	-52.376
11006	-89.262	-68.605	-72.410	-103.494	-34.101	-21.045	-55.674
11042	-97.017	-74.978	-80.097	-105.623	-35.206	-22.357	-57.300
11124	-97.139	-74.190	-80.831	-108.435	-45.843	-21.736	-59.054
11398	-65.492	-50.285	-59.202	-77.389	-31.668	-20.915	-42.439

Venezuela Silvatic	Argentina +Chile	Lowland Bolivia	North Eastern Brazil	Cochabamba	North/Central America	Venezuela Domestic	Venezuela Silvatic
AM3	-79.208	-72.003	-69.833	-99.222	-93.917	-70.328	-55.936
CD45	-83.965	-71.733	-60.385	-103.584	-81.387	-51.940	-40.558
CO22	-95.072	-79.945	-64.211	-99.430	-95.511	-60.143	-50.840
CO57	-90.936	-78.624	-70.018	-102.856	-76.265	-50.561	-41.918
CO75	-88.619	-76.054	-63.569	-98.136	-75.853	-57.785	-44.333
CO84	-85.885	-75.180	-66.181	-95.458	-73.846	-51.849	-38.296
TCSCII	-89.580	-71.291	-68.900	-92.294	-84.185	-59.053	-46.958
361TA	-90.272	-77.447	-78.963	-103.536	-82.502	-49.388	-50.137
458	-65.096	-69.254	-56.136	-96.420	-108.728	-78.507	-67.477
AM6	-91.093	-78.706	-70.878	-108.459	-96.744	-68.364	-49.446
DM1	-89.593	-84.502	-70.808	-112.395	-71.978	-63.449	-49.361
DM4	-88.627	-77.099	-74.189	-104.212	-83.640	-64.220	-45.202
LL2LA	-85.379	-73.787	-62.842	-100.013	-89.778	-66.524	-45.759
M12	-89.230	-68.813	-58.990	-98.361	-70.623	-59.811	-40.102
M13	-88.542	-67.598	-66.838	-102.378	-93.757	-63.374	-47.221
M15	-85.702	-65.023	-59.085	-100.473	-83.408	-63.426	-39.572
M16	-81.036	-68.422	-59.734	-96.944	-79.254	-61.885	-43.182
M18	-81.901	-75.111	-68.689	-105.184	-89.445	-61.736	-43.567
M7	-98.125	-72.889	-68.438	-113.309	-71.233	-56.030	-44.292
PARAMA39	-92.456	-78.736	-73.770	-99.138	-92.694	-66.061	-51.621
AMRPA34	-73.047	-69.691	-61.766	-89.385	-73.904	-60.905	-48.672
PARAMA40	-86.238	-74.636	-64.517	-99.535	-91.103	-68.307	-47.766
PARAMA41	-90.798	-70.428	-61.283	-89.319	-89.299	-68.674	-48.936
RR5	-97.145	-88.530	-76.722	-112.837	-85.151	-61.792	-49.451
SRLB	-76.928	-72.096	-60.981	-90.380	-82.054	-58.820	-37.017
SRP2A	-83.974	-77.020	-71.061	-96.623	-90.221	-64.596	-47.432
SRP2B	-77.027	-68.774	-63.176	-84.657	-74.969	-54.362	-36.947
TERF-1	-94.464	-77.577	-73.571	-108.879	-90.589	-69.372	-50.124
DMSU8	-70.685	-70.961	-66.271	-93.492	-95.178	-69.324	-56.355
DMSUC	-75.983	-68.590	-67.599	-97.239	-82.749	-64.739	-52.701
CALC104	-79.543	-72.321	-62.119	-93.604	-68.876	-53.758	-42.370

Eleven reassignments were made according to −ln likelihood scores in Table 36. Two Chilean isolates, CHILE22 and CHILEWALL were reassigned from the Argentinean population to that in Cochabamba. One isolate from North Eastern Brazil, XE5809, and seven Venezuelan domestic isolates were reassigned to the Venezuelan silvatic population. Finally one Colombian isolate, 458, was reassigned from Venezuelan silvatic to North Eastern Brazil. With the exception of XE5809, all these reassignments are concurrent with the branching order of the isolates in question in Figure 55, and as such XE5809 was retained in the North Eastern Brazilian population during later analyses
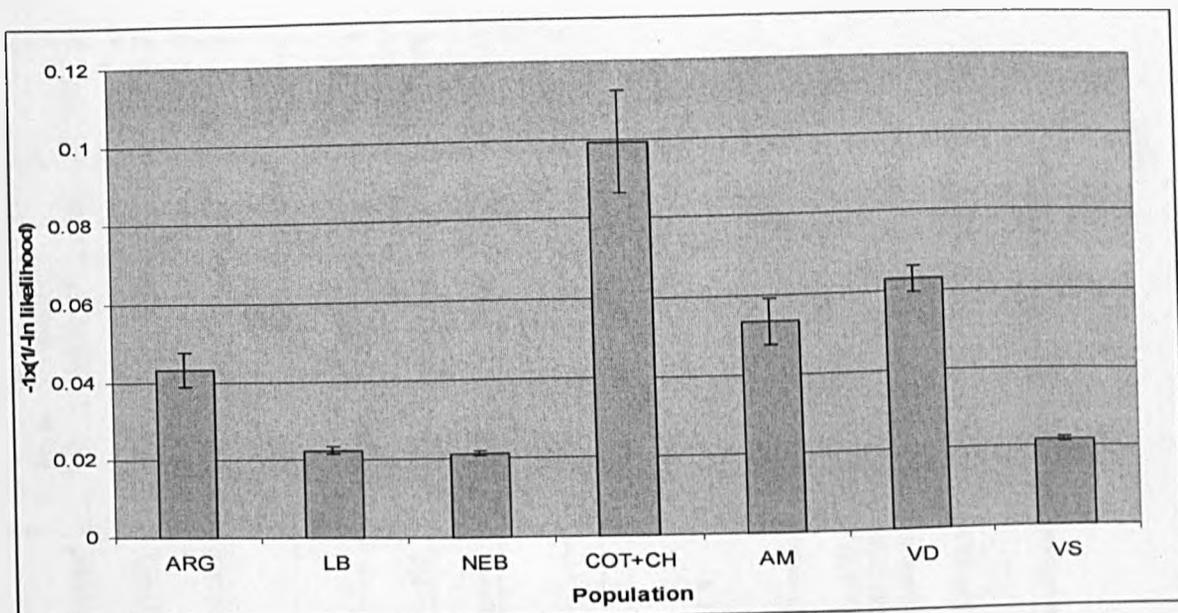
Figure 57 Transformed (-1x (1/-ln likelihood) mean –ln likelihood assignment scores for seven reassigned TCI populations. Error bars represent standard error about the mean.

As a measure of intrapopulation cohesion, the transformed mean of all –ln likelihood scores for each population was calculated once individual isolates had been reassigned (Figure 57). These scores broadly reflect the level of clustering observed clustering in Figure 55. Population cohesion is apparently highest among isolates from COT+CH as well as VD. Lowest mean scores are obtained from silvatic populations in LB, NEB and VS.

5.2.2.1.8 Population parameters

Several population parameters were calculated as in Section 3.7.6.1.4 to asses the genetic characteristics of the seven TCI sub-populations post statistical assignment (Table 37). Calculations were made across all loci, but with suitable Bonferroni correction for type I errors, where appropriate.

Table 37 Table of key genetic parameters for seven TCI populations, post statistical assignment.

Population Name	N/G	PL	PA/S	GD	Ho^a	HE^a	P values^b		% HD^c	%HE^d	F_{IS}	P value^e	% PL in LD^f	% LD physical^g	I_A^h	P value^h
							D	E								
Argentina (Arg)	10/10	27	0.8	0.38	0.535	0.551	0.02	0.98	8.82	2.94	0.035	0.3849	100	n/a	12.37	<0.001
Cochabamba + Chile (COT+CH)	11/11	20	0	0.21	0.406	0.396	0.0256	0.975	3.85	7.69	-0.168	0.8928	100	n/a	2.05	<0.001
Lowland Bolivia (LB)	16/16	32	0.273	0.52	0.467	0.643	0	1	17.50	0.00	0.238	0	21.2	8.2	3.98	<0.001
North Eastern Brazil (NEB)	39/39	32	0.438	0.55	0.383	0.571	0	1	50.00	0.00	0.273	0	0	n/a	2.03	0.287
North/Central America (AM)	7/7	24	0	0.26	0.332	0.445	0	1	0.00	0.00	0.204	0.0405	100	n/a	2.39	0.005
Venezuela Domestic (VD)	13/13	12	0.154	0.24	0.421	0.422	0.0316	0.968	7.14	14.29	-0.297	0.9902	100	n/a	1.21	0.011
Venezuela Silvatic (VS)	37/37	28	1.182	0.55	0.449	0.637	0	1	44.19	0.00	0.247	0	29.8	7.2	1.38	<0.001
Total	133	133											26.1	11.9	2.22	<0.001

N = Number of isolates in population.

G = Number of multilocus genotypes per population.

PL= Number of polymorphic loci.

PA/S = Multiple (3+) alleles per sample, calculated as a proportion of un-cloned isolates. See Section 5.2.2.1.3.

GD – Nei's unbiased gene diversity. Calculated in Microsatellite Toolkit (294)

[a]Mean observed and expected heterozygosity across all loci, calculated using ARLEQUIN v3.1(295)

[b]Calculated in GENEPOP v1.2, using a Monte Carlo algorithm approximation of an exact test, under the assumption of statistical independence of loci. D-deficit, E - excess (See Section 3.7.6.1.4)(297).

[c]Proportion of loci showing a significant deficit in heterozygosity after a Bonferroni correction. Calculated in ARLEQUIN v3.1(295).

[d]Proportion of loci showing significant excess heterozygosity after a Bonferroni correction. Calculated in ARLEQUIN v3.1(295).

[e]Calculated in ARLEQUIN v3.1, as P (Random F_{IS}>=Observed F_{IS}), where a null distribution of population specific F_{IS} values (Random F_{IS}) was generated over 16000 permutations (see Section 3.7.6.1.4) (295).

[f]Calculated in Multilocus v1.3 as the percentage of polymorphic locus pairs in significant linkage disequilibrium (r^A – related to the correlation coefficient r) by comparison to a null distribution of 1000 randomisations drawn from the same dataset (300).

[g]Calculated as the proportion of statistically linked polymorphic loci physically linked on the same contig.

[h]Calculated in Multilocus v1.3 by comparison to a null distribution of 1000 randomizations drawn from the same dataset (see Section 3.7.6.1.4) (300)

Figure 58 Histogram of three genetic parameters across seven TCI subpopulations. Populations associated with the domestic cycle are genetically depauperate and show increased relative levels of observed heterozygosity with respect to exclusively silvatic populations. Open bars – GD, dark grey - H_O, pale grey - H_E. Error bars represent +/- Standard error about the mean across loci (H_O & H_E) and +/- inter-locus standard deviation (GD).

The most genetically diverse populations studied where those from LB, NEB and VS. All three demonstrate the greatest gene diversity (0.52-0.55). Far less diversity is present in ARG, AM, COT+CH and VD (GD=0.21-0.38). A strongly significant overall deficiency in heterozygosity was observed across all loci (global test) in the majority of populations (4/7), by comparison to HW expectations. In two populations, ARG and VD, however, this deficit was marginal, and in a third, COT, a slight excess was observed. No statistical significance could be attached to these phenomena, however. At the level of individual loci, three of the four populations identified with strongly significant heterozygous deficit at a global level (VS, LB, and NEB) showed further signs significant deficit after a Bonferroni correction (See Section 3.7.6.1.4). No loci showed significant deficit in AM after a Bonferroni correction, a likely result of low sample size (N=7). Populations with marginal global p-values (p~0.05) for heterozygous deficit (ARG, VD, & COT) were ambiguous at the level of individual loci. In COT and VD slightly more of loci showed heterozygous excess than deficit while the inverse is true for ARG. COT and VD showed correspondingly negative values for the inbreeding coefficient, F_{IS} and these are indicative of excess heterozygosity in the populations they describe. Associated p-values imply the significance of this observation in VD (p>0.95), but not in COT (p<0.95) (See Section

3.7.6.1.4). Significantly positive F_{IS} values (p<0.05), correlating with excess homozygosity, were observed in LB, NEB, VS and AM. Intriguingly the observed F_{IS} derived from ARG was not significantly different from zero, the expectation of a randomly mating population conforming to HW expectations, although I_A values indicate that recombination is unlikely to be a phenomenon of this population (See below).

Meaningful conclusions can only be drawn from these observations if populations and associated statistics are considered across the entire dataset. Figure 58 describes three key parameters across all seven populations. LB, NEB, and VS all demonstrate a similar profile, whereby significant excess homozygosity correlates with high overall gene diversity. By comparison VD and COT demonstrate greater relative heterozygosity and far lower genetic diversity. AM and ARG represent intermediates between these two extremes. AM, while exhibiting relatively low genetic diversity (GD), demonstrates global homozygosity excess in line with LB, NEB and VS. The opposite is true for ARG, where genetic diversity is fairly high but relative heterozygosity levels are analogous to those in VD and COT.

Indices of association (I_A) provide evidence of multilocus allelic linkage disequilibrium in all populations (p<0.05). Pairwise tests for significant linkage between polymorphic loci within populations, measured by r^d (See Section 3.7.6.1.4), vary from 0% in NEB to 100% in VD, AM, ARG, and COT. As above, populations are essentially split into two groups by this measure. Loci in NEB, VS, and LB demonstrate apparently low levels of pairwise linkage (0-29.8%) while the remaining populations all demonstrate high levels (100%).

Linkage measures could point to recombination as a possible source of variation in three populations. I_A is a highly conservative measure of LD, and the extent of any possible departure from HW equilibrium cannot be inferred from size of its value (Section 3.7.6.1.4)(252). Thus I_A may fail to detect recombination if it occurs at low levels within the dataset. Low % significant pairwise r^d values in LB and VS and NEB, which showed no evidence of recombination under the I_A, therefore, raise the possibility of recombination within these populations.

Before assuming genetic exchange as the likely source of non-significant departures from linkage equilibrium between loci in certain populations, the phenomenon of multi-clonality must be addressed. If a number of microsatellite profiles analysed in this study are drawn, not from single clones, but from multiclonal populations of parasite, then this phenomenon could easily mimic the effect of recombination. Multi-allelic loci were observed in the dataset, and may be the result of multi-clonality (See Chapter 6 for detailed analysis). If the number of multi-allelic loci per sample (Table 37) is taken as an approximate measure of multi-clonality in populations (n.b. multi-clonality can also be present in the form of 'hidden homozygotes', see Section 1.8.1.5, Chapter 1), then all populations within which recombination is suspected are likely to contain multi-clonal samples. Nevertheless, two further populations which show strong evidence of inter-allelic association, VD and ARG, also contain multi-allelic loci, thus potential multi-clonality may not always act to break up multilocus linkage. However, in VS and LB, no relationship between physical linkage and statistical linkage is observed. Briefly, physically linked loci make up ~9.5% of all the possible pair-wise comparisons between loci in the dataset. On the assumption that recombination between linkage groups is more likely to occur than within, one might expect the over-representation of physically linked locus pairs among the total number demonstrating statistical linkage (i.e.>>9.5% of all statistically linked pairs). Instead LB and VS show values of 8.3% and 7.2% respectively, while the overall proportion across all populations is 11.9%, all of which are not much greater than the value expected by chance and thus possibly incompatible with recombination.

5.2.2.1.9 Pairwise population gene flow

Pairwise comparisons on inter-population differentiation (subdivision) were made using the fixation index F_{ST}, as described in Section 3.7.6.1.5 and using ARLEQUIN v3.1 (Table 38)(295). A further statistic, R_{ST}, based on assumptions of stepwise mutation, was also calculated as in Section 3.7.6.1.5 using MICROSAT v1.2 (Table 39) (290). Population groupings of samples are based on those defined by −ln assignment values generated in Section 5.2.2.1.7.

Table 38 F_{ST} **estimates of inter-population differentiation for seven TCI subpopulations based on microsatellite data.** Italics indicate p-values generated from 1000 random permutations leading to a value larger than or equal to that observed. Constructed in ARLEQUIN v3.1(295). All values remain significant after Bonferroni correction.

	ARG	COT+CH	LB	NEB	AM	VD	VS
ARG	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
COT+CH	0.34316	*	0.00000	0.00000	0.00000	0.00000	0.00000
LB	0.20769	0.30428	*	0.00000	0.00000	0.00000	0.00000
NEB	0.20561	0.31917	0.1446	*	0.00000	0.00000	0.00000
AM	0.48992	0.6504	0.34058	0.27871	*	0.00000	0.00000
VD	0.57066	0.71162	0.4111	0.35685	0.25116	*	0.00000
VS	0.22685	0.33062	0.14811	0.10808	0.23938	0.29522	*

F_{ST} values varied considerably between population pairs. High levels of gene-flow ($F_{ST}<0.2$) are observed between VS, LB & NEB (301). Moderate population subdivision ($0.2<F_{ST}>0.33$) is observed between a number of other population pairs, notably LB-ARG, VD-VS & LB-COT+CH. Complete isolation ($F_{ST}>0.33$) is observed in other pairs, especially VD-COT+CH & AM-COT+CH. Much of the most pronounced subdivision occurs between geographically distant populations. Two exceptional cases are, however, worth highlighting. The first is the relatively low level of gene-flow between geographically adjacent populations LB and COT+CH ($F_{ST}=0.30428$), compared with much higher gene flow between LB and more distant populations (e.g. LB-VS $F_{ST}=0.14811$). The second is the moderate subdivision between VD and VS ($F_{ST}=0.29522$), which share a broadly contiguous range (Figure 54), compared to the less pronounced population subdivision between VS and other populations which are located thousands of kilometres away (e.g ARG-VS, $F_{ST}=0.22685$)

Inter-population relationships were used to construct an un-rooted UPGMA tree (Figure 59). Tree topology is similar to that observed using D_{AS} values (Figure 55). The clustering of the silvatic clades NEB, VS and LB, however, is more apparent in this analysis.
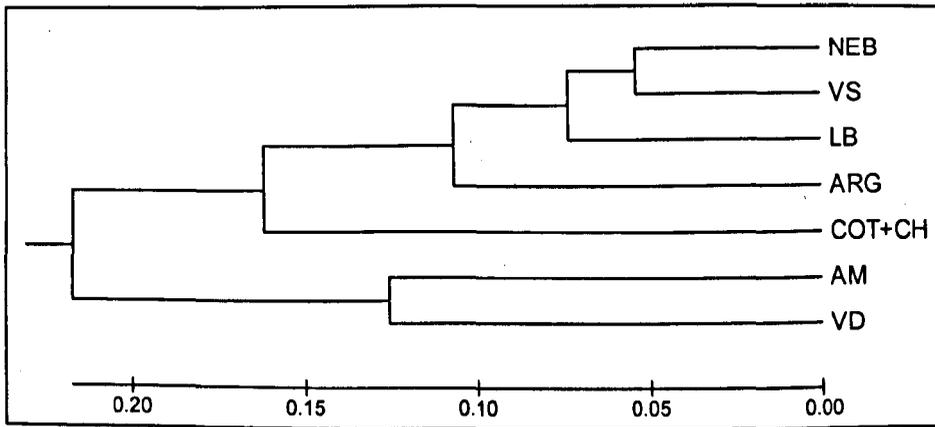
Figure 59 Un-rooted UPGMA tree based on pairwise F_{ST} values between TCI sub-populations constructed using MEGA v4(377).

By contrast to those observed between major *T. cruzi* sub-lineages (Section 5.1.2.1.3), R_{ST} values generated for TCI inter-subpopulation differentiation followed approximately the same pattern as F_{ST} values for the same dataset (Table 39). A linear regression, calculated in the form of a Mantel's test, similar to that described in Section 3.7.6.1.3, showed a strong positive correlation between the two measures (R_{XY}=0.954, P=0.001, R^2=0.9094).

Table 39 Pairwise matrix of Slatkin's R_{ST} values between seven TCI subpopulations based on microsatellite data Values in italics indicate the standard error of the mean over 1000 bootstraps. Constructed in MICROSAT v1.2 (290).

	ARG	COT+CH	LB	NEB	AM	VD	VS
ARG	*	*0.084*	*0.024*	*0.026*	*0.055*	*0.05*	*0.027*
COT+CH	0.25	*	*0.043*	*0.032*	*0.132*	*0.14*	*0.042*
LB	0.078	0.183	*	*0.017*	*0.053*	*0.073*	*0.015*
NEB	0.087	0.196	0.061	*	*0.055*	*0.076*	*0.021*
AM	0.323	0.432	0.213	0.192	*	*0.05*	*0.047*
VD	0.38	0.463	0.293	0.295	0.07	*	*0.075*
VS	0.079	0.134	0.056	0.084	0.141	0.231	*

While the overall level of subdivision estimated is lower (the greatest level of subdivision estimated – R_{ST}= 0.463 – occurs between VD and COT+CH), there is far greater variation between different pairwise values. Population differentiation between NEB, VS and LB, for example, are almost a factor of ten lower than those between more subdivided populations (e.g. COT+CH-LB).

The respective success and failure of R_{ST} to accurately reflect population substructure in this dataset and that in Section 5.1.2.1.3 (inter-lineage) could be accounted for by the

210

relative divergence between the populations involved. The stepwise model of microsatellite mutation is intrinsically more susceptible to the confounding effects of saturation than the IAM (see Section 1.9.4), a phenomenon which is likely to play a greater role when assessing genetic differentiation between major *T. cruzi* lineages, if they are as ancient as molecular evidence suggests, than between populations from the same lineage (see Section 1.8.2).

5.2.2.1.10 Tests for population substructure

A parallel test for population substructure was implemented using STRUCTURE v2.1 (see Section 3.7.6.1.7) (306). The aim of this analysis was to test the validity of earlier *a priori* and statistical population assignment as described in Sections 5.2.2.1.6 & 5.2.2.1.7. Estimation of the appropriate population number (K) was made by reference to a Gaussian curve of transformed –ln likelihood (-1x (1/-ln Pr (K/X))) (Figure 61). The number of populations, K, was allowed to vary between 2 and 13, with 10 replicates per value for K, a burn-in of 100,000, and 500,000 additional iterations of the Markov chain algorithm (See Section 3.7.6.1.7). Transformed –ln likelihood was observed to plateau at a value of $K=9$ (Figure 61). Confirmation for this observation was found by calculating ΔK, an extension of the plateau method that accounts for variance in –ln likliehood values for each value of K and the rate of change between the mean value for each value of K. Similar data were obtained to the plateau method. The graphical output of the optimal value for K is shown in Figure 60.

Figure 60 Graphical output of Structure v2.1 for *K*=9 (306). The run is divided into *K* colours. Vertical bars represent individual isolates. Population order is not necessarily indicative of relatedness. Admixture (bars made up of two or more colours) indicates where isolates may have split affinities between groups.
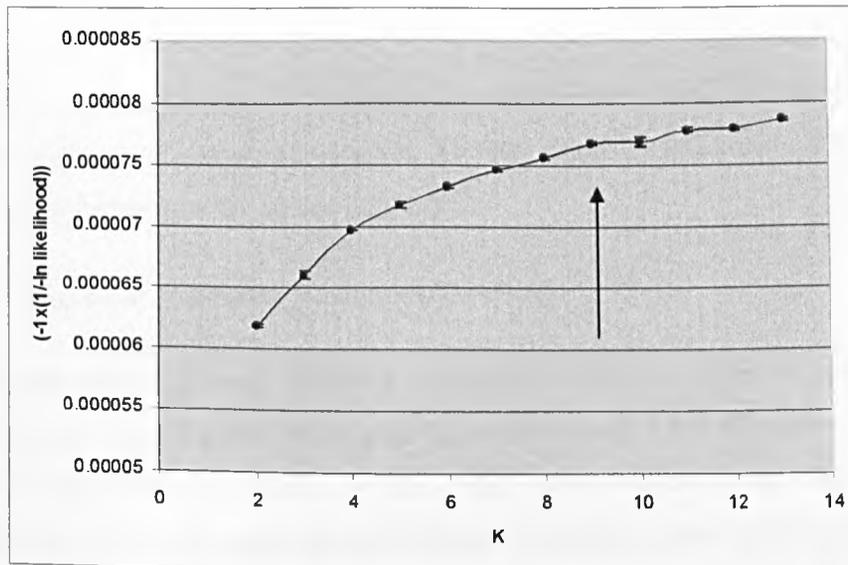
The identities of populations defined by STRUCTURE at the optimal value for K (K=9) show some interesting similarities, as well as differences, by reference to other methodologies outlined previously. Population number was potentially overestimated using other techniques in the case of AM and VD isolates, which show strong kinship as a single population here (Figure 60). By contrast, LB, VS, NEB, and ARG fragment into sub-populations. Perhaps unsurprisingly, all four of these populations showed relatively low population cohesion (<0.05) under statistical population assignment (Figure 57). In most instances (NEB1, NEB2, NEB3, VS1, VS2, & LB1 (Figure 60)) these population fragments are comprised uniquely of isolates from a broader previous assignment (i.e. NEB, VS & LB). ARG, however, appears to have split affinities, with a minority of isolates (n=4) clustering alongside LB isolates (ARG1/LB2), and the remaining six alongside Andean isolates from Cochabamba and Chile (ARG2/COT+CH). This phenomenon is broadly compatible with the clustering of the same isolates in the D_{AS} neighbour-joining tree in Figure 55. Other population assignments within diverse silvatic populations, specifically NEB, seem less compatible with Figure 55. Indeed, on the basis of D_{AS} values, no obvious clustering is observed in NEB, and these STRUCTURE-based population assignments may

represent a level of error due to non-conformance to HW assumptions (see Section 3.7.6.1.7).

At the level of individual isolates, ten of the eleven assignments made in Section 5.2.2.1.7 are supported here. In this analysis, however, XE5809 retains its affinities to NEB (NEB1), reflecting its positioning on the tree in Figure 55.

5.2.2.1.11 Analysis of Molecular Variance (AMOVA)

Three separate AMOVAs were undertaken in ARLEQUIN v.3.1, to test the proportion of the total variance accounted for by the population subdivision inferred in previous sections (295). In all cases, the vast majority (>72%) of all variation occurs at an intra-population level (Table 40). This observation fits with the deep branching within most clades in Figure 55, and correspondingly high within-population MNA values in most populations in Table 37.

Table 40 Analysis of molecular variance (AMOVA) undertaken for TCI subpopulations defined in previous sections. Calculated using ARLEQUIN v3.1(295)

Source of Variation	Degrees of Freedom	Sums of squares	Variance components[a]	% Variation	Fixation Indices[a]	p - values[b]
Grouped by geography & transmission						
Among populations	6	571.194	2.46205 Va	24.43	F_{ST}: 0.24435	0.00000
Within populations	259	1972.039	7.61405 Vb	75.57		
Total	265	2543.233	10.0761			
Post-statistical assignment						
Among populations	6	606.38	2.64751 Va	26.15	F_{ST}: 0.26146	0.00000
Within populations	259	1936.853	7.47820 Vb	73.85		
Total	265	2543.233	10.1257			
Grouped by Structure v2.1						
Among populations	8	688.003	2.79904 Va	27.95	F_{ST}: 0.27946	0.00000
Within populations	257	1854.711	7.21678 Vb	72.05		
Total	265	2542.714	10.01744			

[a] See Section 3.7.6.1.8 for a details of calculation.
[b] Calculated by comparison to a null distribution of 16000 random permutations.

Nonetheless, significant overall F_{ST} values, and, by implication, population subdivision does exist in all cases (p=<<<0.05). An improvement in the proportion of total variation explained by inter-population variation (Va) across different methodologies is also observed. *A priori* population assignment, statistical assignment, and assignment using

STRUCTURE v2.1 account for 24.43%, 26.15%, and 27.95% of total variance respectively.

5.2.2.1.12 Tests for population bottlenecks

Allelic diversity and distribution among populations and between individuals in the same population may be used to determine if a severe and rapid reduction in the effective population size (N_e) has occurred in the recent past of certain groups of individuals. Details of these related tests and their underlying assumptions are described fully in Section 3.7.6.1.9. Three populations were identified where such a phenomenon could have occurred: AM, on the basis of circumstantial biogeographical evidence (See Section 1.6.2.1), as well as COT+CH and VD, one the basis of reduced genetic diversity across a wide geographic area (Figure 56, Table 37). Two descriptive statistics were calculated, the Garza-Williamson (G-W) Index, in ARLEQUIN 3.1 (Figure 62), and allele frequency distribution as in Luikart *et al.*, (1998) (Figure 63) (295, 315, 316).
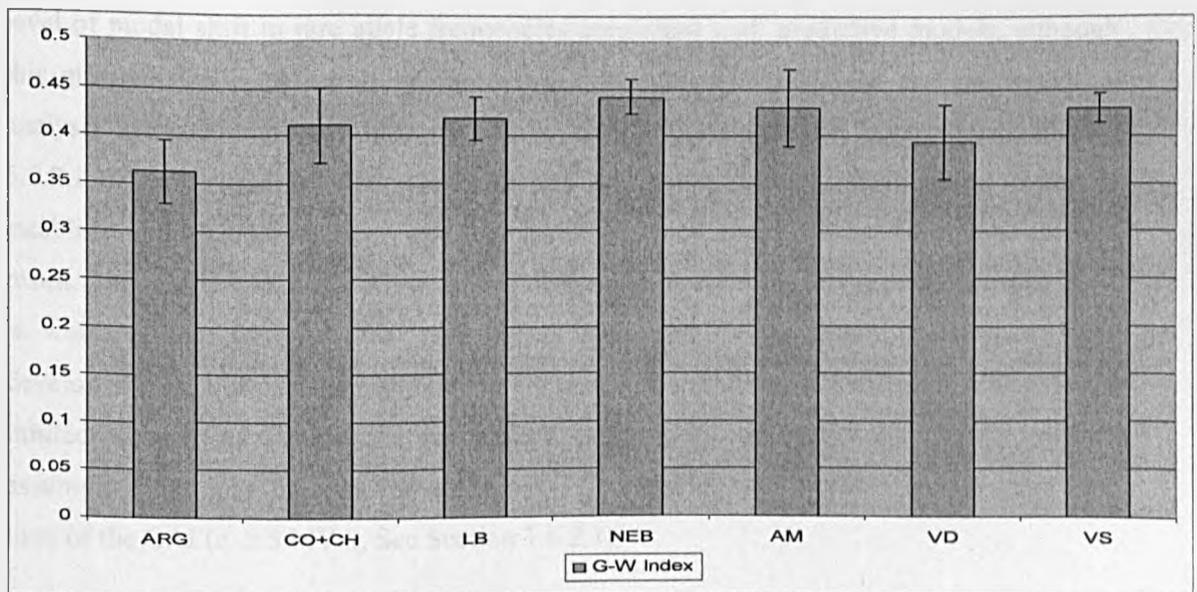
Figure 62 The Garza-Williamson Index calculated for seven TCI subpopulations in ARLEQUIN v3.1(295). Error bars represent +/- standard error about the mean of all polymorphic loci in population.

The G-W index, or M ratio, provides a measure of the drop in allelic diversity against the allelic size range (number of repeats) in a population (See Section 3.7.6.1.9), and should be smaller in populations that have experienced a recent bottleneck than in those for whom N_e has been stationary for a long period of time. While no critical value for M (M_c) was

derived, and thus no test applied, no specific pattern can be observed between suspected bottleneck populations (AM, VD, and COT+CH) and others, aside from a marginal increase in the associated error about the mean, probably linked to low sample size.

Rare allele frequency plots (Figure 63) provide a perhaps more interesting comparison between populations. The three most genetically diverse populations, VS, NEB and LB, all demonstrate allele frequencies consistent with a population at mutation drift equilibrium under an infinite allele model, conceivably indicating long term stability of N_e, even though the associated assumption of random mating is likely to be contravened (Figure 25) (314, 315, 378). Other populations present distinct profiles, consistent with a loss of rare alleles (1-10%) with respect to other allele frequency classes within the population, as well as a greater relative number of alleles approaching fixation (91-100%). In VD and COT+CH this change in rare allele frequency is not entirely consistent with the modal shift as predicted in populations that have undergone a recent bottleneck (1-10% to 11-20% or 21-30%), but some degree of perturbation is observed (315). AM and ARG both do show a level of modal shift in rare allele frequencies consistent with predictive models, although this phenomenon is perhaps more pronounced in AM. STRUCTURE and pairwise D_{AS} analysis indicate that ARG may in fact be split into two distinct populations (Section 5.2.2.1.10), and intra-population substructure is known to be a confounding factor in establishing population bottlenecks (314). Little credence, therefore, can be attached to this result. The modal shift observed in AM is of interest, despite the low sample size (N=7). It is worth noting, however, that the models under which the modal shift theory was developed were designed to examine the signature of a population bottleneck after a fairly limited number of generations (dependent on severity of bottleneck event), another assumption that may be contravened here if TCI migrated into North America around the time of the GAI (c. 2.5 MYA, See Section 1.6.2.1)
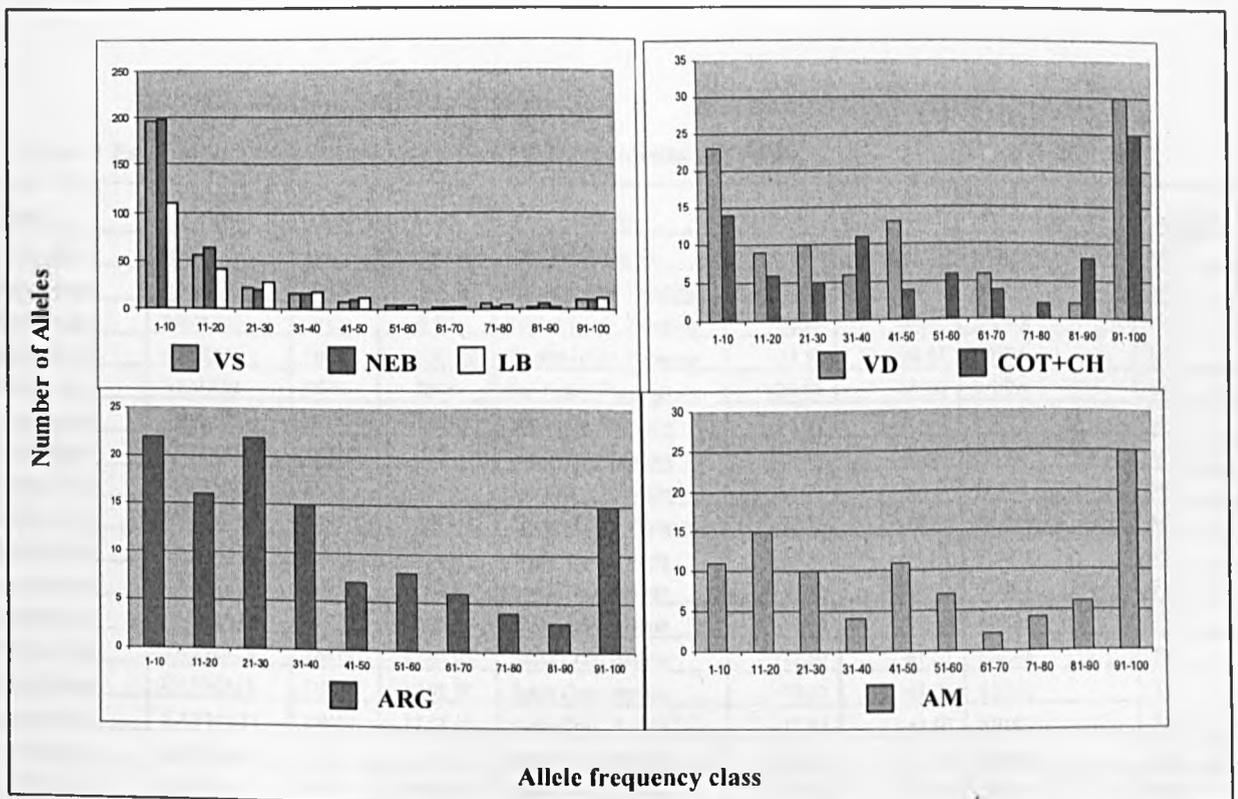
Figure 63 Rare allele frequency plots of for seven TCI populations. Frequency classes (X-axes) correspond to the frequency of an allele in a population (%). Number of alleles (Y-axes) corresponds to the number of alleles in a given frequency class present across all loci in each population.

There is no balance of evidence across the two descriptive statistics detailed here to suggest that any of the populations studied have undergone a recent population bottleneck. Sample size is clearly limiting in most cases. It has been proposed that the number of loci employed can, to an extent, compensate for low sample sizes, and the high number of loci employed here do lend some weight to the results (314) Additionally, it is of note that LB resembles divesrse silvatic populations (NEB and VS) in terms of its allele frequency class distribution (Figure 63), despite containing only half as many samples (N=16).

5.2.2.2 TCIIc sublineage

5.2.2.2.1 Samples analysed

A panel of 53 TCIIc lineages was assembled and subjected to microsatellite analysis. Samples were drawn from at least six species of mammal, the vast majority from order Cingulata. A single triatomine isolate was present, M3-CU, from a *Panstrongylus* sp nymph. All available TCIIc samples were included, all of which were silvatic in origin, and their geographic distribution is shown in Figure 64.

216

Table 41 Panel of 53 TCIIc isolates assembled for microsatellite analysis.

Host	Original Code	Tree Code[e]	Date	Locality	Latitude[b]	Longitude[b]	Population 1[c]/2[d]	Source[e]
Dasypus novemcinctus	85/847	DN1	Unknown	Alto Beni, Bolivia	-15.50	-67.50	NB/NB	MT
Dasypus novemcinctus	ARMA12	DN2	2001	Campo Lorro, Paraguay	-22.33	-58.93	PA/PA	MY
Dasypus novemcinctus	ARMA13	DN3	2001	Campo Lorro, Paraguay	-22.33	-58.93	PA/PA	MY
Dasypus novemcinctus	ARMA18	DN4	2001	Campo Lorro, Paraguay	-22.33	-58.93	PA/PA	MY
Dasypus novemcinctus	ARMA24	DN5	2001	San Pedro, Paraguay	-24.00	-57.00	PA/PA	MY
Dasypus novemcinctus	ARMA25	DN6	2001	San Pedro, Paraguay	-24.00	-57.00	PA/PA	MY
Dasypus novemcinctus	ARMA26	DN7	2001	San Pedro, Paraguay	-24.00	-57.00	PA/PA	MY
Dasypus novemcinctus	ARMA27	DN8	2001	San Pedro, Paraguay	-24.00	-57.00	PA/PA	MY
Dasypus novemcinctus	ARMA9	DN9	2001	Campo Lorro, Paraguay	-22.33	-58.93	PA/PA	MY
Dasypus novemcinctus	CAYMA13	DN11	17.01.05	Santa Cruz, Bolivia	-17.50	-61.50	SB/SB	TS
Dasypus novemcinctus	CAYMA14	DN12	14.01.05	Santa Cruz, Bolivia	-17.50	-61.50	SB/SB	TS
Dasypus novemcinctus	CAYMA17	DN13	17.01.05	Santa Cruz, Bolivia	-17.50	-61.50	SB/SB	TS
Dasypus novemcinctus	CAYMA18	DN14	17.01.05	Santa Cruz, Bolivia	-17.50	-61.50	SB/SB	TS
Dasypus novemcinctus	CAYMA19	DN15	17.01.05	Santa Cruz, Bolivia	-17.50	-61.50	SB/SB	TS
Dasypus novemeinctus	CAYMA21	DN10	17.01.05	Santa Cruz, Bolivia	-17.50	-61.50	SB/SB	TS
Euphractus sexcinctus	CAYMA25	ES2	17.01.05	Santa Cruz, Bolivia	-17.50	-61.50	SB/SB	TS
Euphractus sexcinctus	CAYMA3	ES1	06.01.05	Santa Cruz, Bolivia	-17.50	-61.50	SB/SB	TS
Dasypus sp	CM 17	DN16	1982	Carimaga, Colombia	3.30	-73.00	BVC/BVC	MT
Dasyprocta fugilinosa	CM 25	DF1	1982	Carimaga, Colombia	3.30	-73.00	BVC/BVC	MT
Panstrongylus sp	M3-CU	PS	03.05.04	Barinas, Venezuela	8.48	-70.73	BVC/BVC	TS
Dasypus novemcinctus	M10	DN17	17.05.04	Barinas, Venezuela	8.48	-70.73	BVC/BVC	TS
Dasypus novemcinctus	M5	DN18	09.05.04	Barinas, Venezuela	8.48	-70.73	BVC/BVC	TS
Dasypus novemcinctus	M5631	DN19	c.1980	Marajo. Brazil	-1.00	-49.50	BVC/BVC	LSHTM
Dasypus novemcinctus	M6	DN20	09.05.04	Barinas, Venezuela	8.48	-70.73	BVC/BVC	TS
Dasypus novemcinctus	M8	DN21	15.05.04	Barinas, Venezuela	7.50	-71.23	BVC/BVC	TS
Chaetophractus vellorosus	MA194	CV1	27.01.05	Campo Lorro, Paraguay	-22.33	-58.93	PA/PA	TS/NA
Chaetophractus vellorosus	MA202	CV2	04.08.05	Santa Cruz, Bolivia	-19.21	-63.43	SB/SB	TS/NA
Dasypus novemcinctus	MA204	DN22	05.08.05	Santa Cruz, Bolivia	-19.21	-63.43	SB/SB	TS/NA
Dasypus novemcinctus	MA215	DN23	12.08.05	Santa Cruz, Bolivia	-19.21	-63.43	SB/SB	TS/NA
Dasypus novemcinctus	MA219	DN24	13.08.05	Santa Cruz, Bolivia	-19.21	-63.43	SB/SB	TS/NA
Dasypus novemcinctus	MA220	DN25	13.08.05	Santa Cruz, Bolivia	-19.21	-63.43	SB/SB	TS/NA
Euphractus sexcinctus	MA222	ES3	13.08.05	Santa Cruz, Bolivia	-19.21	-63.43	SB/SB	TS/NA
Euphractus sexcinctus	MA25*	ES4	2003	Campo Lorro, Paraguay	-22.33	-58.93	PA/PA	MY
Euphractus sexcinctus	MA25X*	ES5	2003	Campo Lorro, Paraguay	-22.33	-58.93	PA/PA	MY
Dasypus novemcinctus	MA87	DN26	2003	Campo Lorro, Paraguay	-22.33	-58.93	PA/PA	MY
Dasypus novemcinctus	PARAMA26	DN27	06.08.05	Barinas, Venezuela	8.43	-70.55	BVC/BVC	TS
Dasypus novemcinctus	PARAMA25	DN28	06.08.05	Barinas, Venezuela	8.43	-70.55	BVC/BVC	TS
Dasypus novemcinctus	PARAMA34	DN29	06.08.05	Barinas, Venezuela	8.43	-70.55	BVC/BVC	TS
Dasypus novemcinctus	PARAMA6	DN30	06.08.05	Barinas, Venezuela	8.43	-70.55	BVC/BVC	TS
Dasypus novemcinctus	SAM6	DN31	21.11.04	Santa Cruz, Bolivia	-20.02	-63.02	SB/SB	TS
Dasypus novemcinctus	SJMC10	DN32	13.09.04	Beni, Bolivia	-14.81	-64.6	NB/NB	TS
Dasypus novemcinctus	SJMC4	DN34	09.09.04	Beni, Bolivia	-14.81	-64.6	NB/NB	TS
Dasypus novemcinctus	SJMC19	DN33	09.09.04	Beni, Bolivia	-14.81	-64.6	NB/BVC	TS
Dasypus novemcinctus	SP13	DN40	2003	San Pedro, Paraguay	-24.00	-57.00	PA/PA	MY
Dasypus novemcinctus	SP14	DN41	2003	San Pedro, Paraguay	-24.00	-57.00	PA/PA	MY
Dasypus novemcinctus	SP15	DN42	2003	San Pedro, Paraguay	-24.00	-57.00	PA/PA	MY

217

Table 41 Continued from overleaf

Dasypus novemcinctus	SP16	DN43	2003	San Pedro, Paraguay	-24.00	-57.00	PA/PA	MY
Monodelphis domestica	SP4	MD	2003	San Pedro, Paraguay	-24.00	-57.00	PA/PA	MY
Dasypus novemcinctus	**SJMO18**	DN35	21.06.07	Beni, Bolivia	-15.12	-64.32	NB/NB	TS
Dasypus novemcinctus	**SMA18**	DN36	09.06.07	Beni, Bolivia	-14.13	-65.36	NB/NB	TS
Dasypus novemcinctus	**SMA19**	DN37	09.06.07	Beni, Bolivia	-14.13	-65.36	NB/NB	TS
Dasypus novemcinctus	**SMA8XE**	DN38	08.06.07	Beni, Bolivia	-14.13	-65.36	NB/NB	TS
Dasypus novemcinctus	**SMA9**	DN39	08.06.07	Beni, Bolivia	-14.13	-65.36	NB/NB	TS

Samples in bold were cloned as in Section 3.4.2.
[a] See neighbour-joining tree in Figure 65.
[b] Decimal degrees
[c] *A priori* population assignment (See Section 5.2.2.2.6).
[d] Statistical population assignment.
[e] As in Table 34, apart from **TS/NA** – Isolated during this study in collaboration with Nidia Acosta (LSHTM).
*Isolated by MY from the same mammal using two different techniques (See Section 3.2.4) – MA25: direct inoculation from blood; MA25X: xenodiagnostic isolation.
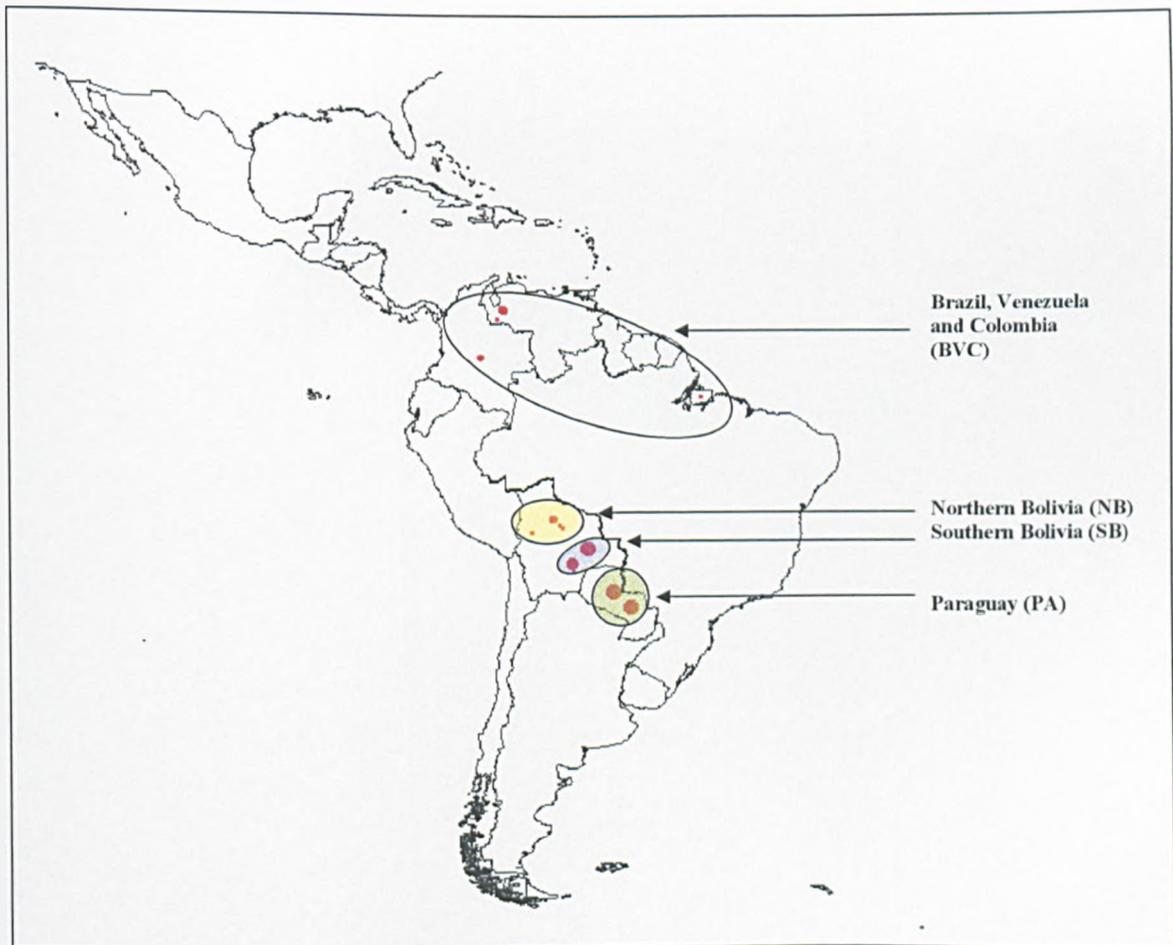
Figure 64 Geographic distribution of TCIIc isolates analysed in this study. Red points indicate sample location and point size is proportional to sample density. Coloured ellipses represent *a priori* population assignment, as detailed in Section 5.2.2.2.6. Population codes correspond to Population 1, in Table 41.

218

5.2.2.2.2 Loci employed in analysis

The panel of microsatellite loci employed to analyse the TCIIc isolates is shown in Table 42. These loci correspond, in part, to those employed in the analysis of TCI (Table 42). Forty-nine variable loci were identified as detailed in Section 3.7.2, while two loci originate from an earlier study (215). As detailed in Section 5.2.2.1.2, monomorphic loci and those that amplified poorly across all strains were excluded.

Table 42 Microsatellite loci employed to study TCIIc intra-lineage diversity (continued overleaf).

Primer Code	Contig ID[a]	Position[b]	Repeat[c]	Forward Primer / Reverse Primer	Primer Code	Contig ID[a]	Position[b]	Repeat[c]	Forward Primer / Reverse Primer
1TETP2b	6986	75669..75701	$(TA)_n$	TGAAGGAGAATTCTCTGCGGT / CTCTCATCTTTTGTTGTGTCCG	vicP1a	8305	987..1016	$(TAA)_n$	CCGCGGTAGAAGAACCATAA / TGCGTATTCACGACGAGAAG
1TETP10b	6996	94129-94191	$(CAA)_n$	ATTCGGCCTGTTTGTATTCG / TTGGTTGTGTTTTACTGCTGC	vicp1b	8305	46598..46638	$(TA)_n$	AACCCGCGCAGATACATTAG / TTCATTTGCAGCAACACACA
1TETP11a	6996	16332-16526	$(GA)_n$	GCAAAGCAACAAAAACAGC / GAGCACACGAAGGGGAATAA	tetp3a	8305	49478..49495	$(TC)_n$	GTACGACGTGGACACAAAC / ACAAGTGGGTGAGCCAAAAG
1TETP5a	6996	42683..42715	$(TC)_n$	CGCTCTCAAAGGCACCTTAC / ATATGGACGCGTAGGAGTGC	netp6a	8305	88534..88549	$(CA)_n$	GTCGCCATCATGTACAAACG / CTGTTGGCGAATGGTCATAA
1TETP6a	6996	28781-28980	$(TG)_n$	TTGCTTGCTTCCACGTGTA / TATTCCCTTTGCCTTTGCTG	vicp6b	8305	87584..87606	$(CA)_n$	GTGTCGTTGCTCCCAAACTC / AAACTTGCCAAATGTGAGGG
famp2a	7118	76285-76410	$(TC)_n$	GACATGTATGCTTGAAACCTCC / TCCATCTCCCTTCACACTCC	tetp7a	8305	15729..15742	$(CA)_n$	ACCCAGAGGGGGAGAAAAAGA / TTTACGGTTGGTTCGTGTGA
nedp4a	7118	10052..10134	$(TTA)_n(GTT)_n$	GAGGTGATGACGAATAAAATTGG / GTCTTTCCGCATATCCGAGA	tetp1a	8305	987..1016	$(TAA)_n$	CCGCGGTAGAAGAACCATAA / TGCGTATTCACGACGAGAAG
vicp7a	7118	12867..12936	$(TG)_n$	TTGCGTGGTTGTTTGTGC / GAGAAGAGGGGGAGGAAGAA	1TETP1b	8328	63326-63566	$(TG)_n$	AAGAGAGGCACTCCCTGTGA / GAGGAAGAGAGGAAGTACAGTTGAGC
1TETP12a	7118	5369-5576	$(TA)_n(GA)_n$	TGTGATCAACGCGCATAAAT / TTCCATTGCCTCGTTTTAGA	1FAMP12a	8328	124985-125219	$(TG)_n$	GCGCGTTATTAACACTCGCT / GCCCGGTATCATTGAAAAGA
vicp2a	7143	88658-88832	$(CT)_n$	CATCAAGGAAAAACGGAGGA / CGGTACCACCTCAAGGAAAG	1TETP3a	8328	44002..44057	$(TA)_n$	AGAAAAAGGTTTACAACGAGCG / CGATGGAAACGTGAAACAA
famp6a	7143	119499..119520	$(TG)_n$	TCGTTCTCTTTACGCTTGCA / TAGCAGCACCAAACAAAACG	1TETP4a	8328	71097-71226	$(GA)_n$	GTCACCACCTAGCGATGACA / ACTGCACAATACCCCCTTTG
nedp8a	7143	48024..48053	$(TG)_n$	GAAACGCACTCACCCACAC / GGTAGCAACGCCAAAACTTTC	1TETP4b	8328	32430-32629	$(TTA)_n$	GAGAGAGATTCGGAAACTAATAGC / CATGTCCCTTCCTCCGTAAA
nedp5a	7206	69979..70016	$(TC)_n$	CCAACATTCAACAAGGGAAA / GCATGAATATTGCCGGATCT	1FAMP7b	8328	83338-83417	$(CA)_n$	CTACCTTCTCTTTCCTCCCTAACC / TTTGCTCTGGACTGCATGC
tetp8a	7206	16233..16257	$(TA)_n$	CGTGTGCACAGGAGAGAAAA / CGTTTGGAGGAGGATTGAGA	1TETPRa	8328	39618-39874	$(CA)_n(TA)_n$	CATGTCATTAAGTGGCCACG / GCACATGTTGGTTGTTGGAA
1TETP10a	7206	82234-82445	$(TAA)_n$	CCGCAGACATTTCTTCGACT / GCTTTTTGTCTTCTGCCGAC	famp1a	7098	64001-64129	$(TG)_n$	GAGCAGATCTTCCTTGTGCC / TGGTGAAATGCACGCATC
1TETP6b	7206	51225-51373	$(TA)_n$	GGAAACACATCACGCAAAGA / AGTTGACAAAGGGGGACATTG	vicp3a	7098	28533-28726	$(TC)_n$	GCCCCCATGCAATAATTTTTA / GTGGTGGGAGAACAACCAAC

220

Primer Code	Contig ID[a]	Position[b]	Repeat[c]	Forward Primer/ Reverse Primer
vicp2b	7853	94443..94463	(TA)n	TGTAACGGTAGGTCTCAATTCG / TTGCACTGTGTATCTCGCC
petp3a	7853	77851..77877	(TC)n	GCCATGTTCCTTCACCAAC / AGGTGTTCCCCTCTTTGGAT
vicp5a	7853	72728-72864	(CT)n(TA)n	GCAGAGACGCACAGACACAT / AAAGTGCCATCCCACCCTC
nedp7a	7853	47741..47761	(TG)n	AGAAGAGAGTCCGGAGTTTTC / TGTGCGGCCAAGAGTTCTAAA
1FAMP9b	7644	200539-200729	(CA)n	CTGTTGGTCCTTTCAGTGCA / AGTCCTACTGCCTCCTTGCA
FAMP11a	8370	18473-18619	(CA)nn(GA)n	ATTCATCGCCGACATTTTA / GCAATTGCTGCTTTCGGTAT
1FAMP14a	8370	15248-15407	(TTA)n	TCACGCTGCCGTACAAGTAG / GCGGTACACCAACATGTACG
famp5a	8646	58996-59147	(CA)n(CA)n	GTGTGTTTGTGTGTGAGAGGC / CCAGTATTCCTCCCCCCTAA
famp7a	8646	17391..17437	(TA)m	CGTCTTGTTGATTTCCCCTT / ACCACCAGGAGGACCATGAAG
famp8a	8646	88680..88708	(TCG)n	TGTACACGGAACAGCGAAG / AACATCCTCCACCTCACAGG
famp8b	8646	12639-12819	(TA)m	TTTGAATGCGAGGTGGTACA / AGTTGACATCCCAAGCAAG
vicp6a	8812	7458..7500	(CA)n	CCCTGATGCTGCAGACTCTT / GCGTAGCGAATTCATTTCC
famp1b*	Unknown	Unknown	(CA)nA(CA)n	ATCCGCTACCACTATCCAC / CCTCTGCGCACACATTCATT
vicp1c	Unknown	Unknown	Unknown	CCGTTCTTCATCACCATCCT / GATCCCGCAATAGGAAAC
famp2c*	Unknown	Unknown	(CT)nTG)n	GTGCATGTTCCATGGCTT / AACAAAATCTAGCGTCTACCATCC
1TETP15	Unknown	Unknown	Unknown	GGTGTTGGCGTGTATGATTG / ATGCTCCGCAACATATTACTCA
TETP2a	Unknown	Unknown	Unknown	GTCGAGCTTCTGTTGTTCCC

[a] [b] [c] As in Table 35.

* Alternative primer codes are famp1b = mclf10; famp2c = scle10 (215)

Loci of known site, by comparison to the online *T. cruzi* genome, are distributed across 13 syntenous regions (contigs) (376). The largest number of loci drawn from any single contig was seven, the smallest one. Where possible di- or tri-nucleotide repeat types were identified using the online database. Microsatellite repeat types for those loci used in the previous study are based on those in the literature (215).

5.2.2.2.3 Missing data, clone correction and multiple alleles.

As in Sections 3.7 and 5.2.2.1.3, after repeated lower temperature annealing temperature re-runs in an attempt amplify problematic loci in some samples, a level of residual missing data remained in the dataset. This value was 8.8%, equating to approximately 4.35 loci per sample. Missing data were evenly distributed between samples and loci and none were excluded from the analysis.

Clone correction was applied to the dataset in GENALEX v6.0, but no identical matching genotypes were identified (292).

The presence of multiple (3+) alleles at single loci was identified within the dataset, indicating the possible presence of multi-clonality. The dataset was screened for physically linked multi-allelic loci indicative of aneuploidy, and none were identified. As in Section 5.2.2.1.3, samples demonstrating multiple alleles were biologically cloned where possible. After cloning, six samples demonstrating multi-allelic loci remained, all of which were uncloned, and for which no live organism was available. The frequency of multiple alleles per sample per population is displayed in Table 44, while the overall proportion of multiple alleles with respect to the total number in the dataset was ~0.15%.

5.2.2.2.4 Distance measures

Pairwise distance between isolates was estimated under an IAM (D_{AS}) and a SMM ($\delta\mu^2$), as with TCI (Section 5.2.2.1.4). Similarly 1000 diploid re-samplings of the dataset were made to avoid excluding multi-allelic samples from the analysis, as in Section 3.7.6.1.1. The mean of the derived distance matrices forms the basis for the D_{AS} tree in Figure 65.
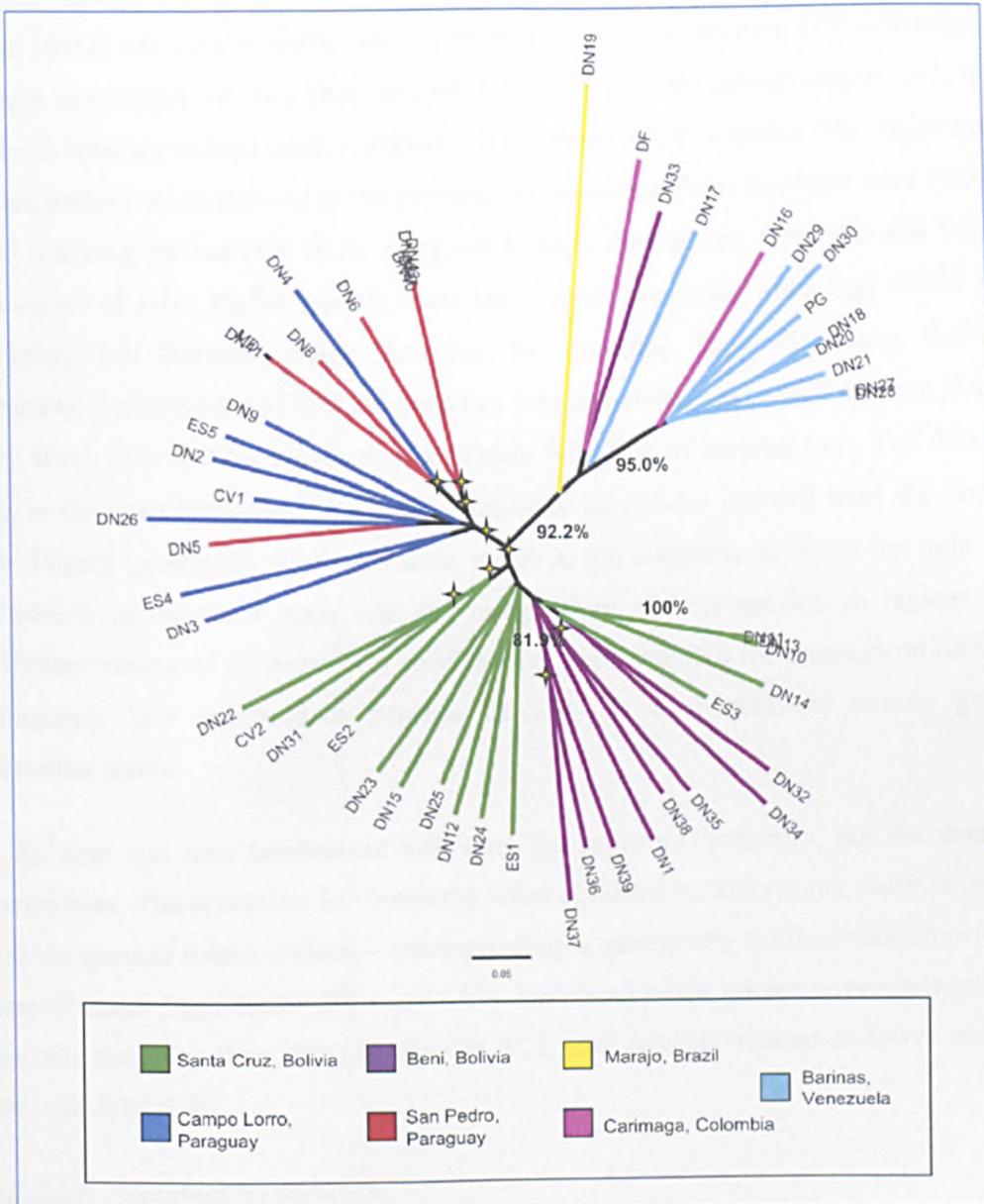
Figure 65 Un-rooted D_{AS} neighbour joining tree based on the multilocus microsatellite profiles of 53 TCIIc isolates. Values indicate D_{AS}-based bootstrap values over 10,000 trees on major clades >75%. Yellow stars indicate multi-allelic topology instability over 1000 trees (<90% recovery, see Section 3.7.6.1.1) Tip labels indicate host/vector origin, as in Table 41.

Intriguingly, despite a lower proportion of multi-allelic samples in the TCIIc dataset by comparison to the TCI dataset, the level of instability across random diploid re-samplings was higher, with a number of branches experiencing <90% stability. This suggests either a non-linear effect of multiple allele frequency on stability and/or that those loci that were multi-allelic in TCIIc were marginally more informative. Again, those clades experiencing

223

greatest instability correspond broadly to those populations in which multi-allelic samples occur (Table 44). D_{AS} bootstrap values, derived from the mean over 100 bootstraps of 100 diploid re-sampled datasets (See Section 3.7.6.1.2), provide strong support (>75%) for a division between isolates from Northern and Southern South America. The major exception to this pattern of divergence is the presence of an isolate from Northern Beni (SJMC19 – DN33) among the northern clade, alongside isolates from Brazil, Colombia and Venezuela. A number of other clades receive bootstrap support, including groupings within Bolivia, Paraguay and Northern South America. No bootstrap support between Bolivia and Paraguay is observed, and internal branches between clades comprised of these isolates are very short. One feature of this tree consistent with that of derived from TCI data (Figure 55), is the deep branching between the majority of isolates isolated from the same study site. Highly genetically similar isolates, while in the minority, do occur but only between isolates from the same study site and independent of host species. In regions where a sufficient variety of different host species have been sampled for comparison (Bolivia and Paraguay), host independent parasite diversity is also observed among genetically dissimilar stains.

A $\delta\mu^2$ tree was also constructed following the same methodology, but the data are not shown here. The maximum $\delta\mu^2$ bootstrap value obtained between major clusters was <20%, and the greatest within clusters – corresponding to genetically similar stains showing 100% support under D_{AS} (Figure 65), was 51.9%, indicating much lower correspondence between the two measures than that observed in TCI, and possibly related to lower sample size (Section 5.2.2.1.4)

5.2.2.2.5 Isolation by distance.

As in Sections 3.7.6.1.3 and 5.2.2.1.5, the relationship between pairwise values for D_{AS} and pairwise values for geographic distance were examined using a Mantel's test in GENALEX v6.0 with a random permutation procedure to test the statistical significance of any correlation (292). When all TCIIc isolates were considered (Figure 66, Plot 1), a significant and positive correlation between genetic and geographic distance was noted ($R_{XY} = 0.687$, p<<0.05), and a fairly good fit of the data to the regression line observed ($R^2 = 0.472$).

A second Mantel's test, involving only isolates derived from *D. novemcinctus*, also demonstrated a significant, but marginally lower, positive correlation between genetic and geographic distance (R_{XY} = 0.658, p<<0.05), with a correspondingly lower fit of the data about the regression line (R^2 =0.433).

As is clear from the physical map of sample localities for TCIIc (Figure 64), the geographic range from which isolates are drawn is considerable smaller than that of TCI (Figure 54). Nor is the sampling of TCIIc as even across this range, with the vast majority of isolates clustered in Bolivia and Paraguay. Nonetheless, greater isolation by distance is apparent between TCIIc isolates than between those from TCI (R_{XY} TCIIc = 0.687, R_{XY} TCI = 0.393), as well as when isolates from *D. novemcinctus* (TCIIc) and *D. marsupialis* (TCI) are considered separately (R_{XY}= 0.658 and 0.429 respectively).
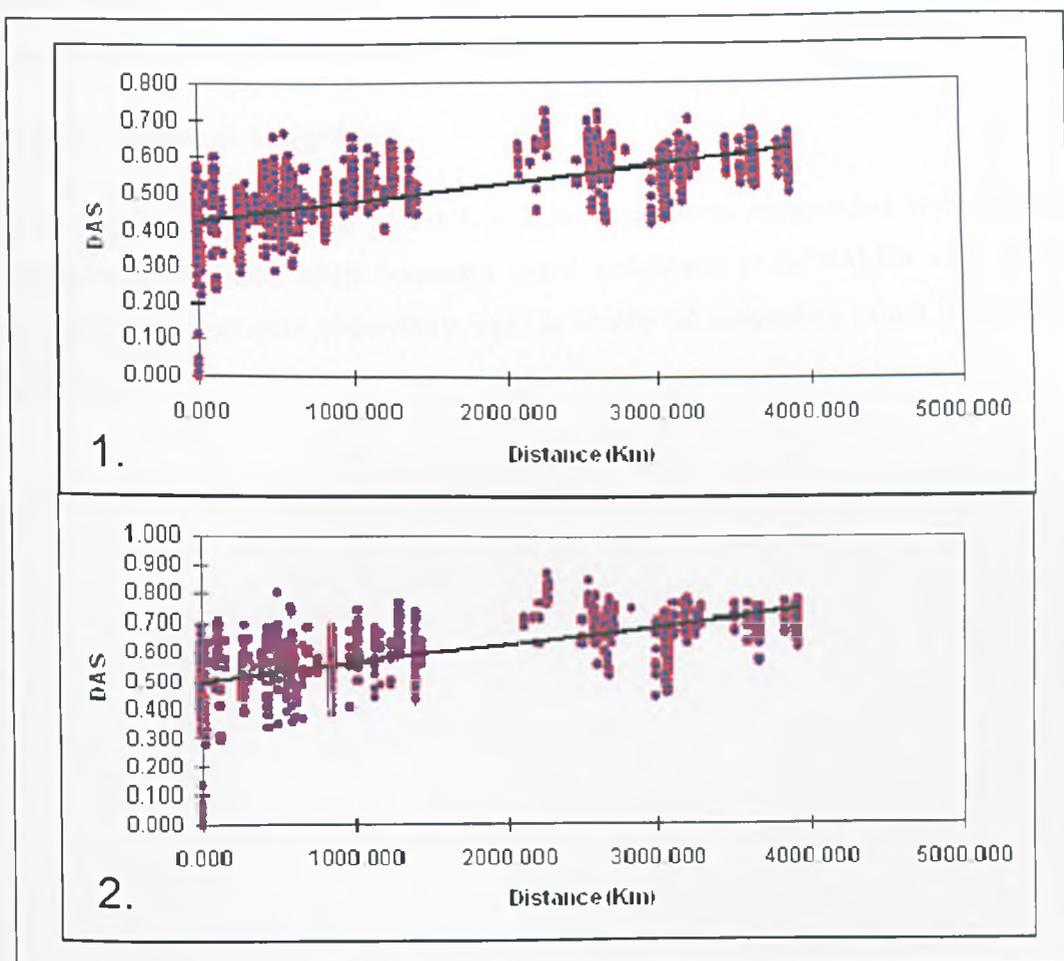
Figure 66 Linear regression (Mantel's test) of microsatellite D_{AS} pairwise genetic distance against geographic distance (km) for TCIIc isolates. Regression lines plotted in black. Plot 1 represents all TCIIc isolates, Plot 2 isolates from *D. novemcinctus* only. See text for correlation statistics.

5.2.2.2.6 *A priori* population assignments

A full list *a priori* population assignments is found in under 'Population 1' in Table 41, and population subdivisions are demarcated on the map in Figure 64. Two factors were considered when making assignments: geographic locality and sample size. As all samples analysed were drawn from the silvatic cycle only, transmission cycle was not considered relevant. Samples were assigned to four populations indicated on the map in Figure 64. Owing to the reduction in the global sample size with respect to TCI, all TCIIc populations

226

were necessarily smaller. Geographical subdivision between populations in Bolivia was deemed large enough to justify a split between sample sites from the north and south of the country. As a result one population, NB, contained only nine samples. The only Brazilian isolate, M5631, was grouped with Venezuelan and Colombian isolates, being marginally geographically closer to these populations than to those in Bolivia and Paraguay.

5.2.2.2.7 Statistical Assignment

As in Sections 3.7.6.1.6 and 5.2.2.1.7, *a priori* population assignments were tested for robustness by statistical allele frequency based assignment in GENALEX v6.0 (292). A compete table of samples, populations and −ln likelihood assignment values is included in Table 43.

Table 43 Allele frequency-based −ln likelihood values for population assignment of individuals within the TCIIc dataset. Values in bold indicate most likely population (least negative −ln likelihood) value for each individual.

Brazil, Venezuela, Colombia	Brazil, Venezuela, Colombia	Southern Bolivia	Northern Bolivia	Paraguay
CM17	**-49.111**	-93.718	-98.967	-95.570
M5631	**-114.417**	-121.241	-118.192	-115.326
M6	**-24.740**	-60.174	-56.680	-66.293
M8	**-26.503**	-61.850	-65.527	-70.378
PARAMA25	**-31.714**	-88.162	-86.769	-90.472
CM25	**-89.316**	-99.983	-92.833	-101.832
PARAMA26	**-27.813**	-92.345	-88.313	-89.864
M3-CU	**-31.826**	-69.281	-66.640	-78.873
PARAMA34	**-46.687**	-98.532	-88.780	-98.761
M10	**-67.971**	-89.092	-80.907	-91.595
PARAMA6	**-38.830**	-94.285	-96.338	-97.266
M5	**-29.189**	-89.698	-85.181	-94.546
Southern Bolivia	**Brazil, Venezuela, Colombia**	**Southern Bolivia**	**Northern Bolivia**	**Paraguay**
CAYMA21	-88.100	**-38.689**	-47.135	-78.510
MA202	-108.272	**-57.412**	-74.114	-61.331
MA204	-108.118	**-58.150**	-89.027	-60.918
MA215	-95.744	**-49.497**	-61.958	-50.816
MA219	-95.855	**-44.694**	-57.157	-71.643
MA220	-95.667	**-44.807**	-55.884	-66.612
MA222	-87.369	**-39.530**	-47.611	-65.658
SAM6	-109.750	**-51.952**	-73.330	-56.334
CAYMA3	-97.261	**-55.244**	-64.884	-75.095
CAYMA13	-76.368	**-30.033**	-42.102	-64.453
CAYMA14	-86.368	**-43.276**	-53.822	-54.471
CAYMA17	-78.361	**-31.704**	-45.532	-68.302
CAYMA18	-79.831	**-40.390**	-53.129	-64.695
CAYMA19	-108.514	-57.472	**-54.649**	-75.983
CAYMA25	-76.034	**-43.208**	-50.026	-49.457

228

Northern Bolivia	Brazil, Venezuela, Colombia	Southern Bolivia	Northern Bolivia	Paraguay
84/857	-95.991	-64.180	-61.952	-81.311
SJMC10	-103.425	-51.257	-50.215	-71.675
SJMC19	-65.863	-91.109	-88.939	-98.808
SJMC4	-94.105	-58.342	-55.135	-73.411
SJMO18	-96.335	-55.925	-53.915	-64.455
SMA18	-97.791	-56.896	-50.625	-73.133
SMA19	-85.800	-61.500	-46.127	-80.230
SMA8XE	-88.518	-48.620	-43.573	-68.425
SMA9	-96.044	-52.784	-50.140	-71.506
Paraguay	**Brazil, Venezuela, Colombia**	**Southern Bolivia**	**Northern Bolivia**	**Paraguay**
ARMA12	-109.956	-62.444	-82.424	-49.711
ARMA24	-106.057	-63.107	-77.814	-44.755
ARMA27	-91.319	-53.893	-69.689	-36.446
SP14	-103.370	-61.334	-81.630	-32.944
ARMA25	-92.915	-59.770	-68.312	-45.815
SP15	-98.311	-56.695	-60.933	-26.574
ARMA26	-87.016	-51.108	-57.246	-23.615
SP16	-101.190	-59.794	-67.739	-27.169
SP4	-106.570	-62.568	-79.410	-34.001
ARMA13	-74.665	-47.996	-63.566	-42.308
ARMA18	-92.590	-58.884	-69.214	-42.244
ARMA9	-86.880	-55.406	-63.515	-45.236
MA194	-94.052	-47.544	-63.378	-44.032
MA25	-84.643	-52.443	-60.132	-44.522
MA25X	-96.038	-61.085	-81.019	-46.006
MA87	-90.066	-58.633	-68.434	-42.312
SP13	-97.010	-54.887	-60.729	-26.645

Two reassignments were made according to the –ln likelihood scores displayed in Table 43. Concordant with its branching order in Figure 65, SJMC19 showed stronger affinities with BVC than with NB. By contrast, the reassignment of CAYMA19 from SB to NB received no support in Figure 65, and this isolate was retained in SB for later analyses. Interestingly statistical assignment encountered difficulties in assigning M5631 confidently to any population, consistent with its divergent status in Figure 65.
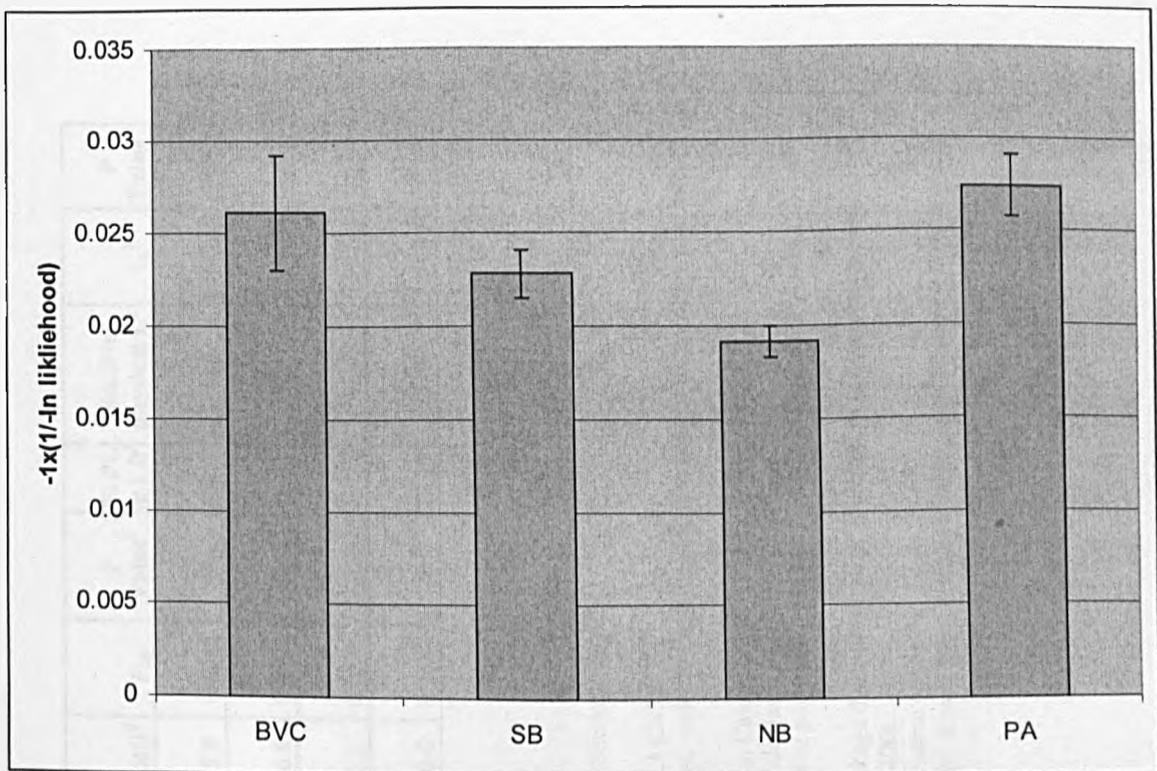
Figure 67 Transformed (-1x (1/-ln likelihood) mean –ln likelihood assignment scores for four reassigned TCIIc populations. Error bars represent +/- standard error about the mean.

Transformed mean –ln likliehood were calculated for each population once individual isolates had been reassigned and provide an approximate measure of population cohesion (Figure 67). A similar analysis when applied to TCI revealed clear differences between populations (Range TCI ~ 0.02-0.1, Figure 57), a far smaller range of values is present here (Range TCIIc ~ 0.019-0.027), and low overall values indicate that population definition is genrally poorer.

5.2.2.2.8 Population parameters

Several population parameters were calculated for each of the four populations post statistical assignment, as in Sections 3.7.6.1.4 and 5.2.2.1.8. The results are listed in Table 44. Genetic diversity values (MNA) varied little by population (4.22-4.78), certainly less so than those derived for TCI (1.73-6.67, Table 37), and showing equivalent levels of allelic richness to LB.

Table 44 Table of key genetic parameters for six TCIIc populations, post statistical assignment.

Population	N/G	PL	PA/S	GD	$H_O{}^a$	$H_E{}^a$	P value[b]		%HE[c]	%HD[d]	F_{IS}	P Value[e]	% PL in LD[f]	%LD in physical L[g]	$I_A{}^h$	P Value[h]
							D	E								
Brazil, Venezuela, Colombia (BVC)	13/13	21	0	0.53	0.327	0.578	0	1	0	15.9	0.4942	0	17.65	5.38	6.72729	<0.001
Southern Bolivia (SB)	15/15	23	0.33	0.56	0.490	0.587	0	1	0	10.6	0.1327	0.00881	16.47	7.25	2.94509	<0.001
Northern Bolivia (NB)	8/8	23	0	0.56	0.504	0.604	0	1	0	2.3	0.1679	0.01374	27.27	9.47	0.95219	0.041
Paraguay (PA)	17/17	16	0.26	0.52	0.438	0.544	0	1	0	10.6	0.2682	0	22.15	5.63	2.38703	<0.001
Total	53/53												12.12	15.0	2.94141	<0.001

N = Number of isolates in population.

G = Number of multilocus genotypes per population.

PL= Number of polymorphic loci.

PA/S = Multiple (3+) alleles per sample, calculated as a proportion of un-cloned isolates. See Section 5.2.2.1.3.

GD – Nei's unbiased gene diversity. Calculated in Microsatellite Toolkit (294)

[a]Mean observed and expected heterozygosity across all loci, calculated using ARLEQUIN v3.1 (295)

[b]Calculated in GENEPOP v1.2, using a Monte Carlo algorithm approximation of an exact test, under the assumption of statistical independence of loci. D – deficit, E - excess (See Section 3.7.6.1.4)(297).

[c]Proportion of loci showing a significant deficit in heterozygosity after a Bonferroni correction. Calculated in ARLEQUIN v3.1(295).

[d]Proportion of loci showing significant excess heterozygosity after a Bonferroni correction. Calculated in ARLEQUIN v3.1(295)

[e]Calculated in ARLEQUIN v3.1, as P (Random F_{IS}>=Observed F_{IS}), where a null distribution of population specific F_{IS} values (Random F_{IS}) was generated over 16000 permutations (Section 3.7.6.1.4)(295).

[f]Calculated in Multilocus v1.3 as the percentage of polymorphic locus pairs in significant linkage disequilibrium (r^d – related to the correlation coefficient r) by comparison to a null distribution of 1000 randomisations drawn from the same dataset (300).

[g]Calculated as the proportion of statistically linked polymorphic loci physically linked on the same contig.

[h]Calculated in Multilocus v1.3 by comparison to a null distribution of 1000 randomizations drawn from the same dataset (see Section 3.7.6.1.4)(300)

As with the majority of TCI populations, all TCIIc populations show an overall deficit in heterozygosity at a global level with respect to HW expectations. This observation is supported at the level of individual loci. Heterozygosity deficit is confirmed by significantly positive values for the inbreeding coefficient, F_{IS} in all populations. Furthermore, F_{IS} value sizes mirror the levels of homozygosity in the respective populations they describe.
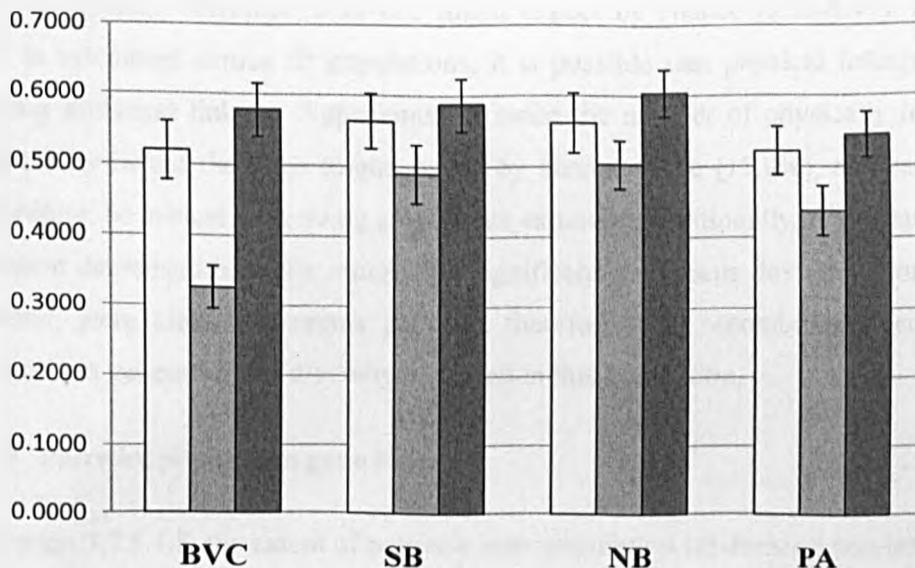
Figure 68 Histogram of three genetic parameters across seven TCIIc subpopulations. Populations associated with the domestic cycle are genetically depauperate and show increased relative levels of observed heterozygosity with respect to exclusively silvatic populations. Open bars – GD, dark grey - H_O, pale grey - H_E. Error bars represent +/- Standard error about the mean across loci (H_O & H_E) and +/- inter-locus standard deviation (GD).

When GD, expected and observed heterozygosity are examined in combination across populations (Figure 68), a similar pattern to that observed in diverse silvatic TCI populations emerges, whereby allelic richness corresponds to excess homozygosity.

Indices of association (I_A) point to strongly significant multi-allelic linkage in three populations - BVC, PA, and SB (p<0.001). In NB, however, multilocus linkage disequilibrium is only marginally significant (p=0.041). As with three TCI populations, VS. NEB and LB, 100% pairwise linkage was not observed in populations showing significant values for I_A. Surprisingly, however, the highest number of linked locus pairs was observed

232

in NB, incongruent with its marginally significant value for I_A. Low sample size is clearly an issue in this population.

Mutli-clonality may, again, act to confound estimates of linkage within TCIIc populations, even though the overall level of multi-allelic loci observed was lower than that in TCI. In this dataset, the proportion of physically linked locus pairs with respect to all possible pairs is 7.5%. At the population level, the proportion of loci both physically and statistically linked is no greater, therefore, than one might expect by chance (5.38-9.47). However, when r^d is calculated across all populations, it is possible that physical linkage may be influencing statistical linkage. Approximately twice the number of physically linked loci are statistically linked than one might expect by chance alone (15.0%), and sample size may, therefore, be critical in deriving an accurate estimate. Additionally, 6/8 isolates in NB, a population demonstrating only marginally significant multilocus deviation from linkage equilibrium, were cloned. It seems possible, therefore, that recombination could have played a role in generating the diversity observed in this population,

5.2.2.2.9 Pairwise population gene flow

As in Section 3.7.6.1.5, the extent of pairwise inter-population subdivision was tested under the IAM, using the fixation index F_{ST}, as well as under the SMM, using R_{ST}, in ARLEQUIN v3.1 and MICROSAT v1.2 respectively (290, 295). Values for F_{ST} are included in Table 45.

Table 45 F_{ST} **estimates of inter-population differentiation between four TCIIc subpopulations based on microsatellite data.** Italics indicate p-values generated from 1000 random permutations leading to a value lager than or equal to that observed. Constructed in ARLEQUIN v3.1(295). All values remain significant after Bonferroni correction.

	BVC	SB	NB	PA
BVC	*	*0.00000*	*0.00000*	*0.00000*
SB	0.28398	*	*0.00356*	*0.00000*
NB	0.28175	0.05112	*	*0.00000*
PA	0.33441	0.16606	0.17486	*

Highest levels of gene flow (F_{ST}<<0.2) were observed between Bolivian populations (NB&SB) consistent with negligible subdivision (301). Connectivity between NB and PA was marginally lower than that observed between SB and PA (0.17486 & 0.16606 respectively), and the overall pattern is consistent with isolation by distance, such that the

233

greatest subdivision is noted between BVC and PA (0.33441). The level of connectivity between NB and BVC, may be underestimated, however, in the light of the assignment of one NB isolate (SJMC19) to the latter population, prior to analysis. A UPGMA tree constructed on the basis of pairwise F_{ST} values is shown in Figure 69.
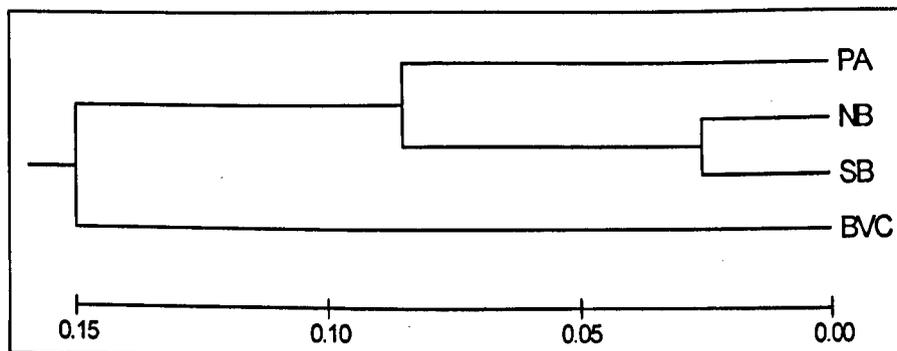
Figure 69 Un-rooted UPGMA tree based on pairwise F_{ST} values between TCIIc sub-populations constructed using MEGA v4(377).

No significant correlation, on the basis of a Mantel's test carried out as in Section 3.7.6.1.3, was observed between R_{ST} and F_{ST} estimates of subdivision between TCIIc populations (p>0.05, Table 46), possibly because of either low population number and/or smaller population size. In light of the signnificant correlation observed between these two statistics in TCI, it seems that saturation seems less likely to be behind this phenomenon in TCIIc, a putatively less ancient lineage (Section 1.8.2 and 5.2.2.1.9).

Table 46 Pairwise matrix of Slatkin's R_{ST} values between four TCIIc subpopulations based on microsatellite data. Values in italics indicate the standard error of the mean over 1000 bootstraps. Constructed in MICROSAT v1.2(290).

	BVC	SB	NB	PA
BVC	*	*0.027*	*0.032*	*0.034*
SB	0.097	*	*0.011*	*0.027*
NB	0.134	0.008	*	*0.030*
PA	0.151	0.097	0.147	*

5.2.2.2.10 Tests for population substructure.

As in Section 3.7.6.1.7, TCIIc isolates were tested in parallel for population substructure in SRUCTURE v2.1, for comparison with previous population assignment methodologies (See Sections 5.2.2.2.6 & 5.2.2.2.7) (306). The number of populations (K) was allowed to vary between 2 and 8, and a Gaussian curve derived from a plot of transformed −ln

likelihood (-1x (1/ln Pr (K/X))) against each value of K. For each value of K, ten independent runs were made, each with a burn in of 100,000, followed by a further 500,000 iterations of the Markov-chain algorithm. The mean of –ln likelihood across runs values is plotted in Figure 71. The Gaussian curve was observed to plateau at $K=6$. Confirmation for this observation was found by calculating ΔK, an extension of the plateau method that accounts for variance in –ln likliehood values for each value of K and the rate of change between the mean value for each value of K. Similar data were obtained to the plateau method. The graphical output for this run is included in Figure 70.
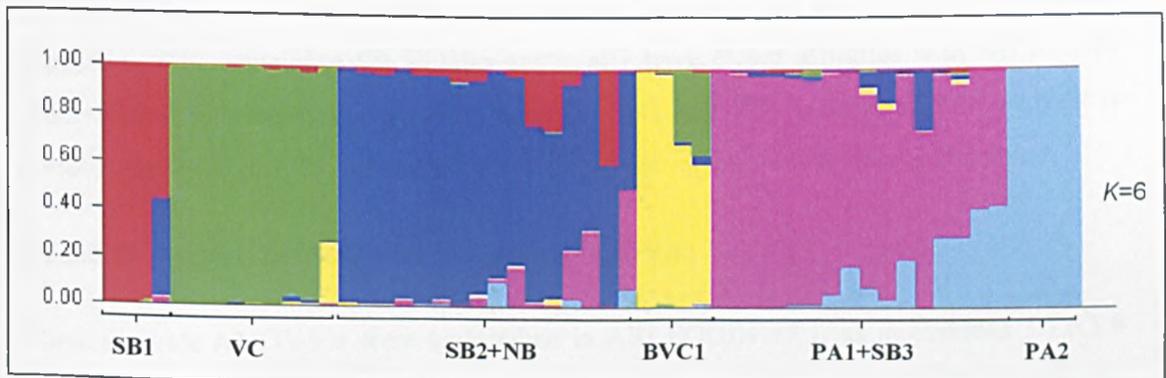
Figure 70 Graphical output of Structure v2.1 for $K=6$ (306). The run is divided into K colours. Vertical bars represent individual isolates. Population order is not necessarily indicative of relatedness. Admixture (bars made up of two or more colours) indicates where isolates may have split affinities between groups.
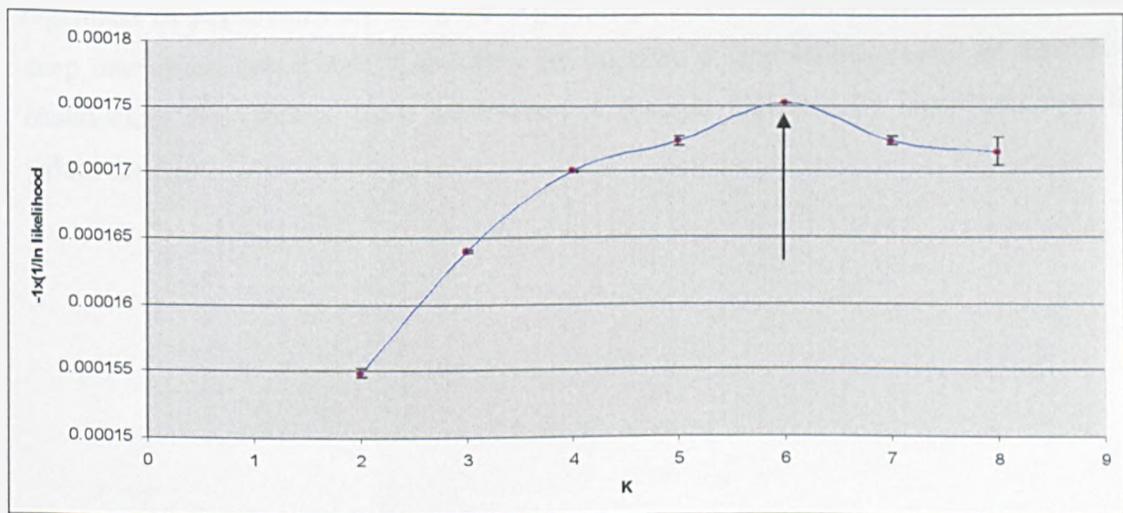
Figure 71 Estimation of the appropriate population number (K) from a Gaussian curve of transformed –ln likelihood scores generated in STRUCTURE v2.1(306). Error bars represent standard error about the mean of ten independent runs for each value for K. The approximate point of plateau (K=6) is indicated by an arrow.

235

Analysis in STRUCTURE reveals considerable subdivision among the populations defined previously, congruent with many features of the tree constructed using pairwise D_{AS} values (Figure 65). Tightly clustering clades within SB and PA are both identified as distinct populations here (SB1 and PA2), while four outliers within the BVC population, including SJMC19, also form a distinct population. This population appears to have formed by default, as all four isolates that comprise it are, in reality, highly divergent from each other, as well as from other populations. Interestingly, a degree of identity is maintained across the Chaco region; such that a subset of PA and SB isolates (PA1 and SB3) form a separate population. The remaining SB isolates apparently have closer affinities with NB isolates (SB2+NB). The subdivision apparent between SB2 and SB3 isolates in this analysis is corroborated in Figure 65, but unlike BVC1, does not receive >75% bootstrap support.

5.2.2.2.11 Analysis of Molecular Variance (AMOVA)

Three separate AMOVA's were undertaken in ARLEQUIN v3.1, as in Sections 3.7.6.1.8 and 5.2.2.1.11, in order to test the proportion of the total variance in the population accounted for by the population groupings made in previous sections (295). As with TCI populations, by far the majority of TCIIc variation (>72%) occurs at the individual level, regardless of populations are assigned. Again, high levels of allelic richness (MNA) and deep inter-taxon branching (D_{AS}) within the majority of populations correlates with this observation. Nonetheless, clear subdivision is present, indicated by significant overall values for F_{ST}.

Table 47 Analysis of molecular variance (AMOVA) undertaken for TCIIc subpopulations defined in previous sections. Calculated using ARLEQUIN v3.1(295)

Source of Variation	Degrees of Freedom	Sums of squares	Variance components[a]	% Variation	Fixation indices[a]	p - values[b]
Grouped by geography						
Among populations	3	95.980	1.08666 Va	**22.71**	F_{ST}: 0.22706	0.000
Within populations	102	377.303	3.69905 Va	**77.29**		
Total	105	473.283	4.78571			
Post-statistical assignment						
Among populations	3	97.505	1.10463 Va	**23.07**	F_{ST}: 0.23067	0.000
Within populations	102	375.778	3.68410 Vb	**76.93**		
Total	105	473.283	4.788873			
Grouped by Structure v2.1						
Among populations	5	127.286	1.34439 Va	**27.98**	F_{ST}: 0.27983	0.000
Within populations	100	354.997	3.45997 Vb	**72.02**		
Total	105	473.283	4.80436			

[a] See Section 3.7.6.1.8 for a details of calculation.
[b] Calculated by comparison to a null distribution of 16000 random permutations.

Across the three methods of population assignment employed, STRUCTURE v2.1 apparently performs best in defining populations that account for the greatest proportion of total variance (27.98%), while frequency based statistical assignment (23.07%) apparently only performs marginally better than *a priori* assignment (22.71%), even when unlikely assignments (e.g. CAYMA19 to NB) are excluded.

5.3 Discussion

5.3.1 Genetic resolution at the inter-lineage level: lessons from sequence and microsatellite data.

Both the sequence and microsatellite datasets presented in this study had the power to resolve most of the six major *T. cruzi* lineages as distinct and well supported monophyletic clades (see Sections 5.1.1.2 and 5.1.2). As in earlier sequencing studies, TCIId and TCIIe *gpi* haplotypes fell equally between the TCIIb and TCIIc, confirming their hybrid status, and TCI and TCIIb were apparently the most divergent clades (e.g. (199, 379)). In the absence of derived haplotypes, TCIIe and TCIId formed distinct but closely related clades on the basis of microsatellite data. The distinct allelic repertoires of these two strains are

reported fully by MLEW, and indicate that their origin is likely to be the result of at least two independent hybridisation events between TCIIb and TCIIc, an observation that has been made previously (199, 216, 375). Interestingly, microsatellite data also indicated that TCI and TCIIb are the most divergent clades (F_{ST} = 0.52053, Table 30), not including hybrids TCIId and TCIIe, in line with earlier studies employing the same type of marker (e.g.(215, 216))

The status of TCIIa is ambiguous across both datasets, where the group received poor bootstrap support from *gpi* sequence (64.3%, Figure 49), and D_{AS} bootstrap support <50% on the basis of microsatellites (Figure 50). The nature of this ambiguity may have arisen, in part, from the considerable divergence within the clade between North American and South American isolates, and *gpi* data suggested that this clade was by far the most heterogeneous (Section 5.1.1.2). Microsatellite data broadly supported this assertion but failed to distinguish, on the basis of pairwise D_{AS} values, the TCIIa clade with strong support from TCI (Figure 50). Nonetheless, the F_{ST} estimate between TCI and TCIIa suggested strong subdivision (0.35368, Table 30), and provides reasonable grounds to uphold their identification as separate lineages. Comparable microsatellite data to support the status of TCIIa in the literature are almost non-existent. Only a single isolate has ever been analysed, CANIII, from which five nuclear loci were amplified, and this strain also clustered more closely with TCI than with any other lineage (216).

The most important incongruity between the two genotyping methodologies reported here relates to the branching order between the predominately silvatic groups TCI, TCIIa and TCIIc. *Gpi* sequence analysis was consistent with a stronger affinity between TCI and TCIIc than TCI and TCIIa, while the inverse was apparent from the ten microsatellite loci analysed. This phenomenon has at least three different possible explanations. Westenberger *et al.*, 2005 reported similar incongruence with relation to TCIIc and TCIIa between genealogies drawn from sequence analysis of several genes across at least four distinct linkage groups (191). According to their interpretation of these data, TCIIa and TCIIc display mosaic genotypes, consistent with a hybrid origin from an early fusion between TCI and TCIIb (Figure 14). Recombination is not the only process that can generate phylogenetic incongruence, however, and the under-representation of key lineages in a dataset, especially TCIIa, may be partly responsible for the instability they observed. A

recent study based on *gpi* sequence undertaken by Broutin *et al.*, 2006 is a case in point (379). Their study, based on the same sequence target as this one but including only 12 *T. cruzi* strains, reported stronger affinity between TCI and TCIIa than between TCI and TCIIc.

A second explanation arises form the difficulty of deriving haplotypes from microsatellite data, especially where full multilocus linkage is unknown, as is the case here (Section 5.1.2.1.1). TCIIc probably clusters further from TCI than is the case in evolutionary terms, because of a clear association with recombinants TCIId and TCIIe. A tree based on microsatellite haplotypes alone would thus, in all likelihood, represent more closely the branching order observed from *gpi* sequence analysis.

Finally, it is highly probable that saturation prevents the drawing of meaningful phylogenetic inference from the deep branching order between major *T. cruzi* lineages established on the basis of microsatellite data. It would equally be disingenuous, therefore, to propose that microsatellites could provide any information regarding ancient recombination events. *T. cruzi* is incredibly diverse, with at least as much sequence divergence between TCI and TCIIb as between *L. mexicana* and *L. major*, consistent with an ancient origin even under even the most conservative estimates (Section 1.8.2) (369). The poor resolution of the major clades under an SMM was consistent with the high level of saturation that must have occurred at microsatellite loci since their divergence (Sections 5.1.2.1.3 and 5.1.2.1.2).

Despite the obvious importance of saturation as a phenomenon when studying microsatellite loci, all previous analyses of inter-lineage diversity using these markers have employed stepwise mutational models (215-217, 380). In most cases, the resulting lineage resolution has been extremely poor by comparison to sequence data. The future role of applying microsatellites to study *T. cruzi* at a species level is thus unclear. Clearly the potential for using these markers for analysing and/or dating deep divergence is extremely limited. Nonetheless, it is shown here that microsatellites could provide a relatively rapid and robust means of assigning strains to at a major genotype level, under an IAM at least, on the basis of a limited panel of markers. Additionally microsatellites can easily distinguish TCIId from TCIIe, an objective less easily accomplished using coding

sequences (Section 5.1.1.2) (191). By comparison to an expanded panel of microsatellite loci and *T. cruzi* isolates, it may even be possible to trace the most likely TCIIb and TCIIc parents of these two hybrid lineages (375).

Closing the gap between sequence and microsatellite analyses in terms of their respective abilities to resolve distant and near-past evolutionary events remains a significant challenge with *T. cruzi*. One obvious solution is to select non-coding, less highly conserved sequences, like those applied to the recent elucidation of *T. rangeli* phylogeography (339). Another means of boosting phylogenetic resolution is by using multiple conserved sequences in combination (MLST), or by combining multiple marker types in a single combined analysis, as has been recently applied to a major taxonomic revision of the *L. donovani* complex (200, 319). The key problem with the latter analysis is that different marker types are still constrained by different respective mutational models, and detailed inference, such as phylogenetic dating, must still be limited to sequence data alone. Finally, heterozygous loci must be assigned as null alleles to allow diploid marker profiles to be concatenated into haplotypes, a strategy that would limit the capacity to resolve extremely heterozygous lineages like TCIId and TCIIe.

5.3.2   Genetic resolution at the intra-lineage level from *gpi* sequence data.

The panel of TCI and TCIIc strains employed to assess the level of intra-lineage *gpi* sequence diversity represents the largest ever assembled for analysis at a single locus, by comparison to previous studies. Additionally, isolates were deliberately selected out of the larger panel used in later microsatellite analysis (see Table 34 & Table 41) to represent, as fully as possible, the variety of different geographic regions, hosts, and vectors sampled.

As in Section 5.1.1.2, *gpi* sequence diversity, equating to five fixed SNPs, clearly distinguished TCI and TCIIc as separate monophyletic clades. TCI showed substantially greater levels of intra-lineage diversity (1.4%) than TCIIc (0.9%), perhaps a function of the much wider geographic sampling of TCI isolates, but also because it is a putatively more ancient lineage with a more diverse range of host species (Section 5.2.1.2, Section 1.8.2) (29, 191). Three tests employed to detect mosaic (recombinant) sequences among those analysed gave negative results (Section 5.2.1.4.1). This outcome is perhaps unsurprising

given the brevity of the sequence fragment. Nonetheless, mosaic sequences have been detected before at this locus, albeit within the known TCIIe hybrid strain CLBrener (183).

The overall aim of this analysis was to examine the potential of *gpi* sequence diversity to identify any geographical or host-specific clustering of isolates. For TCI evidence for geographical clustering among strains and/or haplotypes is highly limited, and very few internal branches received any bootstrap support (Figure 53). Nonetheless, those internal clades that did show >50% bootstrap support corresponded to geographically-linked samples, but not as clearly as one might expect if spatial population subdivision was high (i.e all samples derived from the same study site should theoretically cluster together). At the host and vector level no evidence of clustering was observed at all, and strains isolated from different mammal species clustered together even in well supported clades. Addionally there was no obvious partition by triatomine species, although these were under-represented in the dataset.

Within TCIIc, the level of geographical clustering identified is arguably higher (Figure 53). Here the majority of isolates fell within two relatively well supported (~60%) clades. The first predominantly comprised isolates from Northern Bolivia, Colombia, Venezuela and Brazil; the second exclusively isolates from Bolivia and Paraguay. An important exception was the presence of one isolate from SB (CAYMA19) in the northern clade, and this observation is discussed later in the context of microsatellite data. Similarly, two outlying haplotypes, M10_1 (DN_V2_b) and SJMC19_1 (DN_B07_b), fell outside their expected position with relation to other isolates. The anomalous position of SJMC19 is particularly interesting, whereby the strain appeared to have split ancestry, with one haplotype occurring alongside a northern isolate, the other firmly within the southern clade. This isolate was also one of the few isolates heterozygous at three nucleotide sites, and a possible hybrid origin is certainly worthy of further investigation.

Single locus housekeeping gene sequence diversity studies are probably inadequate in resolving population genetic structure between closely related strains or species. Because these sequences tend to be highly conserved, there is often not sufficient resolution to address detailed biogeogrpahic hypothesise. Nonetheless, *gpi* sequence data have proved successful in resolving spatial clusters within the *L. donovani* complex, even though the

level of sequence diversity across the group is lower than that in both *T. cruzi* subgroups studied here (~0.8%) (319, 381). Thus the low geographic resolution of the current dataset may, in part, be derived from a genuine population phenomenon, as opposed to merely representing a methodological limitation. The distribution of the *L. donovani* complex across Europe, Africa and the Middle East is punctuated by a far richer diversity of geographical features - oceans, seas and mountain ranges - than the gradual ecotypical cline present across the distribution of *T. cruzi* in the Americas. As such, the limited diversity present at the *gpi* locus in *L. donovani* may be far more discontinuous in its distribution than that of *T. cruzi*, despite the fact that, in Asia, Africa, Europe and the Middle East, *L. donovani* is likely to represent a more recent radiation than TCI, and possibly TCIIc, in the Americas (199, 319).

Antigenic genes and intron sequences are perhaps a better target for assessing fine scale population differentiation on the basis of a single locus. Their efficacy in identifying intra-lineage (e.g. *T. gondii*) and intra-species (e.g. *P. vivax*) diversity is proven, especially when the gene is known to be under balancing selection (382, 383). Herrera *et al.*, 2007, sequenced a 350 bp fragment of the mini-exon unstranscribed spacer region from 12 TCI isolates in Colombia, incuding domestic and silvatic strains, and were able to identify 17.5% sequence divergence and four distinct haplotypes, although the ecological and/or geographical affinities of these groups were ambiguous, and sample size too low for meaningful inference (384). Further possible targets within *T. cruzi* could include members of the mucin and MASP gene families; although these genes are frequently present as multiple copies (see Section 1.1.1). In this study, *gpi* was sequenced more as a point of reference for later microsatellite analysis than through any major expectation that it might resolve fine scale population differentiation. Congruence between the two types of marker is discussed later in Section 5.3.5.

5.3.2.1 MLMT as a tool for detecting TCI&TCIIc population genetic structure.

The panel of isolates assembled for the analysis of TCI population genetic structure represents the single largest ever analysed. Attempts to study the genetic diversity of this abundant and widespread genotype are historically limited to less than fifty TCI isolates across a limited geographic range and with poor resolution (e.g (385, 386)). One significant

242

recent study has been conducted, employing over one hundred TCI strains from seven countries (387). This study employed MLEE across 18 enzymatic loci, an extremely laborious task, and the authors were able to detect some diversity within TCI stocks, although limited by comparison to that detected here. The largest number of TIIc isolates analysed in any previous study is 17, and no thorough attempt to assess intra-lineage diversity has ever been made (29).

The present study proves that, using the online *T. cruzi* genome, a large, informative panel of microsatellites can be assembled at high speed and low cost. By focussing on polymorphic loci that amplify robustly across isolates using a standardised protocol, a high throughput multilocus typing system can be developed to study intralineage genetic diversity using a bare minimum of parasite genetic material (~50ng of whole genomic DNA/isolate, see Section 3.7).

5.3.3 The population genetic structure of TCI, as revealed by MLMT.

Pairwise genetic distances (D_{AS}) were astonishingly high between the majority of TCI isolates - no single pair of isolates was identical at all loci. To place this in its proper context, identical multilocus microsatellite genotypes are a common phenomenon in studies of other parasitic micro-organisms including *L. donovani* (number of loci (NL) = 15), *L. tropica* (NL=21), *Coccidioides immitis* (NL=9) and *Batrachochytrium dendrobatidis* (NL =12)(214, 266, 267, 388). Clearly the likelihood of obtaining a multilocus match declines as the number of loci examined increases. Given that most samples differed by approximately half of all loci (D_{AS}~0.5), however, it seems likely that one would have to employ a far lower number of loci before identical multilocus genotypes were observed in the dataset.

The level of spatial structuring in TCI isolates, by reference to individual pairwise genetic (D_{AS}) and geographic estimates, was low, although significant. Isolates fell into three classes: lowland 'Amazonian' samples (the majority - including silvatic isolates from Venezuela and some from Argentina, as well as NEB anf LB), where the extent of diversity observed within a geographic focus was only marginally lower than that between foci; highly similar isolates from the same geographic focus, e.g. COT, CH, and some

243

Argentinean isolates; as well as isolates that were genetically similar despite being drawn from distinct geographic foci e.g VD, VD/AM, COT/CH, and AM.

True isolation by distance (IBD) on continental or local scales is actually a fairly rare observation in parasites that infect humans. Population genetic data from numerous organisms, including *P. falciparum*, *Candida albicans*, and *Shistosoma mansoni*, suggest that rapid human migration serves to disrupt spatial genetic structure (320, 389, 390). It is perhaps unsurprising, therefore, that the level of spatial genetic structure increased when only silvatic isolates from *D. marsupialis* are considered. Three groups of isolates comprised of at least some parasites derived from humans (VD, VD+AM and COT+CH) fell as outliers in Figure 56. If humans are responsible for active dispersal of parasite within or between these populations, they are doing so at a higher rate than is occurring in wild transmission cycles.

The extent of spatial genetic structure that existed in wild parasite populations was potentially not as well defined as that observed on the basis of *cytb* analysis of many South American mammalian TCI reservoir species, including *D. maruspialis*. Patton JL and his co-workers noted marked divergence across Amazonia in 15 different orders of mammal, well aligned to current and paleogeographic physical barriers (344, 391). Additionally, while there was a possible association between some major parasite lineages and different hosts or niches (see Section 4.3), TCI was apparently promiscuous in the silvatic cycle, and no specific associations between particular strains and mammal or vector species were observed.

Assigning individuals to the 'correct' populations represents arguably the greatest challenge in population genetics. Three methodologies were applied here, and each tested using an AMOVA to assess how well the overall diversity was captured by the groupings imposed. Perhaps the most notable re-assignments were those of isolates between VS and VD, and COT and CH. Venezuelan domestic strains were initially assigned in an *a priori* fashion to a distinct population. Such a grouping is potentially contentious in the light of the current understanding of vector population dynamics in the region. *Attalea* species palm proximity has been statistically identified as a risk factor for the colonisation of domestic structures by silvatic *R. prolixus* in Barinas state (65). Additionally, a recent study combining

microsatellites and mitochondrial sequence analysis, identified considerable gene-flow (e.g. Palm – House, Cascabel, Barinas, $F_{ST}=0.009$ $p=0.37$) between *R. prolixus* silvatic and domestic populations in two of the three states studied (Barinas and Portuguesa, 100% and 80% of population pairs respectively) (392). Thus it would have been perhaps more appropriate to assign all Venezuelan isolates to a single population under the assumption that silvatic triatomines invasive to the domestic environment were infecting humans with silvatic-type TCI strains. The three human and single domestic triatomine isolates analysed from Barinas and Portuguesa were all distinct, however, from those recovered from the local silvatic environment. Instead they had far greater affinity, on the basis of all available measures, to other human domestic strains from around Venezuela. Additionally, the estimated level of gene-flow (F_{ST}) between the VD and VS was lower than that between VS and distant populations (e.g. ARG), a remarkable observation, and suggestive of a high degree of population subdivision despite geographic co-occurrence. The epidemiological significance of the separation of domestic and silvatic TCI populations in Venezuela is not clear. Perhaps the inefficiency of the faecal transmission route is important. In this case parasites entering dwellings with silvatic adult triatomines may infrequently contaminate their human occupants. Instead the nymphal offspring of colonising adults, higher in number and requiring multiple blood meals, simply spread pre-existing strains among humans. Alternatively a selectionist argument could be invoked, such that, despite repeated human inoculation with diverse silvatic strains, only a competent subset goes on to establish infection. An analogous situation is thought to arise in *T. brucei*, whereby certain genotypes or subspecies, e.g. *T. brucei rhodiense* are thought to be selectively adapted to human infection, while *T. brucei brucei* is not; thus, mixed infections of these two subspecies are only found in tetse flies (236). This theory is perhaps corroborated by the fact that a number of other domestic triatomines from around Venezuela contained silvatic-type, not human type, *T. cruzi* strains. No information is available, however, as to whether these were adults, for if they were they could well be invasive from the silvatic environment.

While selection is an attractive hypothesis, it does not fit well with the eclectic nature of the majority of TCI isolates with regards to host. Furthermore, it is undeniable that some invasion of strains from the silvatic cycle must have occurred - statistical assignment shows

that two human strains, JR and 9010, do belong to the silvatic population, although vector population dynamics in their respective states of origin, Anzoategui and Trujillo, have not been adequately investigated. Silvatic *R. robustus* is thought to invade houses in Trujillo state, but it is not clear to what extent domestic triatomine colonies are established, or the infection rate of invading individuals (64, 393). Additionally, information regarding the disease status of the patients sampled may also serve to refute a selective hypothesis, and it seems that, in one case at least, a silvatic-type strain was associated with a chronic infection, suggesting that domestic-type strains may not be unique in their capacity to establish long-term infection in humans.

The re-assignment of Chilean isolates (CH) from their *a priori* grouping, on the basis of geographical proximity, with Argentinean isolates, to a new population including more distant Cochabamban isolates, presents an interesting biogeographical problem. COT+CH formed a very tightly clustered and highly cohesive group of samples on the basis of D_{AS} and statistical assignment values, far greater than samples from LB, for example, which all originated from within a ten kilometre radius. Cochabamba and Limari (the origin of the Chilean isolates), on the other had, lie approximately 1600km apart on either side of the Andes. Endemic *T. cruzi* transmission is virtually absent from the inhospitable high Andes (see Section 1.5), and it is entirely possible that infected migrant human populations may be responsible for parasite dispersal in this instance. Paradoxically, the majority of isolates from both Chile and Cochabamba were classified as 'silvatic'. If humans were involved, then at least one of these populations should be domestic or epizootic in origin. In both cases the isolates actually originate from what can only be termed 'borderline silvatic' foci. CHILE C22 and CHILEWALL were both isolated from *Mepraia spinolai* infesting the stone walls of goat corrals surrounding highly endemic communities in the 1980s, albeit alongside the caviomorph rodent *Octodon degus* (394). Similarly, the study site at Cotopachi, from which the COT rodent isolates were collected, is located within 4km of the second largest city in Bolivia (Section 4.1.1.4). In the light of these data it may be worth re-examining the silvatic status of *T. infestans* identified at the same site (43, 44, 116, 222).

Population assignment, especially in relation to STRUCTURE analysis, provided further clues to the broader associations of both COT+CH and VD. In grouping half of the Argentinean isolates alongside those from COT and CH, STRUCTURE indentified a clear

246

south-western TCI clade, also apparent from D_{AS} measures (Figure 72). Of particular interest, however, was the relationship between Venezuelan domestic and North American strains, which were identified as a single population at optimal values for K, as well as by reasonably tight and well bootstrapped clustering on the basis of D_{AS} values. There is no obvious epidemiological explanation for this observation. One can reasonably exclude the possibility of a long range introduction of a strain or strains from silvatic transmission cycles in the Southern USA into Venezuelan domestic transmission cycles, or vice versa. It seems more likely that Central American strains should have a closer affinity to Venezuelan domestic isolates on the grounds of geography alone. TCI is the predominant lineage throughout Central America, although this population is very poorly represented in the current study (387, 395). The single Central American isolate that was analysed, DAVIS 9.90, from Honduras, did cluster with North American isolates, but more samples are required to clarify any association.

TCI populations LB, VS, and NEB, which encompassed the majority of isolates, were by far the least cohesive (by statistical assignment) and most diverse (on the basis of D_{AS} values) studied. Estimates of sub-division between these populations, from both pairwise F_{ST} and R_{ST} measures, were the lowest encountered (0.10808-0.14811, Table 38). At the optimal value for K, some intra-population partitions were made by STRUCTURE, but, with the exception of a link between a subset of LB strains and a subset of Argentinean isolates, these partitions corresponded poorly to other measures employed in this study (e.g. D_{AS}). Two of these populations, LB and NEB, are both drawn from lowland Amazonia. The third, VS, is ecologically more diverse in origin, with most isolates derived from within the Orinoco river basin, but two, from Colombia, isolated from silvatic reservoirs west of the Andes. The level of connectivity between all three populations suggested that no major geographical or ecological barriers exist to hosts and/or vectors, assuming parasite dispersal has occurred exclusively between silvatic transmission cycles. Palm species, especially from genus *Attalea*, the favoured habitat of many arboreal didelphid mammals and *Rhodnius* species triatomines, are widespread throughout all these regions, including Pacific Colombia (100, 269). Additionally, the three mountain ranges or Cordilleras that comprise the portion on the Andes that divides eastern and western Colombia are

interspersed with low interconnecting semi-tropical valleys, arguably a less severe barrier to silvatic gene-flow than the more extensive mountain ranges that occur further south.
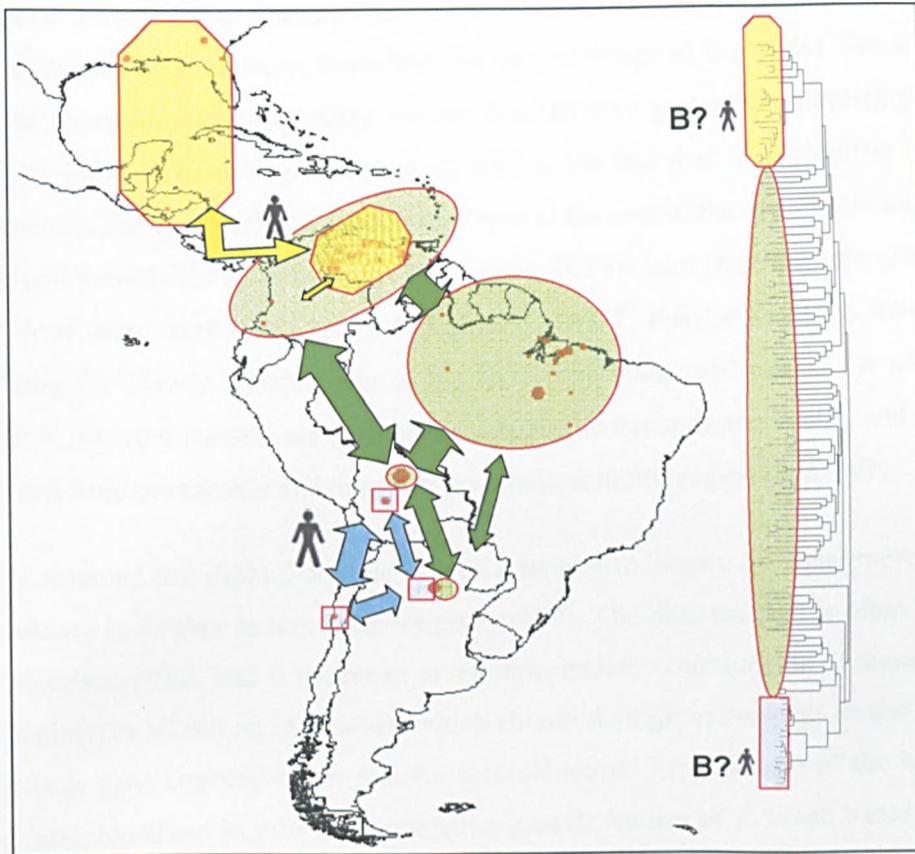
Figure 72 Schematic summary of TCI population dynamics, including putative human involvement in parasite dispersal. Coloured polygons coincide with clades as defined on the D_{AS} NJ tree (right). Coloured arrows indicate connectivity between population groups as a composite between F_{ST}, D_{AS} and population assignment. Man symbols on map indicate where humans may have role in parasite dispersal. Man symbols on tree represent clades containing human samples. "**B?**" represents populations where a population bottleneck may have occurred.

Connectivity between the south-western population group (COT+CH & ARG2, Section 5.2.2.1.10, Figure 72) and the Amazonian and Orinocan core population group (NEB, LB &VS), was typically low. Almost complete isolation was observed between COT+CH and LB (F_{ST}=0.31917). ARG was apparently closer to the core group until subdivision within the population is taken into account (Figure 72). The Chaco region, from which the Argentinean isolates are derived, is much drier than the humid tropical lowlands of the core group, and more reminiscent of the Andean puna at the COT study site (Section 4.1.1.4). Additionally, most isolates from ARG were drawn from *D. albiventris*, a host that was also

found at the COT study site (Section 4.1.1.4), and known to replace *D. marsupialis* in more arid clines, although none of those captured in the current study were infected (156). Thus a shared host and ecotope may partially explain the level of connectivity between COT and some isolates from ARG, disregarding human influence, especially as no major geographical barriers to gene-flow exist along the eastern fringe of the Andes. The level of gene-flow between the remaining ARG isolates and the core group was surprising given their lack of continuity in ecological terms, as well as the fact that *D. albiventris* isolates also occur among this group. Perhaps moist ecotopes to the east of the dry Chaco, including the Panatal and humid Chaco, form a corridor linking TCI isolates from the core group and a proportion of those from ARG (269). Intriguingly, two *T. infestans* isolates from ARG cluster among the silvatic isolates, one in the COT-type clade, and one in the core-type clade. Both *T. infestans* isolates are derived from domestic transmission cycles, and suggest a possible link between silvatic and domestic populations in this region (396, 397).

It is widely assumed that diploid organisms with a long term history of clonal reproduction have a tendency to display excess heterozygosity (299). The phenomenon is often referred to as the Meselson effect, and is the result of the independent accumulation of mutations on separate haplotypes within an individual, which should diverge irreversibly in the absence of mitotic (e.g. gene conversion) or meiotic recombination. No evidence of the Meselson effect has been identified in previous population genetic studies of *T. cruzi*, based on both microsatellite and sequence analysis, and once known hybrid lineages TCIId and TCIIe have been exlcuded (199, 215). Indeed, as in this study, most *T. cruzi* populations demonstrate excess homozygosity. As such, it has been proposed that some genetic mechanism must be at play, and those proposed to explain this phenomenon in *T. cruzi* include a level of recombination within or between individuals (398). Historically, tests for multilocus linkage disequilibrium and/or clonality in *T. cruzi* populations have been hampered by sample size (e.g. (399)), a focus on genetically identical hybrid strains (e.g. (400)), an inability to account for pre-existing sub-divisions (i.e. major lineages) within the population (e.g. (252)), an expectation that any recombination that did occur would result in HW levels of heterozygosity (399), or, unfortunately, all of the above. In this study, inter-locus linkage was assessed not only at an intra-lineage level, but also, as far as possible, within subpopulations of TCI. It seems reasonable that, if recombination was to occur, it

may be most likely to happen between genetically similar (in relative terms) individuals with frequent contact (i.e. from the same geographic region exploiting the same ecological niche), and, given that genetic exchange in *T. cruzi* is unlikely to be Mendelian, it may not result in HW allelelic frequencies (183). No evidence for multilocus linkage equilibrium, was obtained. However some evidence of linkage equilibrium, and therefore possible recombination, was obtained on the basis of low pairwise r^d values from the three 'core' populations, NEB, VS and LB (Section 5.2.2.1.8). These populations correspond to those showing the highest levels of excess homozygosity and genetic diversity, but also of multiple alleles, which unfortunately may have played a confounding role (Section 5.2.2.1.8). Diversity levels in NEB, VS, and LB were far higher than those typically observed from microsatellite studies of comparably-sized domestic and peri-domestic *Leishmania* populations (266, 267, 307). Instead allelic richness levels mirrored more closely those obtained from Brazilian *T. gondii* isolates, where the non-human, albeit domestic mammal, cycle was shown to harbour considerably more diversity than human isolates. Comparable, high resolution population genetic surveys of silvatic parasitic zoonoses are, to date, extremely rare. Still, it is clear in the context of this study that lowland, palm-associated TCI ecotopes represented a considerable pool of parasite diversity with respect to other TCI populations. LB provided an excellent example of where this level of diversity was only weakly dependent on sample size. Despite possessing under half the number of individuals, levels of diversity (GD) were only 5% lower than those of NEB and VS.

Evidence of strong linkage disequilibrium (LD) correlated with reduced allelic richness in the four remaining populations. North American strains were particularly of interest because of the relative recency (c.2.5 MYA, Section 1.6.2.1) of contact between that landmass and South America. Barnabe *et al.*, 2001 suggested that, on the basis of 17 North American TCI strains, not only were levels of diversity sufficient to suggest a pre-historic introduction of TCI into North America, but also comparable to those one might find in an equally sized sample from any other silvatic transmission cycle in South America (340). In the current study, NA isolates were clearly divergent, suggesting long term isolation from the rest of South America, but a marked decline in sample size corrected gene diversity was also observed.

Interestingly ARG, COT+CH, and VD isolates showed moderate increases in heterozygosity with respect to other populations, as well as strong linkage disequilibrium and low gene diversity. All three populations seem to be associated with some type of ecotope or host switch with respect to 'core'. The interpretation of this phenomenon in isolates from ARG is problematic, especially as sub-division existed within the group. Attempts to establish whether this increase in heterozygosity was indicative of a bottleneck in the remaining populations proved inconclusive - most of the standard tests applied to populations that have undergone a bottleneck assume Hardy-Weinberg allelic proportions in the 'parent' population. If TCI, or its progenitor, did emerge alongside didelphid marsupials, whose initial radiation is estimated to coincide with the formation on the Amazon river system c.38MYA (Se Section 1.6.2.1), then, taking into account, the substantial diversity observed in 'core' populations, these could be at least more ancient than South Western and North American 'offshoots', if not actual parents. 'Core' populations clearly convened HW expectations of H_O, but showed rare allele frequency distributions consistent with their high observed diversity and possibly a large, stable N_e. The assumption that populations of organisms exhibiting high levels of genetic diversity could be ancestral to linked populations showing reduced diversity is not a new one. The study of human population genetic diversity best exemplifies this situation, where an astonishingly diverse group of African populations gives rise to serial reduced-diversity bottlenecks in Eurasia, Autralasia and the Americas (401). Most studies investigating and/or modelling population bottlenecks quote a reduction in diversity as a key expectation, other indices aside (310, 314, 317). Perhaps *T. cruzi* populations are simply not amenable to any other measure.

Even if Mendelian genetic exchange does not occur in TCI, a non-Mendelian mechanism has been identified, and evidence for genetic exchange has been tenuously observed before in RAPD data from North Eastern Brazil (181, 183). Perhaps, therefore, TCI bears more than a passing resemblance to other parasitic (e.g. *T. gondii, T. brucei, & P. falciparum*) population genetic structures, whereby panmictic, or at least sexually competent, epidemic populations demonstrating low levels of linkage disequilibrium can give rise to clonal outbreaks in response to new hosts and/or environments (Figure 16)(176, 236, 263). It is, however, impossible to categorically state that genetic exchange is occurring in the 'core'

TCI populations studied here. In fact, the observation that statistical linkage may be independent of physical linkage may be reasonable grounds to suggest that apparent linkage equilibrium is an artefact of multiclonality. Perhaps, therefore, the complete pairwise and multilocus linkage disequilibrium observed in many non-core populations was not a distinguishing feature indicating higher levels of clonality, rather a further indication of reduced diversity. Also of interest was the increased level of H_O and low level of diversity observed in COT+CH and VD which were, in a small way, reminiscent of hybrid strains TCIId and TCIIe (375). If raised heterozygosity in these populations was not the result of a bottleneck — which is difficult to test in a non-Mendelian system - could they be the result of a hybridisation event, in this case between less divergent strains than the TCIIb and TCIIc parents of the known hybrids (i.e. within TCI), and thus showing less pronounced heterozygosity? It has been proposed, largely from circumstantial evidence based on their almost exclusive occurrence in domestic transmission cycles, but also from limited empirical data, that the heterozygous nature of TCIId and TCIIe may confer some advantage to survival in the human host (240). It is conceivable that the two human-associated TCI lineages studied may be governed by the same assumptions. Presently, however, sample size remains an issue, and ideally non-core populations should at least match those classified as 'core' in size, before satisfactory conclusions can be drawn.

5.3.4 The population genetic structure of TCIIc, as revealed by MLMT.

Many fewer TCIIc isolates were available for analysis than for TCI and this strain has been historically extremely poorly sampled by comparison (Section 4.4.2). Principally thanks to a recent study in Paraguay, and with the aid of Nidia Acosta, LSHTM, a panel of strains deemed sufficiently large to address at least some population genetic questions was assembled (29). The analytical approach adopted for the analysis of TCI genetic diversity was also applied to TCIIc genetic diversity, with a view to facilitating direct comparisons between the two datasets.

Identical multi-locus matches were, as with TCI, totally absent from TCIIc. D_{AS} values indicated that levels of diversity, typified by deep branches between taxa, from individual foci were, in some cases, equivalent to those observed in TCI. In two cases, once in SB and once in PA, pockets of 3+ genetically similar isolates were identified, not apparently linked

252

to any specific host, but from the same location. Manual examination of the large cluster at SP indicates that these isolates varied at only a single locus, not including missing data, a far lower level of diversity than was ever noted for TCI.

Genetically similar isolates from the same study site are a norm in parasite population genetics (e.g. (267)). No evidence of widely distributed, genetically related clades was found in TCIIc. As such, unlike TCI, no outliers were identified on the basis of a Mantel's test for IBD. Indeed the level of spatial structure within TCIIc populations was considerably higher than that observed for TCI, possibly, in part, due to a lack of human influence – TCIIc is incredibly rare in the domestic cycle, only five isolates have ever been recorded, one from North-Eastern Brazil and four from Paraguay (29, 171, 402). Like the majority of TCI strains, however, TCIIc strains appeared to be eclectic within their niche, and parasite genetic diversity seemed to be largely independent of host on the basis of available data (only a single vector-derived isolate was available). To examine levels of spatial structure among populations of the major host, *D. novemcinctus*, a second Mantel's test was conducted and can be compared directly with that obtained for the major host of TCI, *D. marsupialis*. In this case the level of spatial structure was still higher among *D. novemcinctus* isolates than among those from *D. marsupialis*. Linking the spatial genetics of TCI and TCIIc to the relative dispersal capacities of their major host species is a complex task. Some studies examining gene-flow between *D. novemcinctus* populations are available to compliment those conducted for *D. marsupialis*, but no paired comparisons have ever been made (347, 348). Still, there is no evidence to suggest that *D. novemcinctus* is any less mobile within its range than *D. marsupialis*, to which its continuing rapid expansion in the US is testament (348). Furthermore, from the limited data available, geographic features like rivers, albeit shallow ones (Rio Paraguay), do not seem to represent significant barriers to gene-flow, and these animals can happily swim large distances (347). Perhaps, therefore, terrestrial TCIIc dispersal is limited with respect to that of arboreal TCI by some other factor, possibly vector behaviour or abundance, or even by the abundance of viable terrestrial secondary hosts.

Population assignment demonstrated extensive instability across almost all of the population groupings made on an *a priori* basis, with mean values for each population calculated by GENALEX equivalent to those achieved for then least cohesive populations

identified for TCI, and widespread re-partitioning on the basis of STRUCTURE. As with TCI, an AMOVA revealed, regardless of assignment methodology, that the majority of variation occured at the between-individual level, which may account for some instability. Furthermore, two levels of hidden population subdivision were observed. The first, involving tight clusters of isolates occurring within geographic foci, had no discernable biological or ecological basis. As mentioned previously, no specific link with a particular host or vector can be identified, and highly divergent isolates were also isolated from the same study sites. The second, most obvious from the reassignment of individuals in STRUCTURE, and arising from an apparent division among some SB isolates (SB2&SB3, Figure 70) could have an ecological basis. The connectivity between a subset of SB isolates and PA isolates may well have arisen from a shared ecotope across the Chaco region which extends from Southern Bolivia into Northern Paraguay (269). The remaining SB isolates had stronger affinities with NB isolates, and thus SB may represent a transitional zone between two (possibly three) parasite populations. The severity of this transition was minimal by comparison to that observed between the two TCI populations identified in Bolivia (COT&LB), and mirrors the gradual transition in ecotope from the moist lowlands of Northern Bolivia to the arid South. Additionally, unlike TCI, a common host, *D. novemcinctus,* is shared across all Bolivian TCIIc populations, a factor which is also likely to increase connectivity. Vector distributions could also play a role. *P. geniculatus* was identified as an associated vector in NB (Chapter 4), but, although efforts have been made to identify the silvatic vector of TCIIc in the Chaco in this and a previous study, none were found (277). Correlating with *gpi* sequence data, CAYMA19 was among those isolates re-assigned by both statistical and STRUCTURE based assignment to a grouping with NB isolates SJMC10 and SJMC4. Other STRUCTURE based re-assignments, however, like that of CAYMA25 to NB, were incongruent with the *gpi* phylogeny.

In general, pair-wise F_{ST} values between TCIIc were consistent with the strong IBD identified using other measures (i.e. D_{AS}). While gene-flow between silvatic TCI isolates was high between the tropical lowlands Northern Bolivia, Colombia, Venezuela and Brazil, no such observation can be made in TCIIc, with one exception. SJMC19, a NB isolate which showed split geographic affinities on the basis of *gpi* sequence analysis, was placed

firmly with the BVC isolates in all MLMT analyses. This observation hints at a possibly stronger link between Northern Bolivia and the rest of Amazonia.

The possible early radiation of TCI into arboreal ecotopes in association with didelphid marsupials was supported by the substantial pool of genetic diversity observed in these ecotopes with respect to outliers (Section 5.3.3). Fossil and molecular evidence suggests that the ancestry of *D. novemcinctus* is closely linked to Northern South America, and that of the *Euphractinae* to dryer ecotopes in the Southern Cone (Section 1.6.2.1) (149). Furthermore, the formation of many of these ecotopes seems to be a fairly recent event (145). If one expected an early emergence of TCIIc, perhaps in association with the *Dasypodidae* in humid tropical clines, these populations should show considerably higher genetic diversity. Instead, no major variation in allelic richness was observed between northern and southern populations.

Excess homozygosity, in common with most TCI populations, was observed in all TCIIc populations, further corroborating previous population genetic studies of *T. cuzi* (199, 215). It seems probable that a similar mechanism is responsible for generating this deviation from HW expectations. Again, efforts to estimate the extent of possible recombination were inconclusive. With the exception on NB, strongly significant multilocus LD was identified in all populations, and, as with TCI, levels of pair-wise physical linkage failed to correlate with statistical linkage at a subpopulation level, calling into question the potential for non-significant I_A values as calculated here to reflect true recombination. However, at the global level, some affect of physical linkage was observed, perhaps due to sample size, but also, if the number of multi-allelic loci/sample did give an approximate measure of multi-clonality (Chapter 6), because these were present as a far smaller proportion of the total dataset. Additionally, low sample size necessitated caution when interpreting the borderline I_A significance of NB, despite the fact that that the majority of isolates were clones. Clearly sample size was a key issue throughout the interpretation of the entire TCIIc dataset. Previous population genetic analyses of parasitic micro-organisms have drawn inference from populations made up of approximately ten individuals, but normally, like with TCI, through comparison with much larger related populations (266, 267, 388).

5.3.5 MLMT and single-locus sequence typing (SLST): a comparison or results for TCI and TCIIc.

The genetic resolution provided by MLMT far outstrips that provided by *gpi* sequence analysis in establishing intra-lineage population structure within both populations studied, perhaps unsurprisingly as microsatellite sequences are thought, in humans and other eukaryotes, to evolve several orders of magnitude faster than coding sequences (e.g. (207)). Additionally, while mutational saturation at microsatellite loci is likely compromise the inference of inter-lineage relationships, correlation between stepwise and infinite allele models, at least at a inter-population level (F_{ST} vs R_{ST}) in TCI, suggested that levels of saturation between populations of the same lineage may be low enough to infer evolutionary hypotheses from their inter-relationship, if sample sizes are large enough (Sections 5.1.2.1.3, 5.2.2.1.9, & 5.2.2.2.9). One principal drawback to microsatellite analysis is reproducibility between sequencing runs, which was assessed here using controls (Section 3.7.5). Conceivably some highly similar multilocus genotypes observed in TCIIc populations could actually represent identical strains. In view of the large number of loci employed, however, it seems likely that any error introduced should only have a marginal effect on the analysis.

Although *gpi* demonstrates far lower genetic diversity than the microsatellite dataset, a limited proportion of this variation correlates between the two datasets. In TCI this correlation was poorer, perhaps because these isolates show less spatial genetic structure, as defined by MLMT. In TCIIc, however, there was an approximate division between Northern South American isolates and those from the South, although far less distinct than is observed on the basis of MLMT.

5.4   Findings and Conclusions: a summary.

In this chapter over 200 isolates drawn from all major *T. cruzi* were analysed. In the first instance *gpi* sequence typing was compared to MLMT to asses their relative ability to distinguish the major *T. cruzi* lineages. Secondly, *gpi* sequence analysis was used to assess intra-lineage population genetic structure within TCI and TCIIc respectively. Finally MLMT typing of 49 loci was employed to tackle the level of population genetic structure in two expanded panels of TCI and TCIIc isolates. Briefly, the major findings and conclusions were:

<u>General</u>

- MLMT and *gpi* SLST can both resolve all major *T. cruzi* lineages (TCI, TCIIa-e), although saturation compromised MLMT analysis. In the future MLMT using selected loci might have a role as a rapid and fairly robust means of distinguish major lineages, especially TCIId and TCIIe, but should not be used to assess inter-lineage evolutionary relationships.

- *Gpi* SLST is not an adequate molecular tool for assessing intra-linage population structure in TCI and TCIIc, although some informative variation in TCIIc was noted. Nonetheless, SLST using more diverse/rapidly evolving sequence targets may have a future role.

- MLMT revealed remarkable structured diversity in both TCI and TCIIc

<u>TCI</u>

- As with other population genetic studies of *T. cruzi*, most populations showed excess homozygosity.

- Spatial genetic structure was weak among the majority of silvatic isolates.

- Arboreal, palm-associated lowland ecotopes harboured the greatest levels of genetic diversity, largely independent of the size of geographic focus, and independent of host/vector species.

- Possible population bottlenecks, based on the reduced genetic diversity in isolates spread over large geographic distances, were identified and

associated with transmission cycle (VD), host and ecology (COT+CH), as well as possibly geography (NA).

- Recombination may occur in palm-associated, diverse populations, but could be artefactual. Outlying, non-diverse populations show signs of tight multi-allelic linkage.

- Evidence of multi-clonality (multi-allelic loci) was identified in the majority of populations.

TCIIc

- As with other population genetic studies of *T. cruzi*, most populations showed excess homozygosity.

- Spatial genetic structure was stronger in TCIIc isolates that that identified in TCI, not necessarily as a result of differential primary host population dynamics, but possibly as a result of differential vector or secondary host dynamics.

- Levels of diversity in some TCIIc populations were similar to those found in diverse TCI populations, also independent of host/vector species, but no 'pools' of diversity were found with respect to particular ecotopes.

- Some evidence of recombination was found among all TCIIc populations, but could be artefactual. Global analysis of the entire dataset, however, did identify a link between physical and statistical linkage.

- Low sample size distributed across all populations was deemed to compromise inference from the TCIIc dataset.

- Evidence of multi-clonality (multi-allelic loci) was identified in half of all populations.

6 Results: TCI multiclonal intra-host population infrastructure.

6.1 Samples analysed and loci employed.

Eight TCI samples, all derived from silvatic mammals, were selected for analysis of clonal population infrastructure at the within-host level (Table 48). Individual samples were selected from populations known to harbour a high number of isolates demonstrating multiple alleles at individual loci (Chapter 5, Table 37). The original number of multi-allelic loci identified in each un-cloned sample is shown in Table 50. Two isolates demonstrating no multi-allelic loci in the un-cloned state were also included, to assess the potential for 'hidden' multi-clonality.

Table 48 Eight TCI samples cloned to assess level of intra-host multi-clonality.

Host	Sample Code	Date	Locality	Latitude*	Longitude*	Isolation Method[a]	Source[b]
Didelphis marsupialis	XE5167	14.09.99	Para, Brazil	-1.71	-48.88	XE	EC
Didelphis marsupialis	XE5740	18.6.02	Para, Brazil	-1.383	-48.86	XE	EC
Didelphis marsupialis	M13	12.6.04	Barinas, Venezuela	7.5	-71.23	B	TS
Didelphis marsupialis	M16	13.6.04	Barinas, Venezuela	7.5	-71.23	B	TS
Didelphis marsupialis	M18	13.6.04	Barinas, Venezuela	7.5	-71.23	XE	TS
Didelphis marsupialis	M7	14.5.04	Barinas, Venezuela	7.5	-71.23	B	TS
Didelphis marsupialis	SJM34	7.9.04	Beni, Bolivia	-14.81	-64.6	XE	TS
Philander opossum	SJM41	9.9.04	Beni, Bolivia	-14.81	-64.6	B	TS

*Decimal degrees.
[a]XE - Xenodiagnosis, B – Direct inoculation of whole blood in culture.
[b]EC – Institute Evandro Chagas, Brazil, TS – This study.

A panel of nine loci were selected from those employed previously to investigate TCI population genetic structure. Loci were deliberately selected as those at which multiple alleles were present among the un-cloned samples under investigation. As such these nine loci represent a non-random, hyper-variable, sample of the forty-eight initially employed.

Table 49 Nine hyper-variable loci selected for analysis of intra-host multi-clonality.

Primer Code	Contig Number[a]	Position[b]	Repeat	Forward Primer/Reverse primer
1FAMP7c	7206	16210-16323	(TA)n	CCAGTTTCACACATACGCAA
				CGTTTGGAGGAGGATTGAGA
1FAMP9c	8370	24890-25024	(TTG)n	ATGGGTGCGAGAGGTATGTC
				TGTCAAAACAGCGGAAAGAA
1TETP12a	7118	5369-5576	(TA)n(GA)n	TGTGATCAACGCGCATAAAT
				TTCCATTGCCTCGTTTTAGA
1TETP4a	8328	71097-71226	(GA)n	GTCACACCACTAGCGATGACA
				ACTGCACAATACCCCCTTTG
1TETP4b	8328	32430-32629	(TTA)n	GAGAGAGATTCGGAAACTAATAGC
				CATGTCCCTTCCTCCGTAAA
famp6a	7143	119499..119520	(TG)n	TCGTTCTCTTTACGCTTGCA
				TAGCAGCACCAAACAAAACG
famp8a	8646	88680..88708	(TCG)n	ACCACCAGGAGGACATGAAG
				TGTACACGGAACAGCGAAG
famp8b	8646	12639-12819	(TA)n	AACATCCTCCACCTCACAGG
				TTTGAATGCGAGGTGGTACA
vicp5a	7853	72728-72864	(CT)n(TA)n	GCAGAGACGCACAGACACAT
				AAAGTGCCATCCCACCCTC

[a] Refers to a syntenous fragment of sequence identified by the *T. cruzi* genome project
[b] Refers to the position of the of the primer binding site, flanking region and repeat motif 5'-3' along the sense strand.

6.1.1 Clonal diversity

Two or more multi-locus genotypes (MLGs) were identified in all populations, including ones for which no multiple alleles were identified in the original un-cloned sample, suggesting a degree of 'hidden' multi-clonality (Table 50). Indeed one of the highest number of distinct MLGs was, paradoxically, observed in an isolate that showed no evidence of multi-allelic loci in the un-cloned state (XE5167, N=10). More predictably, M13, for which seven loci in the un-cloned state were multi-allelic, also demonstrated a high number of MLGs (N=10). Furthermore, XE5167 seems to be an outlier and, excluding this sample, a linear regression demonstrates that the number of multi-allelic loci in the original un-cloned sample is a fairly robust predictor of the number of MLGs in the resulting clonal infrapopulation ($R_{XY} = 0.838$, $R^2 = 0.703$, p=<<0.05).

Table 50 Infrapopulation multi-clonality, mean allelic richness, and ploidy.

CODE	Mammal	Original MAL	N	G	MNA	Clone MAL
XE5167	*D. marsupialis*	0	25	10	3.4	4
XE5470	*D. marsupialis*	2	33	4	2.1	0
M13	*D. marsupialis*	7	29	10	4.8	5*
M16	*D. marsupialis*	0	14	3	1.7	0
M18	*D. marsupialis*	5	32	8	4.5	1
M7	*D. marsupialis*	4	32	7	4.5	2*
SJM34	*D. marsupialis*	4	14	3	3.4	1
SJM41	*P. opossum*	2	32	2	1.5	0
			211	49	3.4125	

Original MAL – Multi-allelic loci present in un-clones sample.
N – Number of clones analysed.
G – Number of multi-locus genotypes.
MNA – Mean number of alleles/locus in infra-population (clone-corrected).
Clone MAL – Number of clones containing multi-allelic loci.
*Includes identical aneuploid genotypes represented more than once in the dataset.

In all eight infrapopulations studied, the presence of a majority or dominant genotype was identified, and providing strong evidence for the reproducibility of the typing technique. Specifically, in all but two infrapopulations, a single majority MLG was present at a greater frequency than all others combined (Figure 73). In M13 and M18, a single MLG, while not occurring in the majority, did have a tendency to dominate, but overall a more balanced clonal composition was observed.
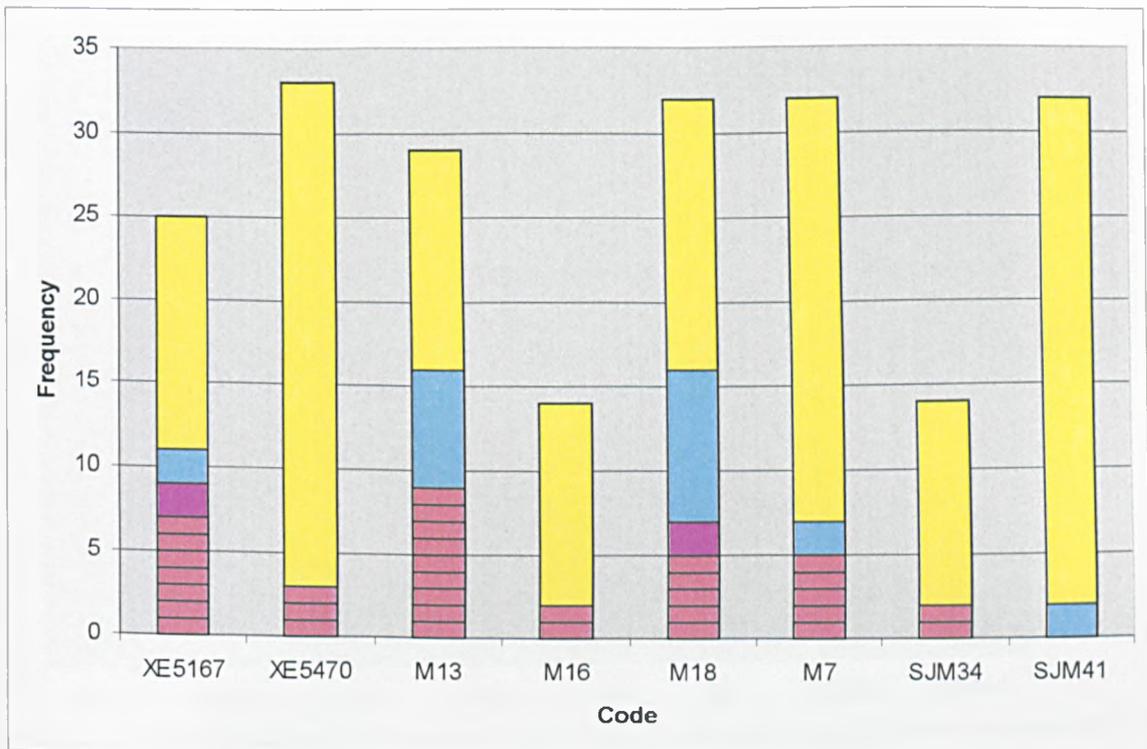
Figure 73 Clonal composition of different within-host infrapopulations. Composite bars indicate frequency of different MLGs per isolate. Each division represents a different MLG frequency class. Yellow bars indicate dominant/majority MLGs (N=2-30), blue and purple bars MLGs at lower frequency (N=2-9), and pink bars MLGs represented by a single clone.

After clone correction, levels of allelic diversity (MNA) were astonishingly high in some infra-populations (e.g M13, MNA = 4.8), equivalent to those observed in some of the more diverse populations TCI studied in Chapter 5, (e.g. LB, MNA= 4.67). Some relationship between allelic richness (MNA) and population of origin was observed (i.e. LB, NEB, VS, Chapter 5). For example, isolates from VS, the most allelically rich at a population level, were similarly the most diverse under the same measure at the infrapopulation level. The relationship between MNA and the number of distinct MLGs per population was related to MLG divergence as well as abundance, and is better understood in the context of mean pair-wise D_{AS} values later in this section (Section 6.1.2).
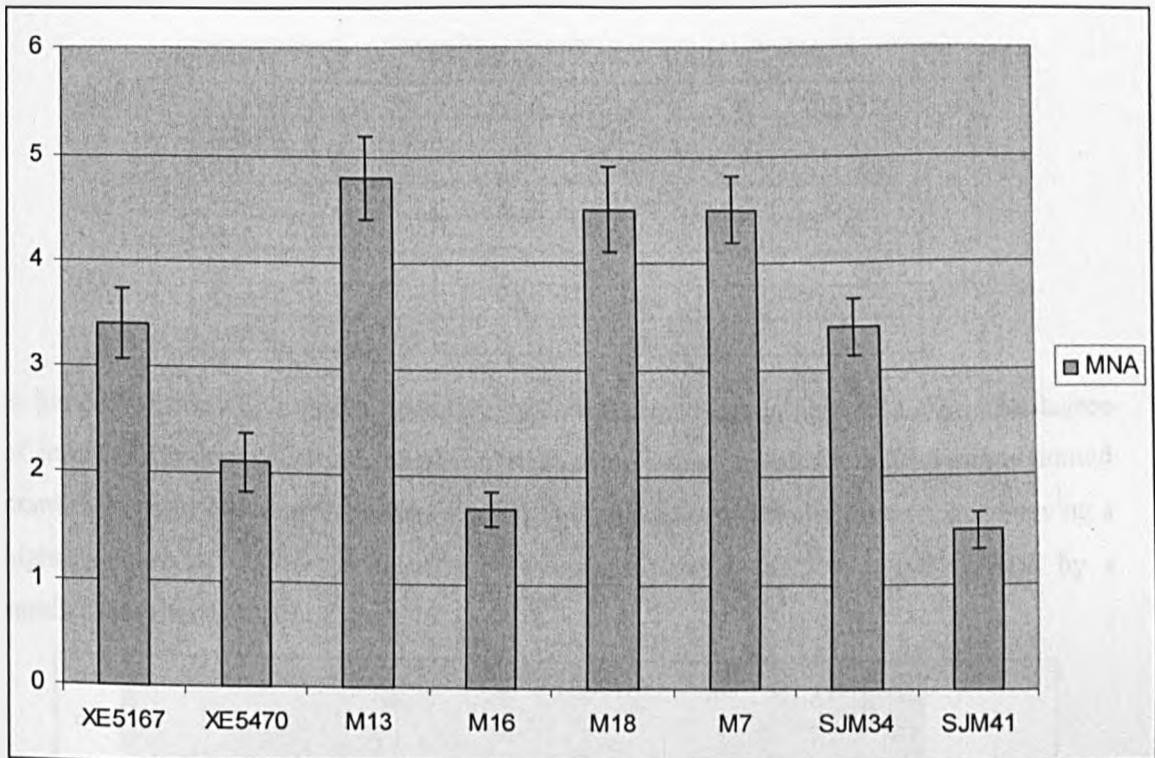
Figure 74 Allelic richness (MNA) among clones drawn from different silvatic mammals. Error bars represent +/- standard error about the mean.

Intriguingly, 5/8 infra-populations analysed contained clones with multi-allelic loci, with some MLGs multi-allelic at up to 8/10 loci and possibly indicative of anueploidy. Only two such MLGs, however, where present at a frequency greater than one, both of which contained only a single multi-allelic locus (M7, N=2; M13, N=13, Table 50). The implications of possible aneuploidy in a subset of clones are dealt with fully in Section 6.1.4.

6.1.2 Clonal divergence

Clonal divergence was measured as a function of the mean pair-wise D_{AS} distance between all MLGs in the infrapopulation (Table 51). Each mean was calculated as in Section 3.7, using 1000 replicates of the randomisation technique previously employed to deal with multi-allelic loci in Chapter 5.

Table 51 Mean D_{AS} values for all eight infrapopulations drawn from silvatic mammals.

CODE	G	Mean D_{AS}	SE
XE5167	10	0.399	0.035
XE5470	4	0.295	0.100
M13	10	0.509	0.042
M16	3	0.067	0.017
M18	8	0.523	0.053
M7	7	0.591	0.074
SJM34	3	0.583	0.267
SJM41	2	0.056	n/a

In general terms, MLG number positively correlated with mean pair-wise D_{AS}. The degree of inter-MLG identity varied, however, with some infrapopulations comprising a limited number of highly divergent strains (e.g. SJM34) and others, clonally diverse, but showing a higher degree of multilocus identity between strains (e.g. XE5167), and typified by a smaller margin of error in Figure 75.
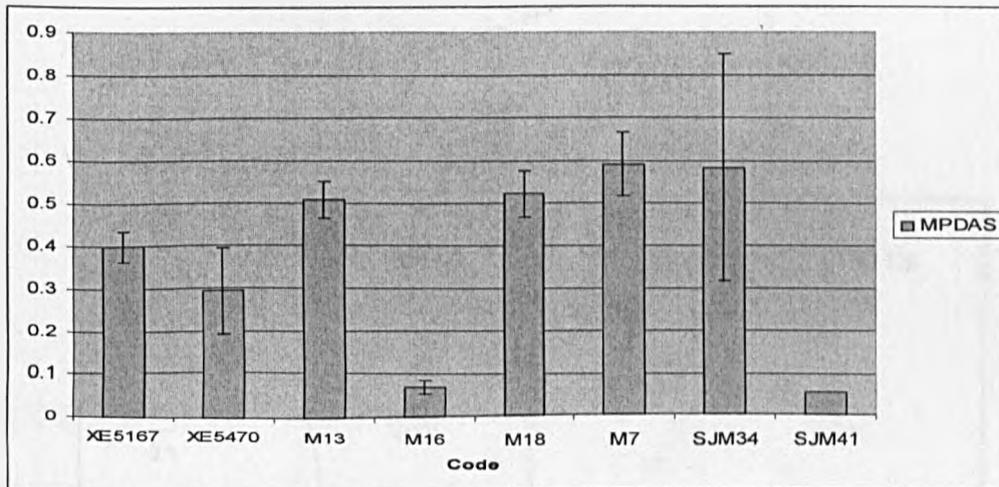
Figure 75 Mean pair-wise D_{AS} values (MPDAS) over all pair-wise comparisons between MLGs within each infrapopulation. Error bars represent +/- standard error about the mean.
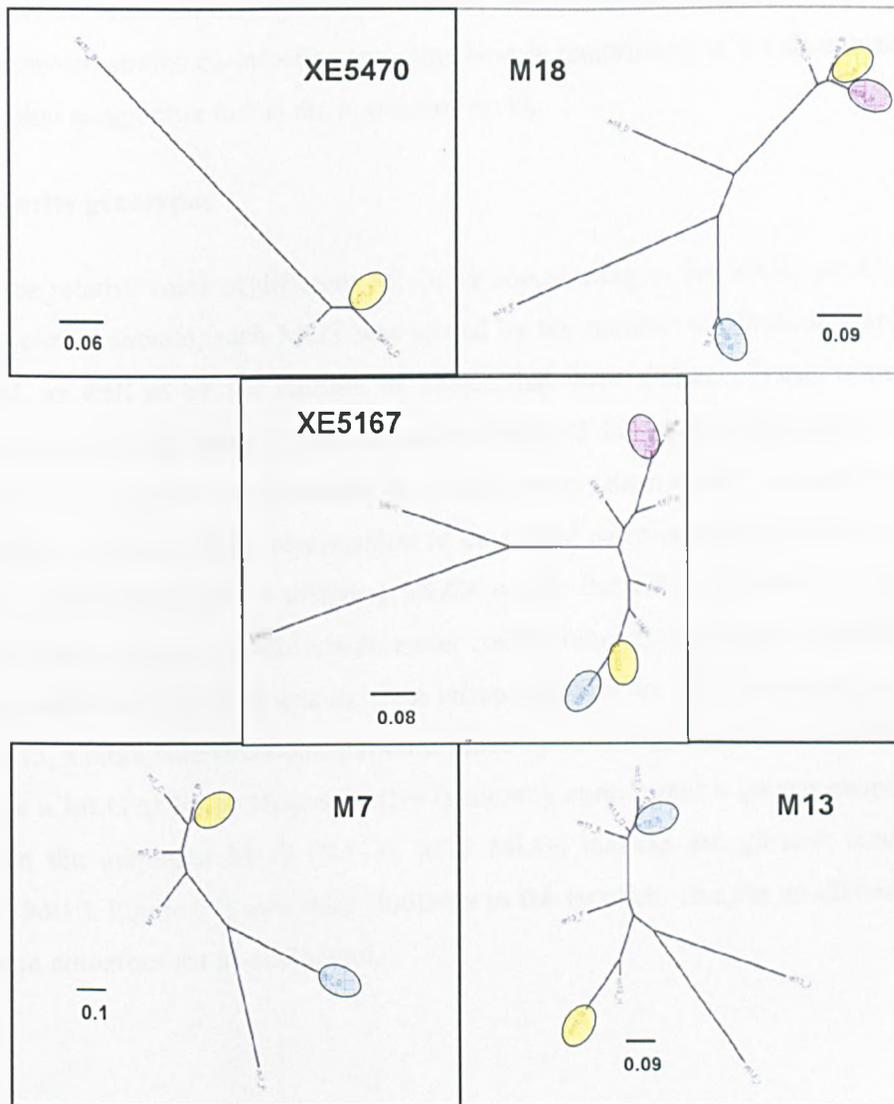
Figure 76 Un-rooted NJ trees based on pair-wise D_{AS} values between MLGs from five parasite infrapopulations where >4 MLGs were identified. D_{AS} values were calculated as in Section 3.7.6.1.2 in MICROSAT and PHYLIP v3.6 (282, 290). Coloured ellipses represent genotypes at a frequency >1 and correspond to Figure 73.

In most cases, infrapopulations comprised a group of closely related strains in conjunction with a limited number of diverse outliers (Figure 76). In four out of five infrapopulations (excluding M13), the majority genotype clusters among closely related strains. Deep branching between strains co-infecting the same host is reminiscent of the deep branching observed within geographic foci at the population level.

6.1.3 Majority genotypes

To assess the relative roles of different MLGs in contributing to the allelic profile of the original un-cloned sample, each MLG was scored by the number of alleles it shared with the original, as well as by the number of alleles that were distinct. These values were combined to provide an index of relatedness and plotted for each MLG (Figure 77). In general terms, the majority or dominant MLG was always most closely related to the un-cloned profile. Thus the allelic composition of un-cloned microsatellite profiles is heavily dependent on the frequency of different MLGs within the infrapopulation, representing closely the highest frequency MLG, with lesser contributions from those at low frequency. In the rare event that MLG frequencies in the infrapopulation are more balanced, as was the case for M13, a more balanced contribution is made by each MLG to the un-cloned profile. In this case a MLG at lower frequency (N=7) actually contributed a greater proportion of alleles than the dominant MLG (N=13). M13 MLGs making the greatest contribution (Clone ID 9&10, Figure 77) owe their similarity to the fact that, like the un-cloned profile, both contain numerous multi-allelic loci.
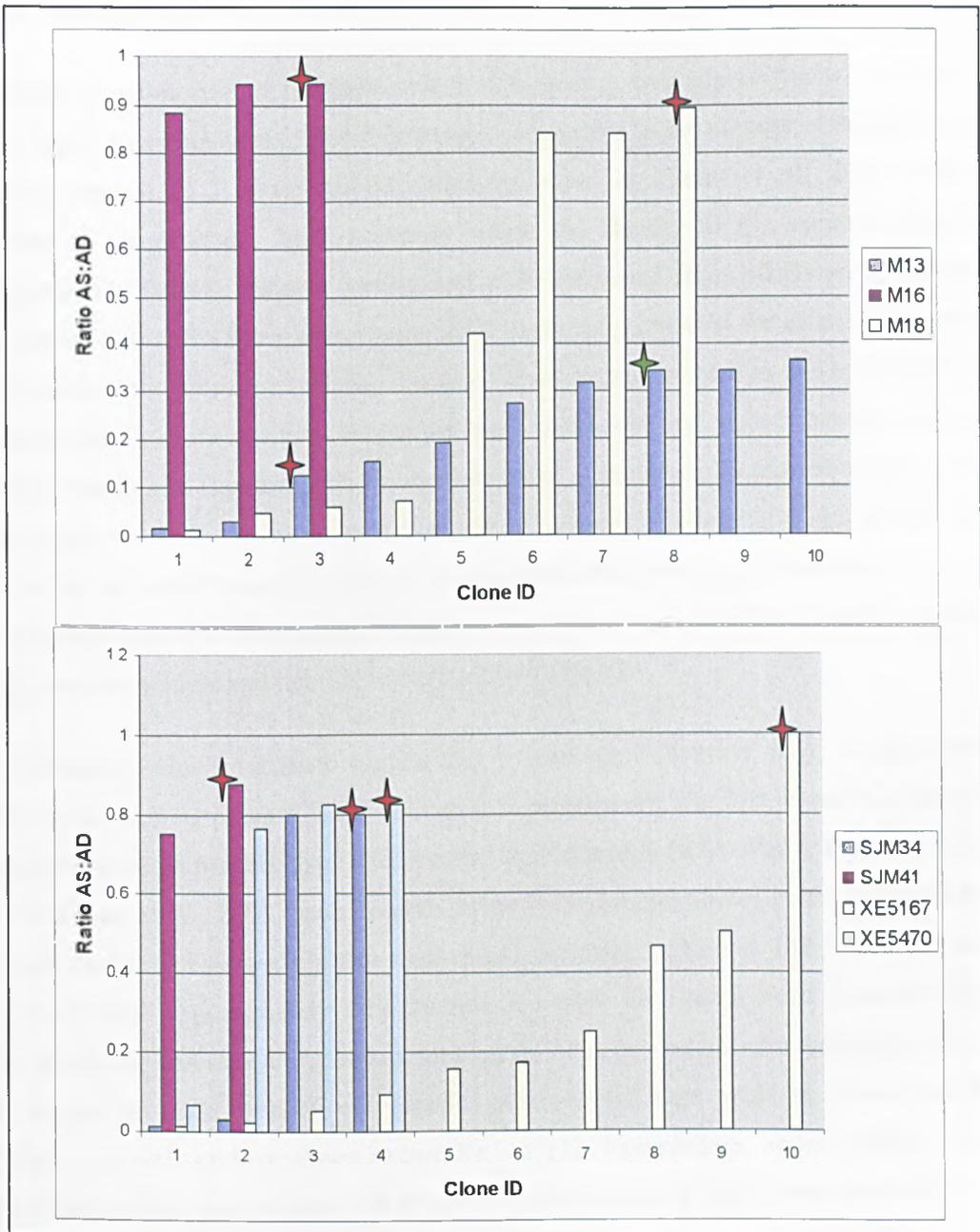
Figure 77 Bar plot demonstrating the relatedness of each MLG to the original, un-cloned multi-allelic profile. Majority or dominant genotypes are designated by red stars. Secondary (N>2) genotypes are represented by green stars. The ratio represented on the Y axes was calculated as the proportion of alleles shared with the un-cloned profile over the number of distinct alleles.

6.1.4 Recombination

To assess possible role of recombination in the dataset, two approaches were taken. The first involved examination of possibly aneuploid samples in the context of the fusion-then-loss mechanism of *T. cruzi* recombination proposed by Gaunt *et al.*, 2003 (183). By contrast to Gaunt *et al.*, 2003, however, where the identity of the parental clones was unequivocally known, putative parents had to be estimated from MLGs recognised to be circulating within the same infrapopulation. The primary criterion for estimating parentage was based on the proportion of shared alleles between putative parents and hybrid(s). Under a fusion-then-loss mechanism, the occurrence of alleles in one or both parents that do not occur in the hybrid is permissible; indeed one set of parental alleles may be missing entirely at a single locus. Assuming no mutation, the occurrence of alleles within a hybrid that do not occur in either parent, however, is not permissible. One such aneuploid profile was indentified from M13 and excluded from this analysis. A complete list of allelic profiles for both putative parents and hybrids is included in Table 52.

Two lines of evidence exist to suggest that at least some 'hybrids' may, in fact, represent error in the cloning process (i.e. 2+ clones may be present). The first relates to relative peak intensity between parental-type alleles in the hybrid profile (XE5167H1, Trace C, Figure 78). As in Gaunt *et al.*, 2003, peak heights across both parental alleles in an aneuploid hybrid should be fairly balanced. In this case, peaks for alleles 132 and 138 (derived from one 'parent') were approximately 1/10 the size of a peak for 126 (derived from the other). If this profile represented a true aneuploid hybrid, then copy number for each allele should be in a ratio of 1:1:2, respectively. Similar observations were made at across loci in this 'clone', as well as to a lesser extent XE5167H2. Nonetheless, approximately balanced parental-type hybrid profiles with respect to their expected ratios were observed in three other aneuploid MLGs (M13H1, M7H1, XE5167H3, Table 52 e.g. Traces A&B, Figure 78).

A second line of evidence relates to linkage between loci thought to have undergone random allelic loss. In Table 52, XE5167H2 has apparently undergone loss of both alleles from one parent (132&138) at locus famp8a. However, no such loss appears to have occurred at an adjacent locus 76kb distant on the same contig, famp8b. Although no figure has been placed upon the expected size of fragments and/or chromosomes lost post-fusion

in *T. cruzi*, it seems reasonable to suggest that relatively closely linked loci should, more likely than not, be lost simultaneously (183). Instead this observation may also represent an artefact of cloning error, possibly involving differential amplification of loci between clones, such that the amplification products for the missing alleles were present, but not at detectable levels.

Table 52 Putative parents and hybrids identified from different infrapopulations under a fusion-then-loss model of genetic exchange.

Primer/ linkage group	M13 P1	M13 P2	M13 H1	M7 P1	M7 P2	M7 H1	XE5167 P1	XE5167 P2	XE5167 H1	XE5167 H2	XE5167 H3
1TETP12a		149	149	149	149	149	147		147	147	147
7118	153		153				153	153	153	153	153
	157		157					155	155	155 L	
Famp6a	138	138	138		136	136	140	140	140	140	140
7143		140	140	138		138					
	142		142	140	140	140					
1FAMP7c	105	105	105	105		105	105	105	105	105	105
7206		107	107	107	107	107		111	111	111	111
	115		115		109	109					
vicp5a		143	143	147		147	145	145	145	145	145
7853		147	L	154	154	154	150	L	L		150
		152	152		156	L					
	154		154								
	156		156								
1TETP4a	111		111	113	113	113	103	103	103	103	103
8328		113	113	117		L	109	109	109	109	109
		115	115								
	117		117								
1TETP4b	167	167	167		167	167	164	L	L		164
8328	179	179	179	170		L	173	173	173	173	173
				176	176	176					
1FAMP9c	97		97	108	108	ND	110		110	110	110
8370	108		108	112	112						
		112	112					112	L	L	
famp8a	126	126	126	132	132	132		126	126[a]	126*	126
8646	129	129	129	132	132	132	132		132[a]	L	132
	132		132				138		138[a]	L	138
famp8b	175	175	175		171	171	173		173	173	173
8646		179	179	175	175	175		175	175	175	175
				177		177					

L – Indicates loss of an allele of allele with respect to putative parents.
ND – No data.
* Possibly incompatible with true hybridisation, see text.
Numbers in bold correspond to parental allele profiles.
[a] Possibly incompatible with true hybridisation, see Figure 78.
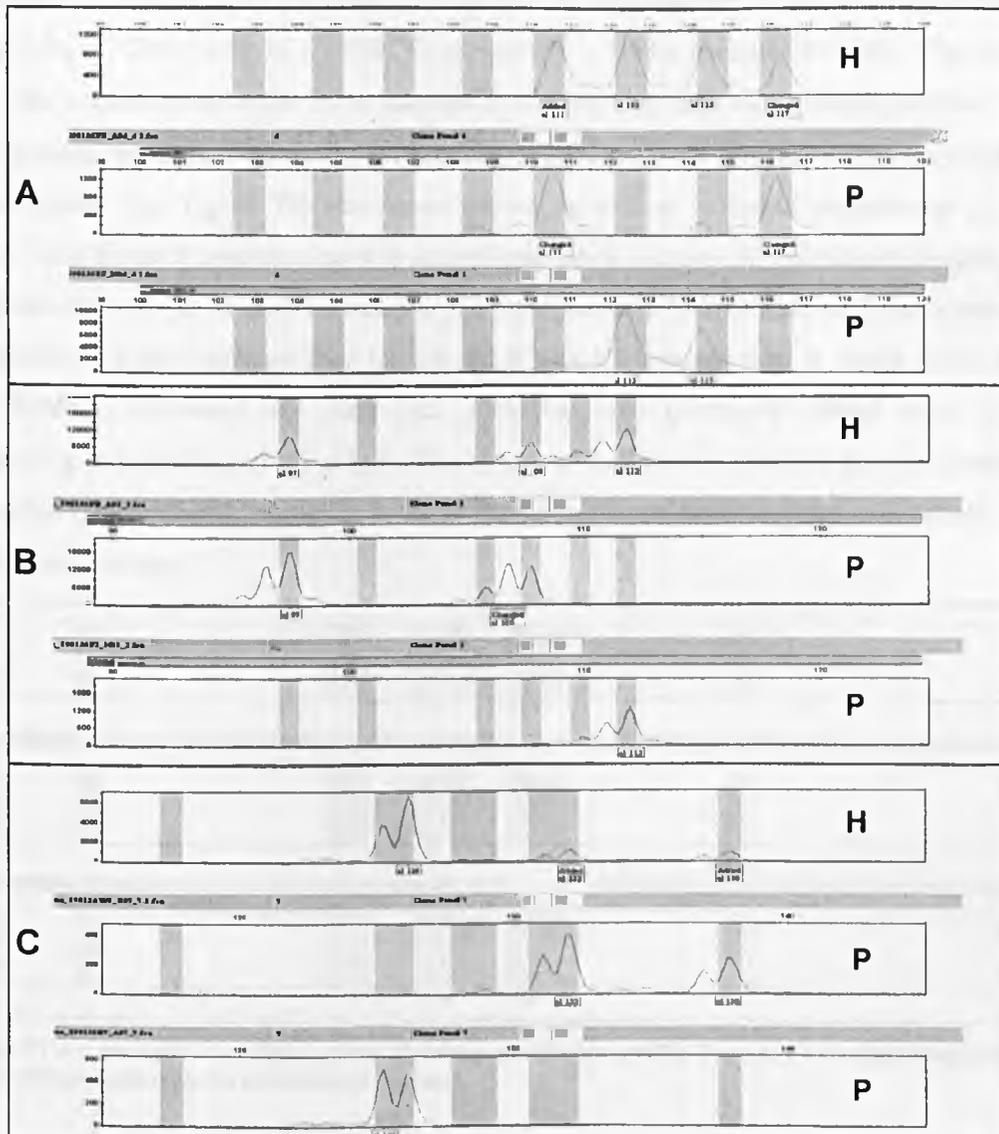
Figure 78 Selected putative parental and hybrid microsatellite traces identified within MLGs drawn from different mammalian hosts. A – M13H1, 1TETP4a. **B** – M13H1, 1FAMP9c. **C** – XE5167H1, famp8a. **H** – putative **hybrid, P** – putative parent. Imbalanced peak intensities in trace C indicate a possible artefact of cloning error, see text for details.

In the second approach employed to examine possible recombination in the dataset, diploid MLGs from the same geographical population (i.e. NEB, LB and VS) were examined for evidence of mutlilocus linkage disequilibrium, correcting for known linkage groups, using the Index of association, I_A, in MULTILOCUS v1.3, as in Section 3.7.6.1.4 (300). In doing so the dataset was also expanded to include several clones analysed in Chapter 5. A full table of allele profiles for all 41 clones examined across the nine loci is included in the

appendices (Table 55). All populations displayed strong evidence of linkage disequilibrium (Brazil, I_A – 2.384, Beni, I_A - 3.559, Venezuela, I_A – 1.984, p-values <<0.05). The dataset was also examined manually in an attempt to identify any rare recombinant profiles. Two observations were made. Initially the number of heterozygous profiles with corresponding homozygotes (e.g. Figure 79) was scored across, as well as within all populations at each locus. Each distinct combination was scored only once. Across all loci across populations, 28 instances were identified, equating to 3.11 per locus. Within populations the number of instances were six, four and four for LB, NEB and VS respectively. A single clone from LB, SJMC3, contained six such loci, including two physically linked locus pairs, suggesting a possible hybrid origin. However, no multilocus putative parents could be identified from within LB or indeed across the dataset, even assuming that some allelic loss might have occurred.
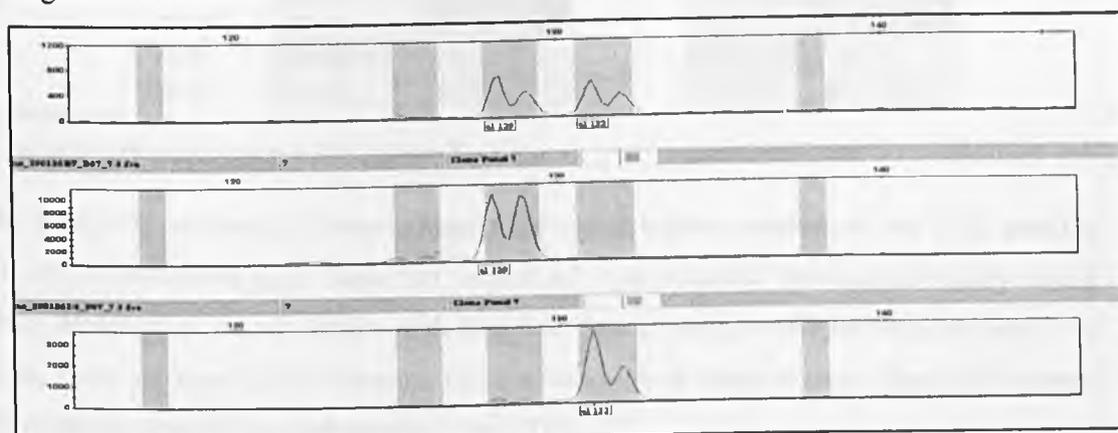
Figure 79 An example of a heterozygous diploid microsatellite profile (top) and corresponding possible parental-type homozygotes (middle and bottom).

The second observation made involved the identification of possible mosaics among physically unlinked, as well as physically linked, locus pairs. Initially, this relationship was examined statistically using r^d. In LB, NEB and VS, 72.2%, 91.6% and 80.55% of all locus pairs demonstrated non-random association respectively, in each case including those physically linked. Secondly, all locus pairs were examined manually for rare break-downs in linkage between MLGs within populations. These were defined as cases were a diplotype at one locus was associated with two or more different diplotypes at another locus within each population. Three examples are provided in Table 53. To reduce the confounding effect of mutation, events were only scored when both alleles at the locus under comparison

were distinct. Finally locus pairs were assessed within populations to detect 'full mosaics', such that all four possible combinations of diplotypes were present.

Table 53 Three examples of possible break down in pairwise linkage within populations.

	7118[a]	7853[a]		7853[a]	8328[a]		8646[a]	8646[a]
Primer Code	1TETP12a	vicn5a		vicn5a	1TETP4b		famp8a	famp8b
A[b]	136/136	147/147	C[b]	143/143	164/164	E[b]	129/129	171/171
B[b]	136/136	150/154	D[b]	143/143	173/173	F[b]	129/129	179/179

[a] Linkage group
[b] Arbitary codes for different MLGs

Table 54 Two 'full' mosaics identified between two primer pairs.

	7206[a]	8370[a]	7853[a]	8370[a]
	1FAMP7c	1FAMP9c	vicn5a	1FAMP9c
5470b[b]	105/107	110/110	143/143	110/110
5167d[b]	105/111	110/110	145/145	110/110
5167i[b]	105/107	112/112	143/143	112/112
5167f[b]	105/111	112/112	145/145	112/112

[a] Linkage group

[b] MLG codes correspond with those in Table 55, Appendices.

In total, 134 instances of non-linkage were noted within populations out of a possible 10,075 comparisons (e.g. Table 53). Eleven of these occurred between physically linked loci. Multi-locus mosaic MLGs, including both hybrid and parental profiles, and across all loci, were not noted. Two instances of possible pairwise mosaics were identified between four MLGs from NEB involving three loci (Table 54).

6.2 Discussion

This study represents the most comprehensive attempt to date to document the level of infrapopulation genetic diversity in *T. cruzi* reservoir hosts infected with a single major lineage, albeit limited to two different host species, *D. marsupialis* and *P. opossum*. Mixed major-genotype infections have been identified from silvatic reservoir hosts in Chile, on the basis of dubious genotyping methodologies, and in Piaui state in Brazil, based on mini-exon gene region PCR-FLP, but evidence of diversity at the intra-genotype level was not observed due to low genetic marker resolution (223, 372). The occurrence of mixed major genotype infections in silvatic triatomines has also been observed, furthermore Yeo *et al.*, 2007 were able to demonstrate the presence of a two distinct TCI genotypes on the basis of

RAPD polymorphism among 20 clones taken from *R. robustus* in Venezuela (276, 370). Efforts to document mutliclonality in naturally occurring *T. cruzi* strains using microsatellites have also been made, and multiple infections with distinct intra-genotypic variants within wild reservoirs, triatomines, and humans were identified (215, 217). In both studies, however, a maximum of only six clones were analysed from any one strain. Additionally, only strains demonstrating multi-allelic MLGs were cloned, and diploid, uncloned strains were assumed to be clonal in composition, a strategy which has also been employed in the context of *T. brucei* population genetics (236). This study demonstrates a fundamental flaw with this assumption, where, although in a minority, strains containing no mutli-allelic loci can in fact harbour numerous distinct MLGs.

It has been proposed that passage from the *in vivo* to the *in vitro* environment is likely to introduce a level of selection, such that a reduction in clonal diversity may be observed (215). This hypothesis is inherently difficult to test without direct analysis of parasite clones from host tissues, while attempts to clone parasites directly from silvatic mammal whole blood have been frustrated by low circulating parasitemia (276). Nonetheless, according to the limited evidence gathered here, some clonal diversity can be maintained in culture despite frequent repassage and periodic cryopreservation. Additionally, no obvious correlation between method of isolation (xenodiagnosis or direct blood inoculation) and MLG diversity was observed. Although clonal diversity may not be modulated by isolation technique, it possibly that clonal identity could be. Some tenuous evidence exists to suggest that different major lineages are observed in chronic human cases according to different isolation technique, but again this could represent a flaw in genotyping methodology (403).

Infrapopulation divergence and diversity statistics are difficult to interpret outside of any ecological or pathological context. Infrapopulation subdivision, at a major lineage level at least, has been recently noted using microsatellites to detect circulating strains in a HIV+ Bolivian patient. Here TCI was exclusively identified in the cerebrospinal fluid, while TCI and TCIId/e were found in circulating blood (404). Tissue specific TCI diversity in silvatic mammalian reservoir species represents an interesting avenue for future study, as are possible temporal fluctuations in infrapopulation MLG composition, even though no specific biological framework, beyond pseudocyst rupture cycles, for such variation currently exists for *T. cruzi*, unlike in *T. brucei* (405). In the meantime is of note that,

among a subset of clones from some strains at least, MLG divergence was extremely high, and mirrors the level of divergence observed at a population level. Still, a degree of clustering was observed even after clone correction, and could represent superinfection by closely related clones, or mutation within the host, although this seems unlikely given the limited lifespan of both marsupial species studied (156). Microsatellite reproducibility could also lie at the heart of this observation, but the widespread recovery of completely identical MLGs does indicate that the technique is fairly robust.

The diversity uncovered here likely represents the tip of the iceberg with respect to the total present: some selection may have occurred during the isolation and culturing of strains, only circulating parasite was sampled at a single point in time, and finally only a limited number of clones were examined from each strain. Nonetheless, in selecting loci known to be hypervariable at population level, a degree of fixed infrapopulation microsatellite variation may have been ignored, and overall levels of diversity, therefore, overestimated. Furthermore, if some multi-allelic MLGs arose from errors in the cloning process, total MLG number may be artificially enhanced, while estimates of allelic richness will remain unaffected.

A key aim of this analysis was to asses the possible impact of multiclonality on the population genetic analysis undertaken in Chapter 5. Most significantly it was shown that a single MLG predominated in each of the eight infrapopulations studied, and in seven populations the dominant MLG bore the closest resemblance to the uncloned profile. Thus, by reference to this study, the uncloned mutilocus microsatellite profiles employed in earlier population genetic analyses may represent closely the most abundant MLG in circulation, with more minor contributions from rarer MLGs, rather than a balanced mixture of all MLGs present (Chapter 5). In population genetic studies of *P. falciparum*, minority haplotypes (as apposed to mixed) are excluded from the analysis of blood stage forms on a quantitative basis, involving comparison of the relative chemofluorescent intensities of different DNA-DNA probes made for known circulating SNPs (406). Given that *T. cruzi* is generally assumed to be diploid, while blood stage *P. falciparum* is haploid, and it is possible that amplification efficiency may vary between alleles at the same locus even in clones, it seems difficult to envisage how a similar strategy could be deployed, apart from in extreme cases (e.g. Figure 78, Trace C)(4). Rare MLGs may be confounding

in population genetic terms, as they are a non-representative sample of overall parasite genetic diversity circulating in a given area. Simplistically, it seems feasible that merely biologically cloning every single isolate is the best approach to eliminate rare, background alleles. However, in this study the overall ratio of minority to majority MLGs was approximately 1:3. Thus, if only a single clone from each sample was taken, approximately every third clone would be unrepresentative of the majority genotype, and, in reality, a minimum of three clones should be taken and analysed from each sample to achieve an accurate result, an extremely laborious task. Thus microsatellite analysis of uncloned parasite strains, accompanied by a software approach to deal with multiple alleles, as detailed in Chapter 5, is likely to be the most unbiased approach to sampling each host infrapopulation. In doing so, however, the accuracy of estimation of some population parameters may be affected. As discussed in Chapter 5, caution must be taken when interpreting linkage statistics, as mixed infections can mimic the effect of recombination. Also, mutliclonality may also marginally increase the observed number of heterozygotes. Preliminary analysis of diploid MLGs presented in Section 6.1.4 (also see Table 55, Appendices) indicates that within-population excess homozygosity is also a feature of the clonal dataset from LB, VS, and NEB (data not shown).

At face value, aneuploid allelic profiles presented here provide persuasive evidence of hybridisation similar to that encountered by Gaunt *et al.*, 2003 (183). Additionally, the reservoir host seems a likely site of recombination in the light of this previous study, given that experimental hybrids were derived after passage through Vero cell lines. Aneuploid 'hybrid' and diploid or aneuploid 'parental' MLGs were identified in three infrapopulations, and in one, XE5167, three hybrids were identified for the same set of parents demonstrating various degrees of allelic loss. Multiple locus 'aneuploids' were in a minority, however, and, as discussed in Section 6.1.4, these could fall within the margin of error of the cloning process. The efficiency of the solid media cloning technique has only been tested once in identifying a polyclonal infection of the same major lineage, and found variation on the basis of a single RAPD primer in a single clone (1/20 clones) (276). Inoculum cell density was carefully monitored in this study to prevent growing colonies merging and the inoculum cell culture kept at log phase to minimise clumping. However, even with the most assiduous effort to ensure that colonies are derived from a single

founding organism, some clumping may still occur. The likelihood of picking a mixed genotype colony is probably also dependent on the number of clones picked, as well as the relative frequencies of distinct MLGs in the infrapopulation. Nonetheless, three 'aneuploid' MLGs demonstrated balanced amplification across alleles at individual loci, as well as no linkage anomalies (Section 6.1.4), and could be recombinants. Further analysis is needed to confirm this observation.

Once 'aneuploid' MLGs had been excluded, widespread LD, on the basis of I_A and pairwise r^d values was observed among the remaining diploid genotypes, including additional clones from the analysis in Chapter 5, in line with the widely held belief that natural *T. cruzi* populations reproduce clonally (228, 230). If recombination is rare in *T. cruzi*, however, both linkage statistics employed may fail in its detection. Manual examination of these profiles provided tantalising evidence of possible recombination, in the form of corresponding homozygotes and heterozygotes at individual loci, as well as apparent breakdowns · in linkage between pairs of loci, and two possible mosaics. Unfortunately, however, the intrinsically high rate of microsatellite mutation could well act to confound the detection of recombination, either by mimicking the Mendelian-type and mosaic-like patterns, if not frequencies, of allelic diversity, or by rapidly erasing evidence of past recombination.

6.3 Findings and conclusions: a summary.

- The presence of multiple and divergent MLGs within many TCI strains isolated from wild mammals supports previous evidence that multiclonality is likely to be a widespread phenomenon in naturally occurring *T. cruzi* populations.

- Un-cloned TCI strains that demonstrate no multi-allelic can harbour substantial 'hidden' diversity, although there is a general correlation between the number of un-cloned multi-allelic loci and the level of multiclonality in the associated infrapopulation.

- In almost all populations, the presence of a majority genotype was detected.

- Aneuploidy detected in three clones, and corresponding with circulating diploid 'parent' MLGs, may be consistent with fusion-the-loss hybridisation, while others could represent error in the cloning process.

- High levels of multilocus and pairwise linkage disequilibrium were detected among diploid MLGs isolated from the same geographical area. Some patterns of variation consistent with genetic exchange were observed on manual inspection of the data, but could also have arisen through mutation

7   General discussion

The principal aim of this project was to establish the potential of microsatellites as a tool to investigate the molecular epidemiology of *T. cruzi*. First, the 'state-of-the-art' with respect to the current understanding of the distribution and ecology of major silvatic *T. cruzi* genotypes in the Americas was established, including new isolates collected during fieldwork undertaken in this study in both Venezuela and Bolivia. Secondly, a panel of microsatellite markers was developed, with nuclear sequence data for comparison, to resolve the six major *T. cruzi* genotypes, and latterly expanded to resolve the molecular epidemiology of two major silvatic genotypes: TCI and TCIIc, using samples drawn from across their distribution in the Americas. Finally the phenomenon of multi-clonality, suggested by the presence of multiple (3+) microsatellite alleles at single loci, was also addressed in the context of TCI.

7.1   Overview of achievements

The field survey yielded >140 silvatic *T. cruzi* isolates that could be genotyped to a described lineage. These were combined with >1000 genotype records drawn from the literature, representing the largest *T. cruzi* genotype dataset ever assembled. The power of microsatellites to resolve the major *T. cruzi* lineages was tested across a representative panel of 48 strains assembled by MLEW (LSHTM) using ten microsatellite markers, and compared to sequence data. An expanded panel of ~50 microsatellites was then used to resolve the molecular epidemiology of TCI and TCIIc across a representative cohort of 135 and 53 strains respectively, again the largest such analysis ever attempted. *Gpi* sequence data from a subset, comprising 50 TCI and TCIIc strains in total, were obtained for comparison. To address the phenomenon of multi-clonality, 211 clones drawn from eight TCI isolates were analysed using nine microsatellite markers.

7.2   Thematic conclusions

7.2.1   Domestic disease transmission in Northern Bolivia: important or irrelevant?

The main objective of the field survey was to establish the prevalence and identity of circulating silvatic genotypes in those regions studied. Unlike at other fieldwork locations, however, the status of domestic transmission in Northern Bolivia was largely unknown, and

278

this was deemed sufficient justification to undertake a pilot study to identify possible local domestic and peri-domestic vector species (Section 1.5)(40). Two triatomine species, *T. sordida* and *R. robustus* were found infesting structures, and a further species, *R. pictipes*, was also found at peri-domestic foci (Table 25). Neither *Rhodnius* species has been found before at domestic or peridomestic foci in Bolivia, while the occurrence of *R. pictipes* at any site, silvatic or domestic, confirms the presence of this species in the country, whose local status had been the subject of some confusion in the literature (Section 4.4.1.2)

An immediate risk to the local population could be identified in some areas. *T. cruzi* was isolated from domestic and peridomestic vectors, including genotypes TCI and TCIIb, d, or e. (Section 4.2.1). Furthermore, the absence of TCIIb, d or e from the local silvatic cycle, at least that so far examined, and their common occurrence in domestic transmission cycles elsewhere in Bolivia, suggests that at least some of these triatomines had fed on humans (Section 4.4.1.2) (248, 400). Further implication of humans as the source of these lineages could be obtained by PCR-based blood-meal analysis of DNAzol-extracted intestinal homogenate, using a mitochondrial target to discriminate between different mammalian or avian hosts (e.g. (407)).

Domestic and peridomestic distribution and abundance was uneven across the study sites examined in Northern Bolivia (Section 4.1.1.1). Specifically, *T. sordida* was only identified in the southern portion of the study area, and all species of domestic and peridomestic triatomines were more abundant in this region. *T. sordida* was also the only triatomine species to behave highly anthrophilically i.e. colonising cracks in adobe walls inside dwellings in some communities. This, and the fact that no *T. sordida* silvatic populations were identified, lead to speculation that this species is being dispersed with human populations, possibly from known domestic foci further South (333).

Silvatic vector populations, principally *Rhodnius spp.*, throughout the region demonstrated high levels of *T. cruzi* infection, representing a significant risk of adventitious infection from the silvatic environment and consistent with other studies throughout the Amazon region (77). Corresponding TCI genotypes between silvatic and domestic transmission cycles suggested a link. However, as demonstrated by later analyses using microsatellite markers in the context of Venezuelan silvatic and domestic isolates, TCI populations from

domestic transmission cycles, especially from the human host, can be distinct from those in the local silvatic environment (Section 5.2.2.1). TCI is also known to occur in domestic transmission cycles throughout Bolivia (227, 248). Microsatellite analysis could now identify whether: 1) TCI genotypes found in domestic vectors in El Beni do originate silvatically, or 2) that they have been introduced by migrant human populations from other endemic areas.

In conclusion, rates of domestic and peri-domestic triatomine infestation in El Beni are low by comparison to areas of Bolivia traditionally considered to be hyperendemic for domestic disease transmission (e.g. (90)). However, a level of domestic transmission is present and represents an emergent threat to human health as peasant farmers migrate to this region from the resource-poor west of the country. There is currently no active scheme for disease surveillance in the region (Cortez, M *personal communication*). On the basis of this pilot study, further more comprehensive investigation is now urgently required, including detailed combined serological, parasitological and vector surveys. In tandem, community surveillance initiatives must also be encouraged. Few householders had any clear understanding of the nature of the disease and its transmission. Any further empirical studies would ideally include a complimentary survey of local knowledge, attitudes and practices with respect to Chagas disease, like others successfully employed elsewhere in South America, and with a view to gauging the level of community education required to promote sustained local surveillance (e.g. (408))

7.2.2 Niche-vector-host-genotype associations: fixed or fuzzy?

The debate regarding the ecological and evolutionary processes that underlie the current diversity of major *T. cruzi* genotype lineages has been ongoing since the first discovery of genetic variation in the parasite in the late 1970s (165). This study is the first to exhaustively survey all published *T. cruzi* records from the silvatic environment in context with original genotype data collected from the field (Chapter 4). The major contribution made by the field data was the observation that TCIIc, previously thought to occur only south of the Amazon, actually occurs as far north as Venezuela and, by implication, possibly throughout most of the silvatic distribution of *T. cruzi*. Other *T. cruzi* genotype

data obtained from the field fitted well with the historical data in terms of niche, host, vector and geographical distribution.

Historical sampling bias was found to have strongly skewed the dataset, such that some sources, especially didelphid marsupial hosts, were greatly over-represented. While, to an extent, this phenomenon was attributed to over-sampling of this certain species, *T. cruzi* prevalence rates gathered in this study, as well as previously, did demonstrate greater infection in some mammal species than in others. Circumstantially, it was suggested that opportunistic insectivory may predispose some mammals to infection, providing a more efficient means of transmission from vector to host than the faecal route (Section 4.4.1.1).

No single *T. cruzi* genotype was restricted to a specific host or vector species, ruling out tight co-speciation. Instead it was hypothesised that *T. cruzi* genotype diversification could be governed by niche. Pan-species trypanosome phylogenies point to vector and host cliques, largely defined by niche and termed as 'host fitting', as a strong correlate with diversification (221). Conceivably the same is true for genetic subdivisons within *T. cruzi*. TCIIc showed the clearest affinity with niche, and has been isolated exclusively from terrestrial transmission cycles, with *D. novemcinctus* and *Panstrongylus sp.* as major carriers, but other secondary terrestrial hosts were also implicated (Section 4.3). Sympatric TCI isolates from arboreal transmission cycles isolated during fieldwork underline the role of ecology in subdividing *T. cruzi* genotypes (Chapter 4). However, while the vast majority of TCI isolates were associated with arboreal didelphid marsupials, associated secondary hosts and vectors, a small but important subset were derived from arid, terrestrial ecotopes in association with rodents and possibly *Triatoma spp.* vectors. The significance of this observation in terms of parasite diversification was explored later in the context of microsatellite analysis (Section 5.2.2.1).

While TCI does not apparently occur sympatrically at terrestrial and arboreal sites, there is evidence to suggest that TCIIa does, especially in Brazil, and this lineage fits poorly with an ecological diversification paradigm (Section 4.4.2) (164). Lineages TCIIb, d and e, while of considerable importance in terms of domestic transmission, were virtually absent from the silvatic cycle, and where they do occur, they are occasionally linked to epizootic outbreaks (Section 4.4.2).

This study presents the most comprehensive survey of silvatic genotypes to date, yet the picture is still woefully incomplete. A number of avenues for further research are open. Targeted and extensive field surveys to establish the true vector, niche and host associations of TCIIb, TCIId, TCIIe and TCIIa are now urgently required. Genotype distribution maps collated in this study approximately identify where such lineages may be found (Figure 46 and Figure 48), and microsatellite analysis should have a role in distinguishing truly silvatic parasite populations from epizootic outbreaks by reference to local domestic strains. Empirical studies investigating infectivity of different *T. cruzi* strains in different silvatic hosts should also be envisaged where ecological associations are well documented in the field i.e. particularly involving TCI and TCIIc. Critically such studies should also involve parasite passage through the appropriate vector species as well as host, with a view to establishing which, if any, parasite lineages demonstrate specific adaptation. Similar studies are currently underway in the USA, but are limited in the number of strains under comparison (Yabsley, M *et al.*, *unpublished*). Any such study should ideally involve the comparison of a representative panel of strains from each lineage if meaningful conclusions are to be drawn.

7.2.3 *T. cruzi* diversity at the inter-lineage level: do microsatellites have the power?

Using a limited panel of ten microsatellite markers, the six major *T. cruzi* groups could be readily distinguished with reasonable bootstrap support under an IAM, with the possible exception of TCIIa (Figure 50). Deep branching patterns, specifically among TCI, TCIIa and TCIIc, were incongruent between microsatellite and *gpi* sequence analysis (Section 5.3.1). It was suggested that microsatellite loci were probably too sensitive to homoplasy to resolve ancient evolutionary relationships. Nonetheless, the technique could be a relatively rapid and robust future means of genotyping new field strains, especially due to the ease with which it differentiated between hybrid lineages TCIIe and TCIId (Section 5.3.1).

Expense is a major consideration when transferring genotyping technologies for use in the field. While microsatellites clearly have to the power to distinguish all major *T. cruzi* lineages, reagent and equipment costs are liable to prohibit their adoption as a standard genotyping protocol. RFLP typing strategies are more likely to fulfil this role in the future,

and a variety of new RFLP targets are now available for development, including one used for genotyping in this study (e.g.(191, 192) Section 3.5.3).

7.2.4   TCI and TCIIc: biogeography and eco-epidemiology

The microsatellite dataset drawn from the panel of TCI and TCIIc isolates assembled has facilitated one of the most detailed population genetic analyses of a zoonotic parasite from its wild transmission cycle ever attempted (Chapter 5). Features among 'core' silvatic populations in TCI: weak isolation by distance, excess homozygosity and high genetic diversity were common to TCIIc (Table 37 and Table 44). At a continental scale, *D. marsupialis* TCI isolates showed less spatial structuring than that observed in TCIIc from *D. novemcinctus* (Figure 56 and). It is not clear whether this can be exclusively linked to different primary host population dynamics. Mammal tissue samples taken during the study could provide enough genetic material to facilitate a comparative study examining gene flow between different host populations. A rapidly evolving mitochondrial marker seems an excellent candidate to achieve this aim, and variable targets have been identified for both species (344, 348).

Although spatial structuring was relatively weak between 'core' populations, levels were higher than those observed for predominantly human parasites. Sub-division- between Lowland Bolivia and North Eastern Brazil TCI isolates (c. 3000 km apart, $F_{ST} = 0.1446$, Table 38), for example, was equivalent to that observed recently between *P. falciparum* populations from Dakar and the Horn of Africa (c. 6000km apart, $F_{ST} = 0.109 - 0.194$)(409). It was suggested that parasite dispersal between sites occurs at a lower rate when humans are not involved.

Also shared between the TCI 'core' and TCIIc populations, was an apparent lack of genetic structure by host and possibly vector (where samples were available i.e. TCI), ruling out co-speciation, and providing evidence that both genotypes behave eclectically within their respective arboreal and terrestrial niches (Figure 55 and Figure 65). Where diversification did occur, at least in silvatic TCI rodent samples, niche again may have been the governing factor. The genetic subdivision between TCI rodent isolates from Cochabamba and nearby 'core' TCI isolates from Lowland Bolivia was far greater than that one might expect

through geographical vicariance alone (Table 38). Perhaps the pattern of divergence between arboreal and terrestrial TCI isolates mirrors that which could have occurred earlier in the evolutionary history of *T. cruzi* in South America to give rise to TCI and TCIIc. Further examples of terrestrial rodent-associated TCI populations like those identified from Piaui and the Pantanal in Brazil would be required to support this assertion(223, 354).

A competing explanation exists to explain the diversification of rodent TCI isolates from Bolivia - human influence. Domestic and silvatic isolates are closely genetically linked locally. Furthermore, the high level of identity between Chilean and Cochabamban isolates is practically inexplicable in the absence of human involvement, and the reduction in diversity observed across a wide geographic area is consistent with some kind of population event like a founder effect or population bottleneck (Section 5.3.3). The switch to a human host is also associated with diversification and a possible bottleneck in TCI isolates in Venezuela. In this case, a high level of subdivision was maintained between most human TCI isolates and local silvatic strains, despite evidence that domestic and silvatic vector populations, at least in some endemic areas in Venezuela, are genetically indistinguishable (392).

Both TCI 'bottleneck' populations showing low levels of genetic diversity and suspected of a recent switch of host and ecotope, i.e. Venezuelan domestic strains (VD, Population 2, Table 34) as well as Cochabamban and Chilean strains (COT+CH, Population 2, Table 34) also demonstrated excess heterozygosity with respect to 'core' populations. This observation was tentatively linked to the high levels of heterozygosity and reduced diversity observed in TCIId and TCIIe as possible evidence of a recent fusion event (Section 5.3.3)(375). It has long been postulated, largely based on circumstantial evidence, that the hybrid nature of TCIId and TCIIe may confer some advantage to survival in the human host (250). Clearly sample size is an issue in both TCI 'bottleneck' populations, and future work must focus on collecting more isolates from both regions before any satisfactory conclusions can be drawn. Additionally, without an empirical basis for argument, it is impossible to determine whether the reduced diversity and raised heterozygosity in VD and COT+CH is attributable to selective or neutral population processes. As discussed in Section 3.7.6.1.9, a population bottleneck generated by a selective sweep and one caused by a random sampling event are theoretically

indistinguishable at neutral loci. In Venezuela at least, where sampling from domestic and silvatic transmission cycles is ongoing in conjunction with detailed clinical data collection, there is excellent potential to expand the current study (Carrasco, H. *personal communication*). A number of biological characteristics of these strains could be compared with those from the silvatic environment, including virulence and infectivity *in vitro* and *in vivo*, as well as clinical outcome of infection in humans. Finally, adequate consideration must be made regarding the implications this study may have in terms of disease control in Venezuela, especially if poor transmission from silvatic vectors, rather than selection within the human host, is to blame for the lack of silvatic-type *T. cruzi* strains in the human population. In areas of established endemicity, limited national funds may be better spent on maintaining domestic vector eradication programmes rather than protective measures (bed nets, window and door screens etc) aimed at sheltering house holders from infection through opportunistic feeding by silvatic vectors (64).

7.2.5 Multiclonality: fascinating but frustrating.

The presence of multiple, genetically distinct, clones within *T. cruzi* strains analysed in this study was postulated from multi-allelic loci in Chapter 5, and confirmed, in TCI at least, through biological cloning and analysis of infrapopulation diversity in Chapter 6. As a general measure, the number of multi-allelic loci in the uncloned profile correlated well with the resulting number of MLGs in the associated infrapopulation ($R_{XY}=0.838$, p<<0.05, Section 6.1.1). However, a degree of hidden diversity was present, such that one strain demonstrating no multi-allelic loci actually comprised a considerable number of distinct MLGs (XE5167). This observation is of critical importance to previous population studies, especially in *T. cruzi* (microsatellites) and *T. brucei* (minisatellites), where exclusively diploid profiles from uncloned isolates were universally assumed to be clonal(215, 217, 236, 380). In the context of this study, it is now important to extend infrapopulation analysis to uncloned strains from non-diverse TCI populations like VD and COT+CH. The expectation of a correlation between low multi-allelic locus frequency with low multi-clonality should remain, but further empirical support is necessary to fully corroborate the observed reduction in diversity. A similar rationale should now also be applied to TCIIc.

In Chapter 5, it was proposed that mutliclonality might take the form of 'hidden homozygosity' within heterozygous uncloned profiles. However, it was demonstrated in Chapter 6 that the clonal composition in those infrapopulations studied was rarely balanced. Instead, a majority genotype was present in almost all infrapopulations, to which the uncloned profile was most closely related, with minor contributions from other, rarer MLGs. Thus the level of homozygosity observed across the uncloned dataset is not artefatual and is likely to represent closely that of the dominant genotype within each infrapopulation (Section 6.1.3).

The level of diversity and multiclonality identified from those infrapopulations studied was astounding. Moreover, given that the strains had already been serially repassaged and cryopreserved, both of which will act to reduce diversity, the level identified is likely to represent a fraction of that originally present, and, by implication, an even smaller fraction of the total diversity of parasites within the mammal. This observation, concurrent with high overall infection prevalence among those mammals sampled (e.g. 75% in *D. marsupialis* in southern El Beni, Table 6), is consistent with incredibly intense silvatic disease transmission. It is hard to reconcile this level of parasite abundance and diversity with an inefficient faecal transmission route. Perhaps oral transmission is allowing parasite diversity to be efficiently transmitted *en masse* from vector to host. This, ecological, mechanism for the maintenance of extraordinary levels of neutral parasite diversity, is perhaps more parsimonious than one invoking some mechanistic phenomenon. Balancing selection, through linkage between microsatellite loci and antigenic genes for example, could explain some variation, but in the absence of clear mapping of the chromosomal location of either antigenic genes or microsatellite loci, or prior knowledge of which genes may be under such selection, such theories are best left until the genomes of the relevant lineages are properly mapped and annotated. The shotgun sequencing of the entire genome for the TCI reference strain Silvio X10, for example, is currently underway (Andersson B, *personal communication*).

There is some empirical evidence to suggest the importance of the oral route in *T. cruzi* transmission. Recent studies in mouse models demonstrate that *T. cruzi* may possess a group of mucin-like a surface molecules highly resistant to the proteolytic conditions of the stomach, and acidic conditions may even potentiate invasion of intestinal mucosae by

metacyclic parasites (410). Targeted study examining opportunistic insectivory in key reservoir species, possibly through analysis of gut contents, could be relevant. Exoskeletal remains derived from the intestines of didelphid marsupials have allowed the classification of insects to Family level (411). Perhaps, with the aid of molecular tools, the level of taxonomic resolution could be improved.

The presence of multiclonality in uncloned samples in this study has frustrated population genetic analysis of the dataset, as well as prohibiting meaningful interpretation of linkage disequilibrium statistics (Section 5.3). However, the software approach developed to address the challenge posed by multi-allelic loci allowed the extent to which they may have disrupted the analysis to be assessed, and overall they represented a tiny portion of the total dataset. These problems were far outweighed by the important insight mutliclonality provided with regards to the amount of diversity potentially present in *T. cruzi* populations at the host level.

7.2.6   Recombination: do we have the right tools?

The frustrating role that multiclonality played in interpreting linkage statistics in Chapter 5 is not a fault specific to microsatellites. In future, biological cloning should be considered standard in any such study, even though the process is laborious and numerous clones must be taken to avoid sampling minority genotypes (Section 6.2). However, as discussed in Chapter 6, microsatellite markers may be intrinsically unsuited to detecting recombination events, especially it they are rare and putative parents are unknown. Mutation rates may be too high, either falsely generating recombinant allelic combinations or rapidly erasing the signal of past recombination events.

If recombination was to occur in *T. cruzi* populations, it must do so between strains occupying the same geographic distribution, niche, vector, and/or host. Thus, searching for recombinant clones among parasite strains known to be in close contact, i.e. in the same vector or host organism, remains a viable strategy. Known hybrids TCIId and TCIIe both fall within geographical distribution of parents TCIIb and TCIIc, although the precise silvatic affinities of TCIIb are not yet known. Improvements must be made with respect to the use of the correct molecular tools, and MLST seems to be a vital parallel approach in

the light of high microsatellite mutability. Lewis & Yeo (*personal communication*) recently demonstrated that some North American TCI lineages, including a subset of those analysed in this study, actually possess maxicircle DNA more closely related to local TCIIa lineages than to other TCI isolates. In this case, however, both nuclear halpotypes, based on *gpi* sequence, appear to be TCI in origin. Inter-lineage recombination - between TCI and TCIIa - is suspected, which fits with a shared ecological niche. A similar approach, involving both nuclear and mitochondrial markers, could be envisaged to study the extent of intralineage recombination in TCI and TCIIc, *again possibly within parasite populations inhabiting individual hosts or vectors*. In this case, an expanded set of molecular markers should be considered to provide sufficient informative variation. These should include several single-copy house-keeping nuclear genes such as superoxide dismutase, glutathione peroxidase, trypanothione reductase and dihydrofolate reductase–thymidylate synthase, among others (412, 413). Ideally, a number of such genes should be contiguous to facilitate the identification of intergenic breakpoints and linkage-distance effects. Chromosome three is among the most accurately mapped regions of the *T. cruzi* genome, has a number of genes that could be suitable candidates, and also contains a number of di- and tri-nucleotide microsatellite repeat sequences (414). To allow for fine-scale differentiation, a number of more variable nuclear genes would also have to be sequenced, possibly including the mini-exon gene spacer region, as well as antigenic targets (384). Finally, to provide information regarding uniparental inheritance of the kinetoplast during hybridisation events, at least one mitochondrial target should be incorporated in the panel (e.g. cytochrome oxidase subunit II (COII) or NADH dehydrogenase subunit 1)(199).

A number of putative fusion recombinants were identified in Chapter 6. Observed aneuploidy in the microsatellite profiles of these clones should correlate with raised genome content if these represent true and recent hybrids of same type generated experimetally by Gaunt *et al.*, (2003) (i.e. non-meiotic) (183). Lewis (2008) has successfully identified raised ploidy levels in experimentally-derived TCI hybrids using fluorescence activated cell sorting, and perhaps a similar approach is applicable here (183, 375).

7.2.7 Microsatellite analysis as a tool for examining *T. cruzi* diversity: advantages and limitations

If microsatellite analysis is to have a secure future as an epidemiological tool for examining *T. cruzi* diversity, a clear assessment of the relative merits of this technique with respect to this particular parasite must be made. Limitations common to all population genetic analyses involving microsatellites, i.e. saturation and a poorly defined mutational model, are obviously important, but not of specific interest here. Reproducibility, frequently a concern with microsatellites by comparison to other genotyping methods, was apparently not a major issue in this study, as indicated by the ease with which identical MLGs were recovered in Chapter 6. Nonetheless, some margin of error is inherent in all such analysis.

T. cruzi is primarily problematic because it deviates so dramatically from the assumptions that underlie so much population genetic theory. As such, *T. cruzi* genetic diversity is not amenable to most analytical tools developed to study organisms conforming to HW expectations. In this study, a number of these tools were employed, but care was taken to limit inference made from any one technique. Similar caution has been lacking from recent studies of related organisms (e.g. (266, 267). Theoretical advances have been made with respect to population genetic models in predominantly clonal organisms (299). Unfortunately *T. cruzi* is far from a model clonal organism, to which the lack of excess heterozygosity in most populations studied here is testament (Chapter 5). Nonetheless, in the absence of a sound population genetic model, this study demonstrates that microsatellites can still provide an incredible amount of information with respect to the molecular epidemiology of *T. cruzi*. Defining at what level *T. cruzi* populations are fundamentally structured does, however, still represent a problem with respect to this study. In the absence of a robust means of assigning individuals to populations, as well as sample size issues, this analysis currently only really represents a preliminary study. Only with a far greater number of isolates and a parametric, though non HW, means of defining populations, will the true population genetic structure of this parasite be revealed and the confident interpretation of population parameters allowed.

Using the panel of microsatellite markers developed here, investigation can now be theoretically extended across a variety of different geographical sites and epidemiological

scenarios, and incoprating many more samples. Before doing so, the panel of existing markers must be streamlined, reduced, and tailored towards the specific research question.

In terms of defining population structure and subdivision, the lack of identical MLGs observed in Chapter 5 from both TCI and TCIIc suggests that a reduction in locus number could be achieved without excessive loss of resolutive power. Doing so would also dramatically lower the cost of analysis. The primary criterion for marker selection should be how informative they are in terms of defining discrete populations. Informative loci can be identified from the current panel through trial and error by selective deletion, but also *in silico* using software like Banks *et al*'s (2003) WHICHLOCI (415). Secondary criteria involve robust amplification of loci across strains, minimal stutter patterns, and a lack of slippage (Section 3.7.5).

To provide insights into other mechanisms, recombination for example, a number of new microsatellite targets would have to be developed and, as is Section 7.2.6, combined with sequence analysis. Analysis of multiple microsatellite loci of known linkage, again possibly derived from chromosome three, should allow multilocus microsatellite haplotypes to be derived through simulation using PHASE (205). Any putative recombination events or signal derived could be confirmed in parallel with linked gene sequences from the same region.

Overall, this study demonstrates that microsatellite markers, while possessing limitations, are an invaluable tool for studying the molecular epidemiology of *T. cruzi*. Also, in parallel with sequence data, multiple linked microsatellite markers could begin to unravel some of the genetic mechanisms involved with *T. cruzi* diversification. Critically, using this work as a basis, future analysis should be able to provide information that is of more than merely academic interest. Microsatellites could not only be key to finally unravelling fundamental mechanics of *T. cruzi* transmission, but also a vital asset in planning effective strategies to control Chagas disease.

7.2.8   Future outlooks

The genotyping techniques developed in this study now provide an excellent platform for future work. A number of possible studies and strategies were alluded to in the preceding sections. Only the major lines of future research are summarised here.

Primarily, research effort will be focused on gathering more TCI isolates from humans in Venezuela, as well as isolates from rodents from rocky ecotopes in Bolivia and Brazil. These strains will be subjected to microsatellite analysis and data generated included in the dataset presented here, which will provide greater insight in the population processes involved in their diversification. Crucially, live strains, either gathered from the field or from cryopreserved stocks held in Bolivia (Dr. Mirko Rojas Cortez), Brazil (Dr. Anna Jansen), and Venezuela (Dr. Hernan Carrasco) will be obtained for comparisons of infectivity and virulence *in vitro* and *in vivo*.

Secondarily, pilot microsatellite studies investigating silvatic and domestic TCI transmission dynamics will be rolled out to other countries where this lineage co-occurs in both cycles, specifically in Bolivia (Dr. Mirko Rojas Cortez) and Colombia (Professor Felipe Guhl), among others, as a means of epidemiological tracking using a specially tailored panel of informative markers.

Thirdly, fieldwork will be undertaken to examine the silvatic ecological affinities of ambiguous strains TCIIa (in Venezuela, Colombia and Northern Brazil), and TCIIb, TCIId, and TCIIe (in South-Eastern Brazil, Paraguay and Bolivia). Again, after basic genotyping, microsatellites will have a role in distinguishing truly silvatic strains from epizootic outbreaks by reference to those occurring in local domestic cycles.

Finally, once basic research questions regarding parasite transmission dynamics and major lineage ecological affinities have been addressed, targeted parallel MLMT and MLST analysis can be employed to investigate the genetic mechanisms of diversification among the panels of strains assembled, doing justice to the wealth of genetic tools becoming available as *T. cruzi* enters the genomic era, and addressing fundamental aspects of *T. cruzi* biology.

8 Appendices

8.1 References for silvatic genotype meta-analysis.

Apt, W. Aguilera, X. Arribada, A. Gomez, L. Miles, M. A. Widmer, G. (1987) Epidemiology of Chagas' disease in northern Chile: isozyme profiles of *Trypanosoma cruzi* from domestic and sylvatic transmission cycles and their association with cardiopathy *Am J Trop Med Hyg* 37, 302-7.

Barnabe, C., 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, 73-9.

Brisse, S., Barnabe, C. Tibayrenc, M. (2000) Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis *Int J Parasitol* 30, 35-44.

Carrasco, H. J. Frame, I. A. Valente, S. A. Miles, M. A., (1996) Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi Am J Trop Med Hyg* 54, 418-24.

Ceballos, L. A. Cardinal, M. V. Vazquez-Prokopec, G. M. Lauricella, M. A. Orozco, M. M. Cortinas, R. Schijman, A. G. Levin, M. J. Kitron, U. Gurtler, R. E. (2006) Long-term reduction of *Trypanosoma cruzi* infection in sylvatic mammals following deforestation and sustained vector surveillance in northwestern Argentina *Acta Trop* 98, 286-96.

Clark, C.G. & Pung, O.J. (1994) Host specificity of ribosomal DNA variation in sylvatic *Trypanosoma cruzi* from North America *Mol Biochem Parasitol* 66, 175-9.

Cortez, M. R. Pinho, A. P. Cuervo, P. Alfaro, F. Solano, M. Xavier, S. C. D'Andrea, P. S. Fernandes, O. Torrico, F. Noireau, F. Jansen, A. M. (2006) *Trypanosoma cruzi*

(Kinetoplastida: Trypanosomatidae) ecology of the transmission cycle in the wild environment of the Andean valley of Cochabamba, Bolivia *Exp Parasitol* 114, 305-13.

Dedet, J. P. Chippaux, J. P. Goyot, P. Pajot, F. X. Tibayrenc, M. Geoffroy, B. Gosselin, H. Jacquet-Vialet, P. (1985) Natural hosts of *Trypanosoma* cruzi in French Guiana. High endemicity of zymodeme 1 in wild marsupials *Ann Parasitol Hum Comp* 60, 111-7.

Dereure, J. Barnabe, C. Vie, J. C. Madelenat, F. Raccurt, C. (2001) *Trypanosomatidae* from wild mammals in the neotropical rainforest of French Guiana *Ann Trop Med Parasitol* 95, 157-66.

Diosque, P. Barnabe, C. Padilla, A. M. Marco, J. D. Cardozo, R. M. Cimino, R. O. Nasser, J. R. Tibayrenc, M. Basombrio, M. A. (2003) Multilocus enzyme electrophoresis analysis of *Trypanosoma cruzi* isolates from a geographically restricted endemic area for Chagas' disease in Argentina *Int J Parasitol* 33, 997-1003.

Fernandes, A. J. Chiari, E. Rodrigues, R. R. Dias, J. C. Romanha, A. J. (1991) The importance of the opossum (*Didelphis albiventris*) as a reservoir for *Trypanosoma cruzi* in Bambui, Minas Gerais State *Mem Inst Oswaldo Cruz* 86, 81-5.

Fernandes, A. J. Diotaiuti, L. Dias, J. C. Romanha, A. J. Chiari, E. (1994) Relationships between *Trypanosoma cruzi* transmission cycles in the county of Bambui, Minas Gerais, Brazil. *Cad Saude Publica* 10, 473-80.

Fernandes, C. D. Murta, S. M. Ceravolo, I. P. Krug, L. P. Vidigal, P. G. Steindel, M. Nardi, N. Romanha, A. J. (1997) Characterization of *Trypanosoma* cruzi strains isolated from chronic chagasic patients, triatomines and opossums naturally infected from the State of Rio Grande do Sul, Brazil *Mem Inst Oswaldo Cruz* 92, 343-51.

Gurgel-Goncalves, R. Ramalho, E. D. Duarte, M. A. Palma, A. R. Abad-Franch, F. Carranza, J. C. Cuba Cuba, C. A. (2004) Enzootic transmission of *Trypanosoma cruzi* and T. rangeli in the Federal District of Brazil *Rev Inst Med Trop Sao Paulo* 46, 323-30.

293

Hernandez, R. Herrera, J. Bosseno, M. F. Breniere, S. F. Espinoza, B. (2001) *Trypanosoma cruzi:* data supporting clonality in Mexican stocks *J Parasitol* 87, 1178-81.

Herrera, L. D'Andrea, P. S. Xavier, S. C. Mangia, R. H. Fernandes, O. Jansen, A. M. (2005) *Trypanosoma cruzi* infection in wild mammals of the National Park 'Serra da Capivara' and its surroundings (Piaui, Brazil), an area endemic for Chagas disease *Trans R Soc Trop Med Hyg* 99, 379-88.

Lewicka, K. Breniere-Campana, S. F. Barnabe, C. Dedet, J. P. Tibayrenc, M. (1995) An isoenzyme survey of *Trypanosoma cruzi* genetic variability in sylvatic cycles from French Guiana *Exp Parasitol* 81, 20-8

Lisboa, C. V. Dietz, J. Baker, A. J. Russel, N. N. Jansen, A. M. (2000) *Trypanosoma cruzi* infection in *Leontopithecus rosalia* at the Reserva Biologica de Poco das Antas, Rio de Janeiro, Brazil *Mem Inst Oswaldo Cruz* 95, 445-52

Lisboa, C. V. Mangia, R. H. Luz, S. L. Kluczkovski, A., Jr. Ferreira, L. F. Ribeiro, C. T. Fernandes, O. Jansen, A. M. (2006) Stable infection of primates with *Trypanosoma cruzi* I and II *Parasitology* 133, 603-11.

Marquez, E. Arcos-Burgos, M. Triana, O. Moreno, J. Jaramillo, N. (1998) Clonal population structure of Colombian sylvatic *Trypanosoma cruzi J Parasitol* 84, 1143-9.

Miles, M. A. Povoa, M. M. de Souza, A. A. Lainson, R. Shaw, J. J. Ketteridge, D. S. (1981) Chagas's disease in the Amazon Basin: Ii. The distribution of *Trypanosoma cruzi* zymodemes 1 and 3 in Para State, north Brazil *Trans R Soc Trop Med Hyg* 75, 667-74.

Miles, MA Toye, PJ Oswald, SC Godfrey, DG (1977) The identification by isoenzyme patterns of two distinct strain-groups of *Trypanosoma cruzi*, circulating independently in a rural area of Brazil *Trans Roy Soc Trop Med Hyg* 71, 217-225

Miles, M. A. Souza, A. Povoa, M. Shaw, J. J. Lainson, R. Toye, P. J. (1978) Isozymic heterogeneity of *Trypanosoma* cruzi in the first autochthonous patients with Chagas' disease in Amazonian Brazil. *Nature* 272, 819-21.

Montamat, E. E. De Luca d'Oro, G. Perret, B. Rivas, C. (1991) Characterization of *Trypanosoma cruzi* from Argentina by electrophoretic zymograms *Acta Trop* 50, 125-33.

Montilla, M. M. Guhl, F. Jaramillo, C. Nicholls, S. Barnabe, C. Bosseno, M. F. Breniere, S. F. (2002) Isoenzyme clustering of *Trypanosomatidae* Colombian populations *Am J Trop Med Hyg* 66, 394-400.

Murta, S. M. Gazzinelli, R. T. Brener, Z. Romanha, A. J. (1998) Molecular characterization of susceptible and naturally resistant strains of *Trypanosoma cruzi* to benznidazole and nifurtimox *Mol Biochem Parasitol* 93, 203-14.

Pinho, A. P. Cupolillo, E. Mangia, R. H. Fernandes, O. Jansen, A. M. (2000) *Trypanosoma cruzi* in the sylvatic environment: distinct transmission cycles involving two sympatric marsupials *Trans R Soc Trop Med Hyg* 94, 509-14.

Pinho, A.P. Cuervo, P. Gripp, E. Costa, B. Herrera, H. D'Andrea, P. Pa'dua, T. Rademaker, V. Fernandes, O. Cupolillo, E. Jansen, A.M. (2003) Isoenzymatic characterization and genetic diversity of *Trypanosoma cruzi* isolates of wild animals from Pantanal, Brazil *Rev Inst Med Trop S Paulo* 45, 12

Povoa, M. M. de Souza, A. A. Naiff, R. D. Arias, J. R. Naiff, M. F. Biancardi, C. B. Miles, M. A. (1984) Chagas' disease in the Amazon basin IV. Host records of *Trypanosoma cruzi* zymodemes in the states of Amazonas and Rondonia, Brazil *Ann Trop Med Parasitol* 78, 479-87.

Samudio, F. Ortega-Barria, E. Saldana, A. Calzada, J. (2007) Predominance of *Trypanosoma cruzi I* among Panamanian sylvatic isolates *Acta Trop* 101, 178-81.

Sanchez-Guillen Mdel, C. Barnabe, C. Tibayrenc, M. Zavala-Castro, J. Totolhua, J. L. Mendez-Lopez, J. Gonzalez-Mejia, M. E. Torres-Rasgado, E. Lopez-Colombo, A. Perez-Fuentes, R. (2006) *Trypanosoma cruzi* strains isolated from human, vector, and animal reservoir in the same endemic region in Mexico and typed as *T. cruzi I,* discrete typing unit 1 exhibit considerable biological diversity. *Mem Inst Oswaldo Cruz.* 101, 585-90.

Saravia, N. G. Holguin, A. F. Cibulskis, R. E. D'Alessandro, A. (1987) Divergent isoenzyme profiles of sylvatic and domiciliary *Trypanosoma cruzi* in the eastern plains, piedmont, and highlands of Colombia *Am J Trop Med Hyg* 36, 59-69.

Solari, A. Munoz, S. Venegas, J. Wallace, A. Aguilera, X. Apt, W. Breniere, S. F. Tibayrenc, M. (1992) Characterization of Chilean, Bolivian, and Argentinian *Trypanosoma cruzi* populations by restriction endonuclease and isoenzyme analysis *Exp Parasitol* 75, 187-95.

Steindel, M. Toma, H. K. Ishida, M. M. Murta, S. M. de Carvalho Pinto, C. J. Grisard, E. C. Schlemper, B. R., Jr. Ribeiro-Rodrigues, R. Romanha, A. J. (1995) Biological and isoenzymatic characterization of *Trypanosoma cruzi* strains isolated from sylvatic reservoirs and vectors from the state of Santa Catarina, Southern Brazil *Acta Trop* 60, 167-77.

Steindel, M. Kramer Pacheco, L. Scholl, D. Soares, M. de Moraes, M. H. Eger, I. Kosmann, C. Sincero, T. C. Stoco, P. H. Murta, S. M. de Carvalho-Pinto, C. J. Grisard, E. C. (2008) Characterization of *Trypanosoma cruzi* isolated from humans, vectors, and animal reservoirs following an outbreak of acute human Chagas disease in Santa Catarina State, Brazil *Diagn Microbiol Infect Dis* 60, 5-32.

Thomaz Soccol, V. Barnabe, C. Castro, E. Luz, E. Tibayrenc, M. (2002) *Trypanosoma cruzi:* isoenzyme analysis suggests the presence of an active Chagas sylvatic cycle of recent origin in Parana State, Brazil *Exp Parasitol* 100, 81-6.

Travi, B. L. Jaramillo, C. Montoya, J. Segura, I. Zea, A. Goncalves, A. Velez, I. D. (1994) *Didelphis marsupialis*, an important reservoir of *Trypanosoma (Schizotrypanum) cruzi* and *Leishmania (Leishmania) chagasi* in Colombia *Am J Trop Med Hyg* 50, 557-65.

Yeo, M. Acosta, N. Llewellyn, M. Sanchez, H. Adamson, S. Miles, G. A. Lopez, E. Gonzalez, N. Patterson, J. S. Gaunt, M. W. de Arias, A. R. Miles, M. A. (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, 225-33.

Zalloum, L. Gomes, M. L. Kinoshita, A. T. Toledo, M. J. Prioli, A. J. de Araujo, S. M. (2005) *Trypanosoma cruzi*: two genetic groups in Parana state, Southern Brazil *Exp Parasitol* 111, 55-8.

8.2   Original references for samples used in this thesis

Apt, W. *et al.*, (1987) Am J Trop Med Hyg 37, 302-7.	CHILEC22, CHILEWALL
Barnabe, C. *et al.*, (2001) Exp Parasitol 99, 73-9.	VFRA1 CL1
Barnabe, C. *et al.*, (2001) Acta Trop 78, 127-37.	92101601P CL1, SAIMIRI 3 CL1, STC10RCL1, 92090802P, 93070103P, 92101601P, USAOPOSSUM, USAARMA, 92122102R
C. Barnabe *unpublished*	P251CL7, 361TA, 458, DAVIS 9.90, P234, P238, P234, PII(BOL)
Breniere, S. F. *et al.*, (1991) Trans R Soc Trop Med Hyg 85, 62-6.	CBB CL2,
Breniere, S. F. *et al.*, (1998) Acta Trop 71, 269-83.	TU18 CL2,IVV CL4
Brisse, S. *et al.*, (1998) Mol Biochem Para 92, 253-263.	CL BRENER,
Carrasco, H. J. *et al.*, (1996) Am J Trop Med Hyg 54, 418-24	PII(CJ007), PII(CJ005), CJO20, CJO33
H. Carrasco *unpublished*	JR CL4, 7570, 8104, 9010, 9354, 10775, 10801, 11006, 11042, 11124, 11398, 11541, 11713, 11804, 11838, 11881, AM3, AM6, AMRPA34, BARM, CALC104, CASABONIF, CD45, C022, CO57, CO75, CO84, DM1, DM4, DMSU8, DMSUC, JR, PGN900, RR5, TCSCII, TERF-1, TEV55, V1, V2, V4
Diosque, P. *et al.*, (2003) Int J Parasitol 33, 997-1003.	PALDA1, PALDA20, PALDA21, PALDA22, PALDA3, PALDA4, PALDA5, PALDAV2^3, TEDA,
P. Diosque *unpublished*	PAH179 CL5, EPV20-1CL1, LHVACL4
Miles, M., *et al.*, (1977) Trans R Soc Trop Med Hyg 71, 217-225.	ESM CL3, RITACL5
Miles, M. A. *et al.*, (1978) Nature 272, 819-21.	CAN III CL1
Miles, M. A. *et al.*, (1981) Trans R Soc Trop Med Hyg 75, 667-74.	M5631CL5
Miles, M. A. *et al.*, (1981) Lancet 1, 1338-40.	X10/1
M. Miles *unpublished*	XE1313, XE1342, XE1381, XE1383, B2084, B2085, XE2913, XE2929, B3159, XE3309, B3726, XE3776, XE3981, XE4389, XE4682, XE4993, XE5011, XE5012, XE5017, XE5164, XE5165, XE5167, B5770, B5781, B5787, XE5809, XE5847, XE5495, XE6004, SABP19CL1, B1947, B2026, B2077, B2310, B2311, XE5740, IM4810
N. Acosta *unpublished*	MA194, MA202, MA204, MA215, MA219, MA220, MA222
Povoa, M. M. *et al.*, (1984) Ann Trop Med Parasitol 78, 479-87.	JA2 CL2.2
Saravia, N. G. *et al.*, (1987) Am J Trop Med Hyg 36, 59-69.	CM17, CM 25
Solari, A. *et al.*, (1992) Exp Parasitol 75, 187-95.	85/847
Souto, R. P. *et al.*, (1996) Mol Biochem Parasitol 83, 141-52.	BUG 2148 CL1,
M. Llewellyn *unpublished*	COTMA22, COTMA38, COTMA47, COTMA55, COTMA9, LL2LA, M12, M13, M15, M16, M18, M7, PARAMA39, PARAMA40, PARAMA41, SJM18, SJM22, SJM23, SJM26, SJM3, SJM32, SJM33, SJM34, SJM35, SJM37, SJM39, SJM40, SJM41, SJMC12, SJMC3, SJMC7, SANRAFLB, SANRAFP2A, SANRAFP2B, ARMA9, CAYMA13, CAYMA14, CAYMA17, CAYMA18, CAYMA19, CAYMA21, CAYMA25, CAYMA3, M3-CU, M10, M5, M6, M8, PARAMA26, PARAMA25, PARAMA34,PARAMA6,SAM6,SJMC10,SJMC4, SJMC19,SJMO18,SMA18,SMA19,SMA8XE,SMA9
Tibayrenc, M. & Ayala, F. J. (1987) Parasitol Today 3, 189-90.	10 R26,M6421 CL6,TULA CL2,
Tibayrenc, M. & Miles, M. A. (1983) Trans R Soc Trop Med Hyg 77, 76-83.	SC43CL1,92.80 CL2,C8 CL1,
Widmer, G. *et al.*, (1987) Mol Biochem Parasitol 23, 55-62.	SAXP18 CL5,VINCH101 CL1
Yeo, M. *et al.*, (2005) Int J Parasitol 35, 225-33.	ARMA12,MA25*,MA25X*,MA87,SP13,SP14,SP15,SP16,SP4, ARMA13,ARMA18,ARMA24,ARMA25,ARMA26,ARMA27,
Yeo, M. *et al.*, (2007) Int J Parasitol 37, 111-20.	B187 CL10,ERA CL2,X10610 CL5,POT7A CL1,POT7B CL5,PARA4CL3, PARA6 CL3,CHACO2 CL3
M. Yeo *unpublished*	CHACO 23 COL4,CHACO9 COL15,CHACO17 COL1,

8.3   Infrapopulation analysis appendices (overleaf)

298

Table 55 Allelic profiles across forty-one diploid MLGs on the basis of nine microsatellite loci

Population	Clone	Linkage 7118 / 1TETP12a[a]	7143 / famp6a[a]	7206 / 1FAMP7c[a]	7853 / vicp5a[a]	8328 / 1TETP4a[a]	8328 / 1TETP4b[a]	8370 / 1FAMP9c[a]	8646 / famp8a[a]	8646 / famp8b[a]	FREQ[b]
LB	sjm34d	136/136	140/142	113/115	150/154	109/109	167/170	106/110	129/129	171/171	1
LB	sjm34a	136/136	140/142	113/115	150/154	109/109	167/170	106/110	129/129	171/177	14
LB	sjmc3	136/136	142/142	111/115	147/150	107/109	164/170	106/110	126/129	171/179	n/a
LB	sjm41b	153/155	131/131	107/107	145/152	107/107	173/173	106/106	126/126	175/183	2
LB	sjm41a	153/155	131/131	107/111	145/152	107/107	173/173	106/106	126/126	175/183	30
LB	sjm37	136/149	133/133	113/113	143/143	107/109	176/176	94/106	126/126	175/187	n/a
LB	sjm22	136/153	131/142	115/119	152/152	107/109	167/170	110/110	129/129	177/179	n/a
LB	sjm39	136/136	140/140	111/113	147/147	107/109	164/164	110/110	129/129	179/179	n/a
NEB	5167c	136/136	140/140	105/107	143/143	103/105	164/173	110/110	132/132	173/173	1
NEB	5167d	153/155	140/140	105/111	145/145	103/109	173/173	110/110	126/126	173/173	14
NEB	5167e	147/153	140/140	105/105	145/150	103/109	164/173	110/110	132/138	175/175	2
NEB	5167f	153/155	140/140	105/111	145/145	103/109	173/173	112/112	126/126	173/173	2
NEB	5167i	149/149	138/140	105/107	143/143	113/115	167/178	112/112	126/129	175/179	1
NEB	5470a	136/136	140/140	105/107	143/143	103/105	164/173	110/110	132/132	173/173	30
NEB	5470b	136/136	140/140	105/107	143/143	103/105	164/164	110/110	132/132	173/173	1
NEB	5470c	ND	140/140	105/107	143/143	103/105	164/173	110/110	132/132	173/173	1
NEB	5470d	153/155	140/140	105/111	145/145	103/109	173/173	110/110	126/126	173/173	1
NEB	b2026	147/155	138/142	103/105	150/150	103/109	167/176	100/110	126/132	171/175	n/a
NEB	xe1313	145/147	138/142	105/109	143/143	105/111	173/173	110/110	126/129	177/177	n/a
NEB	xe5165	151/151	140/142	105/111	145/150	109/109	170/173	110/110	129/132	173/175	n/a
VS	m13a	153/157	138/142	105/115	154/156	111/117	167/178	97/108	129/132	175/175	7
VS	m13g	153/157	138/142	105/115	156/156	111/117	167/167	97/108	129/132	175/175	1
VS	m13h	153/155	140/140	105/111	145/145	103/109	173/173	112/112	126/126	173/173	1
VS	m13i	153/157	138/142	105/115	156/156	111/117	167/178	97/108	129/132	175/175	1
VS	m13j	153/157	138/142	105/115	152/154	111/117	167/178	97/108	ND	175/175	1
VS	m16a	155/157	136/138	107/109	152/152	113/117	170/170	112/112	129/129	171/177	12
VS	m16b	155/157	136/138	107/109	152/156	113/117	170/170	112/112	129/129	171/177	1
VS	m16c	155/157	138/138	107/109	152/152	113/117	170/170	112/112	129/129	171/177	1
VS	m18a	149/155	138/138	105/115	147/147	113/117	176/179	108/112	129/129	171/177	16
VS	m18c	149/155	138/138	105/115	147/147	113/117	173/179	108/112	129/129	171/177	1
VS	m18e	147/153	140/140	105/105	145/150	103/109	173/173	110/110	132/138	175/175	1
VS	m18f	149/155	138/138	105/115	147/147	113/117	176/176	108/112	129/129	171/177	2
VS	m18g	153/157	138/142	105/115	154/156	111/117	167/178	97/106	129/132	175/175	1
VS	m18h	153/157	138/142	105/115	154/156	111/117	167/178	97/108	129/132	175/175	9
VS	m18i	149/151	136/138	105/107	143/147	111/113	ND	106/112	118/129	175/175	1
VS	m7a	149/149	136/140	107/109	154/156	113/113	167/176	108/112	132/132	171/175	25
VS	m7c	ND	ND	107/109	154/156	113/113	167/173	108/112	ND	175/175	1
VS	m7e	149/149	138/140	105/107	147/154	113/117	170/176	108/112	132/132	175/177	1
VS	m7f	149/149	136/140	107/109	154/156	113/113	176/178	108/112	132/132	171/175	1
VS	m7g	153/155	ND	105/111	145/145	103/109	173/173	110/110	126/126	173/173	1
VS	para39	149/155	136/140	109/109	152/152	115/117	170/176	97/110	129/129	175/179	n/a

[a] Primer code. [b] Frequency of MLG within infrapopulation, ND - No Data
Boxed and bold diplotypes indicate heterozygotes with corresponding homozygotes at the same locus across all populations.
Italicised and underlined diplotypes indicate homozygotes that correspond to heterozygotes at the same locus across all populations

9 References

1.   de Souza, W. (2003) In *American Trpanosomaisis* eds. Tyler, KM, Miles MA (Kluwer publishing).

2.   Thomas, S., Martinez, L. L., Westenberger, S. J. & Sturm, N. R. (2007) A population study of the minicircles in *Trypanosoma cruzi*: predicting guide RNAs in the absence of empirical RNA editing *BMC Genomics* **8,** 133.

3.   Telleria, J., Lafay, B., Virreira, M., Barnabe, C., Tibayrenc, M., *et al.* (2006) *Trypanosoma cruzi*: sequence analysis of the variable region of kinetoplast minicircles *Exp Parasitol* **114,** 279-88.

4.   Gibson, W. C. & Miles, M. A. (1986) The karyotype and ploidy of *Trypanosoma cruzi Embo J* **5,** 1299-305.

5.   Henriksson, J., Aslund, L., Macina, R. A., Franke de Cazzulo, B. M., Cazzulo, J. J., *et al.* (1990) Chromosomal localization of seven cloned antigen genes provides evidence of diploidy and further demonstration of karyotype variability in *Trypanosoma cruzi Mol Biochem Parasitol* **42,** 213-23.

6.   El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., *et al.* (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease *Science* **309,** 409-15.

7.   El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., *et al.* (2005) Comparative genomics of trypanosomatid parasitic protozoa *Science* **309,** 404-9.

8.   Tait, A., Turner, C. M., Le Page, R. W. & Wells, J. M. (1989) Genetic evidence that metacyclic forms of *Trypanosoma brucei* are diploid *Mol Biochem Parasitol* **37,** 247-55.

9.   Tyler, K. M., Olson, C. L. & Engman, D. M. (2003) in *American Trypanosmiasis*, eds. Tyler, K. M. & Miles, M. A. (Kluwer publishing)

10. Branche, C., Ochaya, S., Aslund, L. & Andersson, B. (2006) Comparative karyotyping as a tool for genome structure analysis of *Trypanosoma cruzi Mol Biochem Parasitol* **147**, 30-8.

11. Hope, M., MacLeod, A., Leech, V., Melville, S., Sasse, J., *et al.* (1999) Analysis of ploidy (in megabase chromosomes) in *Trypanosoma brucei* after genetic exchange *Mol Biochem Parasitol* **104**, 1-9.

12. Andrade, L. O., Machado, C. R. S., Chiari, E., Pena, S. D. J. & Macedo, A. M. (2002) *Trypanosoma cruzi*: role of host genetic background in the differential tissue distribution of parasite clonal populations *Experimental Parasitology* **100**, 269-275.

13. Sibley, L. D. A., Andrews N.W. (2000) Cell invasion by un-palatable parasites *Traffic* **1**, 100-6.

14. Buscaglia, C. A., Campo, V. A., Frasch, A. C. & Di Noia, J. M. (2006) *Trypanosoma cruzi* surface mucins: host-dependent coat diversity *Nat Rev Microbiol* **4**, 229-36.

15. Tarleton, R. L. (2003) in *American Trypanosmiasis*, eds. Tyler, K. M. & Miles, M. A. (Kluwer, Massachusetts)

16. Campbell, D. A., Westenberger, S. J. & Sturm, N. R. (2004) The determinants of Chagas disease: connecting parasite and host genetics *Curr Mol Med* **4**, 549-62.

17. Kumar, S. & Tarleton, R. L. (1998) The relative contribution of antibody production and CD8+ T cell function to immune control of *Trypanosoma cruzi Parasite Immunol* **20**, 207-16.

18. Tarleton, R. L., Grusby, M. J., Postan, M. & Glimcher, L. H. (1996) *Trypanosoma cruzi* infection in MHC-deficient mice: further evidence for the role of both class I- and class II-restricted T cells in immune resistance and disease *Int Immunol* **8**, 13-22.

19. Ferreira, M. S. (1996) The acquired immunodeficiency syndrome and endemic diseases in Brazil *Rev Soc Bras Med Trop* **29**, 531-5.

20. Castro, C., Prata, A. & Macedo, V. (2001) [A 13-year clinical study on 190 chronic chagasic patients from Mambai, Goias, Brazil] *Rev Soc Bras Med Trop* **34**, 309-18.

21. Tarleton, R. L. (2003) Chagas disease: a role for autoimmunity? *Trends Parasitol* **19**, 447-51.

22. Marin-Neto, J. A., Cunha-Neto, E., Maciel, B. C. & Simoes, M. V. (2007) Pathogenesis of chronic Chagas heart disease *Circulation* **115**, 1109-23.

23. Punukollu, G., Gowda, R. M., Khan, I. A., Navarro, V. S. & Vasavada, B. C. (2007) Clinical aspects of the Chagas' heart disease *Int J Cardiol* **115**, 279-83.

24. da Silveira, A. B., Lemos, E. M., Adad, S. J., Correa-Oliveira, R., Furness, J. B., *et al.* (2007) Megacolon in Chagas disease: a study of inflammatory cells, enteric nerves, and glial cells *Hum Pathol* **38**, 1256-1264.

25. Virreira, M., Serrano, G., Maldonado, L. & Svoboda, M. (2006) *Trypanosoma cruzi*: typing of genotype (sub)lineages in megacolon samples from bolivian patients *Acta Trop* **100**, 252-5.

26. Miles, M. A., Cedillos, R. A., Povoa, M. M., de Souza, A. A., Prata, A., *et al.* (1981) Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas' disease? *Lancet* **1**, 1338-40.

27. Clark, C. G. & Pung, O. J. (1994) Host specificity of ribosomal DNA variation in sylvatic *Trypanosoma cruzi* from North America *Mol Biochem Parasitol* **66**, 175-9.

28. Apt, W., Aguilera, X., Arribada, A., Gomez, L., Miles, M. A., *et al.* (1987) Epidemiology of Chagas' disease in northern Chile: isozyme profiles of *Trypanosoma cruzi* from domestic and sylvatic transmission cycles and their association with cardiopathy *Am J Trop Med Hyg* **37**, 302-7.

29. 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**, 225-33.

30. Montalvo-Hicks, L. D., Trevenen, C. L. & Briggs, J. N. (1980) American trypanosomiasis (Chagas' disease) in a Canadian immigrant infant *Pediatrics* **66**, 266-8.

31. Riera, C., Guarro, A., Kassab, H. E., Jorba, J. M., Castro, M., *et al.* (2006) Congenital transmission of *Trypanosoma cruzi* in Europe (Spain): a case report *Am J Trop Med Hyg* **75**, 1078-81.

32. Miles, M. A., Souza, A., Povoa, M., Shaw, J. J., Lainson, R., *et al.* (1978) Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas' disease in Amazonian Brazil *Nature* **272**, 819-21.

33. Walter, A., Lozano-Kasten, F., Bosseno, M. F., Ruvalcaba, E. G., Gutierrez, M. S., *et al.* (2007) Peridomicilary habitat and risk factors for Triatoma infestation in a rural community of the Mexican occident *Am J Trop Med Hyg* **76**, 508-15.

34. Cecere, M. C., Vasquez-Prokopec, G. M., Gurtler, R. E. & Kitron, U. (2006) Reinfestation sources for Chagas disease vector, *Triatoma infestans*, Argentina *Emerg Infect Dis* **12**, 1096-102.

35. Miles, M., Yeo, M. & Gaunt, M. (2004) in *The Trypanosomiases*, eds. Maudlin, I., Holmes, P. & Miles, M. (CABI publishing)

36. Ramirez, C. J., Jaramillo, C. A., del Pilar Delgado, M., Pinto, N. A., Aguilera, G., *et al.* (2005) Genetic structure of sylvatic, peridomestic and domestic populations of *Triatoma dimidiata* (Hemiptera: Reduviidae) from an endemic zone of Boyaca, Colombia *Acta Trop* **93**, 23-9.

37. Shikanai-Yasuda, M. A., Marcondes, C. B., Guedes, L. A., Siqueira, G. S., Barone, A. A., *et al.* (1991) Possible oral transmission of acute Chagas' disease in Brazil *Rev Inst Med Trop Sao Paulo* **33**, 351-7.

38.   Coura, J. R., Junqueira, A. C., Fernandes, O., Valente, S. A. & Miles, M. A.
(2002) Emerging Chagas disease in Amazonian Brazil *Trends Parasitol* **18,** 171-
6.

39.   Borges, E. C., Dujardin, J. P., Schofield, C. J., Romanha, A. J. & Diotaiuti, L.
(2005) Dynamics between sylvatic, peridomestic and domestic populations of
Triatoma brasiliensis (Hemiptera: Reduviidae) in Ceara State, Northeastern
Brazil *Acta Trop* **93,** 119-26.

40.   Matias, A., de la Riva, J., Martinez, E., Torrez, M. & Dujardin, J. P. (2003)
Domiciliation process of *Rhodnius stali* (Hemiptera: Reduviidae) in Alto Beni,
La Paz, Bolivia *Trop Med Int Health* **8,** 264-8.

41.   Carrasco, H. J., Torrellas, A., Garcia, C., Segovia, M. & Feliciangeli, M. D.
(2005) Risk of *Trypanosoma cruzi* I (Kinetoplastida: Trypanosomatidae)
transmission by *Panstrongylus geniculatus* (Hemiptera: Reduviidae) in Caracas
(Metropolitan District) and neighboring States, Venezuela *Int J Parasitol* **35,**
1379-84.

42.   Fitzpatrick, S. (2005) (The London School of Hygiene and Tropical Medicine,
London), PhD thesis

43.   Cortez, M. R., Emperaire, L., Piccinali, R., Gurtler, R. E., Torrico, F., *et al.*
(2007) Sylvatic *Triatoma infestans* (Reduviidae, Triatominae) in the Andean
valleys of Bolivia *Acta Trop* **102,** 47-54.

44.   Noireau, F., Cortez, M. G., Monteiro, F. A., Jansen, A. M. & Torrico, F. (2005)
Can wild *Triatoma infestans* foci in Bolivia jeopardize Chagas disease control
efforts? *Trends Parasitol* **21,** 7-10.

45.   Munoz, J., Portus, M., Corachan, M., Fumado, V. & Gascon, J. (2007)
Congenital *Trypanosoma cruzi* infection in a non-endemic area *Trans R Soc
Trop Med Hyg*.

46.   Torrico, F., Alonso-Vega, C., Suarez, E., Rodriguez, P., Torrico, M. C., *et al.*
(2004) Maternal *Trypanosoma cruzi* infection, pregnancy outcome, morbidity,

and mortality of congenitally infected and non-infected newborns in Bolivia *Am J Trop Med Hyg* **70**, 201-9.

47. Blanco, S. B., Segura, E. L., Cura, E. N., Chuit, R., Tulian, L., *et al.* (2000) Congenital transmission of *Trypanosoma cruzi*: an operational outline for detecting and treating infected infants in north-western Argentina *Trop Med Int Health* **5**, 293-301.

48. (2006) Chagas disease after organ transplantation--Los Angeles, California, 2006 *MMWR Morb Mortal Wkly Rep* **55**, 798-800.

49. Schmunis, G. A., Zicker, F., Pinheiro, F. & Brandling-Bennett, D. (1998) Risk for transfusion-transmitted infectious diseases in Central and South America *Emerg Infect Dis* **4**, 5-11.

50. Anon (2007) Blood donor screening for chagas disease--United States, 2006-2007 *MMWR Morb Mortal Wkly Rep* **56**, 141-3.

51. Raasi, A. & Luquetti, A. (2003) in *American Trypanosmiasis*, ed. Tyler, K. M. Miles, M.A. (Kluwer publishing).

52. Priotto, G., Fogg, C., Balasegaram, M., Erphas, O., Louga, A., *et al.* (2006) Three drug combinations for late-stage *Trypanosoma brucei gambiense* sleeping sickness: a randomized clinical trial in Uganda *PLoS Clin Trials* **1**, e39.

53. Harder, A., Greif, G. & Haberkorn, A. (2001) Chemotherapeutic approaches to protozoa: kinetoplastida--current level of knowledge and outlook *Parasitol Res* **87**, 778-80.

54. Wilkinson, S. R., Taylor, M. C., Horn, D., Kelly, J. M. & Cheeseman, I. (2008) A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes *Proc Natl Acad Sci U S A* **105**, 5022-7.

55. Garcia, S., Ramos, C. O., Senra, J. F., Vilas-Boas, F., Rodrigues, M. M., *et al.* (2005) Treatment with benznidazole during the chronic phase of experimental Chagas' disease decreases cardiac alterations *Antimicrob Agents Chemother* **49**, 1521-8.

56. Viotti, R., Vigliano, C., Lococo, B., Bertocchi, G., Petti, M., *et al.* (2006) Long-term cardiac outcomes of treating chronic Chagas disease with benznidazole versus no treatment: a nonrandomized trial *Ann Intern Med* **144**, 724-34.

57. Croft, S. L., Barrett, M. P. & Urbina, J. A. (2005) Chemotherapy of trypanosomiases and leishmaniasis *Trends Parasitol* **21**, 508-12.

58. Dias, J. C., Silveira, A. C. & Schofield, C. J. (2002) The impact of Chagas disease control in Latin America: a review *Mem Inst Oswaldo Cruz* **97**, 603-12. ·

59. http://www.who.int/tdr/publications/tdrnews/news65/chagas.htm.

60. http://www.who.int/tdr/research/progress/chag_afr/regional.htm.

61. http://www.who.int/tdr/research/progress/chag_afr/southcone.htm.

62. http://www.paho.org/english/AD/DPC/CD/dch-andina-vi-2005.htm.

63. Rojas de Arias, A., Lehane, M. J., Schofield, C. J. & Maldonado, M. (2004) Pyrethroid insecticide evaluation on different house structures in a Chagas disease endemic area of the Paraguayan Chaco *Mem Inst Oswaldo Cruz* **99**, 657-62.

64. Kroeger, A., Villegas, E., Ordonez-Gonzalez, J., Pabon, E. & Scorza, J. V. (2003) Prevention of the transmission of Chagas' disease with pyrethroid-impregnated materials *Am J Trop Med Hyg* **68**, 307-11.

65. Sanchez-Martin, M. J., Feliciangeli, M. D., Campbell-Lendrum, D. & Davies, C. R. (2006) Could the Chagas disease elimination programme in Venezuela be compromised by reinvasion of houses by sylvatic *Rhodnius prolixus* bug populations? *Trop Med Int Health* **11**, 1585-93.

66. Ceballos, L. A., Cardinal, M. V., Vazquez-Prokopec, G. M., Lauricella, M. A., Orozco, M. M., *et al.* (2006) Long-term reduction of *Trypanosoma cruzi* infection in sylvatic mammals following deforestation and sustained vector surveillance in northwestern Argentina *Acta Trop* **98**, 286-96.

67. Schmunis, G. A. (1999) Prevention of transfusional *Trypanosoma cruzi* infection in Latin America *Mem Inst Oswaldo Cruz* **94 Suppl 1**, 93-101.

68. Wendel, S. & Gonzaga, A. L. (1993) Chagas' disease and blood transfusion: a New World problem? *Vox Sang* **64**, 1-12.

69. Moraes-Souza, H., Bordin, J. O., Bardossy, L., MacPherson, D. W. & Blajchman, M. A. (1995) Prevention of transfusion-associated Chagas' disease: efficacy of white cell-reduction filters in removing *Trypanosoma cruzi* from infected blood *Transfusion* **35**, 723-6.

70. Ramirez, L. E., Lages-Silva, E., Pianetti, G. M., Rabelo, R. M., Bordin, J. O., *et al.* (1995) Prevention of transfusion-associated Chagas' disease by sterilization of *Trypanosoma cruzi*-infected blood with gentian violet, ascorbic acid, and light *Transfusion* **35**, 226-30.

71. Dias, J. C. (1987) Control of Chagas disease in Brazil *Parasitol Today* **3**, 336-41.

72. Schofield, C. J. & Dias, J. C. (1999) The Southern Cone Initiative against Chagas disease *Adv Parasitol* **42**, 1-27.

73. Moncayo, A. (1999) Progress towards interruption of transmission of Chagas disease *Mem Inst Oswaldo Cruz* **94**, 401-4.

74. Feliciangeli, M. D., 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**, 44-9.

75. Feliciangeli, M. D., Sanchez-Martin, M. J., Suarez, B., Marrero, R., Torrellas, A., *et al.* (2007) Risk Factors for *Trypanosoma cruzi* Human Infection in Barinas State, Venezuela *Am J Trop Med Hyg* **76**, 915-921.

76. Anon, A. (2004) International meeting on suveillance and prevention of Chagas disease in the Amazonian region (Manuas, Brazil).

77. Aguilar, H. M., Abad-Franch, F., Dias, J. C., Junqueira, A. C. & Coura, J. R. (2007) Chagas disease in the Amazon Region *Mem Inst Oswaldo Cruz* **102 Suppl 1,** 47-56.

78. Valente, S. A., Valente, V.C, Neto, H.F. (1999) Considerations on the Epidemiology and Transmission of Chagas Disease in the Brazilian Amazon *Mem Inst Oswaldo Cruz* **94,** 395-398.

79. http://www.cdc.gov/chagas/factsheets/detailed.html

80. http://www.who.int/tdr/publications/tdrnews/news65/chagas.htm.

81. Schofield, C. J., Jannin, J. & Salvatella, R. (2006) The future of Chagas disease control *Trends Parasitol* **22,** 583-8.

82. Ache, A. & Matos, A. J. (2001) Interrupting Chagas disease transmission in Venezuela *Rev Inst Med Trop Sao Paulo* **43,** 37-43.

83. Moncoya, A in Global epidmiology of infectious diseases eds. Murray, J.L. Lopez, A.D. & Mathers C.D. (WHO publications)

84. Ramirez-Perez, J. (1987) Revision de los triatominos (Hemiptra: Reduviidae) en Venezuela. *Boletin de la direccion de Malariologia y Saneamiento Ambiental* **27,** 118-146.

85. Feliciangeli, M. D., Carrasco, H., Patterson, J. S., Suarez, B., Martinez, C., *et al.* (2004) Mixed domestic infestation by *Rhodnius prolixus* Stal, 1859 and Panstrongylus geniculatus Latreille, 1811, vector incrimination, and seroprevalence for *Trypanosoma cruzi* among inhabitants in El Guamito, Lara State, Venezuela *Am J Trop Med Hyg* **71,** 501-5.

86. de Lima, H. C., J. Rodríguez, A. de Guglielmo, Z. Rodríguez, N. (2006) [Trypanosomatidae of public health importance occurring in wild and synanthropic animals of rural Venezuela] *Biomedica* **26,** 42-50.

87. http://www.imfstatistics.org/imf/

88.   http://www.msf.org/msfinternational/invoke.cfm?objectid=F7D228EA-E018-0C72-09F866FA4EB2CFB6&component=toolkit.article&method=full_html.

89.   Becht, D.K. (2001) Kissing the assasin bug goodbye. Perspecives In Health (PAHO) Vol 6.

90.   Pless, M., Juranek, D., Kozarsky, P., Steurer, F., Tapia, G., *et al.* (1992) The epidemiology of Chagas' disease in a hyperendemic area of Cochabamba, Bolivia: a clinical study including electrocardiography, seroreactivity to *Trypanosoma cruzi*, xenodiagnosis, and domiciliary triatomine distribution *Am J Trop Med Hyg* **47,** 539-46.

91.   http://www.iadb.org/idbamerica/Archive/stories/1999/eng/e499j2.htm

92.   Pizarro, J. C., Lucero, D. E. & Stevens, L. (2007) PCR reveals significantly higher rates of *Trypanosoma cruzi* infection than microscopy in the Chagas vector, Triatoma infestans: high rates found in Chuquisaca, Bolivia *BMC Infect Dis* **7,** 66.

93.   Carrasco, R., Miguez, H., Camacho, C., Echalar, L., Revollo, S., *et al.* (1990) Prevalence of *Trypanosoma cruzi* infection in blood banks of seven departments of Bolivia *Mem Inst Oswaldo Cruz* **85,** 69-73.

94.   Klein, H. S. (2003) *A Concise History of Bolivia* (Cambridge University Press, Cambridge).

95.   Bargues, M. D., Klisiowicz, D. R., Panzera, F., Noireau, F., Marcilla, A., *et al.* (2006) Origin and phylogeography of the Chagas disease main vector *Triatoma infestans* based on nuclear rDNA sequences and genome size *Infect Genet Evol* **6,** 46-62.

96.   Noireau, F., Bosseno, M. F., Carrasco, R., Telleria, J., Vargas, F., *et al.* (1995) Sylvatic triatomines (Hemiptera: Reduviidae) in Bolivia: trends toward domesticity and possible infection with *Trypanosoma cruzi* (Kinetoplastida: Trypanosomatidae) *J Med Entomol* **32,** 594-8.

97.   Noireau, F., Gutierrez, T., Flores, R., Breniere, F., Bosseno, M. F., *et al.* (1999) Ecogenetics of *Triatoma sordida* and *Triatoma guasayana* (Hemiptera: reduviidae) in the Bolivian chaco *Mem Inst Oswaldo Cruz* **94**, 451-7.

98.   Brutus, L., Schneider, D., Postigo, J., Delgado, W., Mollinedo, S., *et al.* (2007) Evidence of congenital transmission of *Trypanosoma cruzi* in a vector-free area of Bolivia *Trans R Soc Trop Med Hyg*.

99.   Miles, M. A. Yeo, M. Gaunt, M.W (2003) In *American Trypanosmiasis* eds. Tyler K.M. & Miles M.A. (Kluwer publshing).

100.   Gaunt, M. & Miles, M. (2000) The ecotopes and evolution of triatomine bugs (triatominae) and their associated trypanosomes *Mem Inst Oswaldo Cruz* **95**, 557-65.

101.   Lent, H. & Wygodzinksy, P. (1979) Revision of the Triatominae and their significance as vectors of Chagas disease. *Bull Am Mus Nat Hist* **163**, 123-520.

102.   Carcavallo, R. U., Curto de Casas, S. I., Sherlock, I. A., Giron, I. G., Jurberg, J., *et al.* (1999) *Atlas of Chagas disease vectors in the Americas* (Editora Fiocruz)

103.   Maldonado-Capriles, J. (1990) Systematic Catalogue of the Reduviidae of the World (Insecta: Heteroptera) *A special edition of Caribbean Journal of Science.*

104.   Gaunt, M. W. & Miles, M. A. (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks *Mol Biol Evol* **19**, 748-61.

105.   Hypsa, V., Tietz, D. F., Zrzavy, J., Rego, R. O., Galvao, C., *et al.* (2002) Phylogeny and biogeography of Triatominae (Hemiptera: Reduviidae): molecular evidence of a New World origin of the Asiatic clade *Mol Phylogenet Evol* **23**, 447-57.

106.   Marcilla, A., Bargues, M. D., Abad-Franch, F., Panzera, F., Carcavallo, R. U., *et al.* (2002) Nuclear rDNA ITS-2 sequences reveal polyphyly of *Panstrongylus species* (Hemiptera: Reduviidae: Triatominae), vectors of *Trypanosoma cruzi Infect Genet Evol* **1**, 225-35.

107. Marcilla, A., Bargues, M. D., Ramsey, J. M., Magallon-Gastelum, E., Salazar-Schettino, P. M., *et al.* (2001) The ITS-2 of the nuclear rDNA as a molecular marker for populations, species, and phylogenetic relationships in Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease *Mol Phylogenet Evol* **18**, 136-42.

108. Gorla, D. E., Dujardin, J. P. & Schofield, C. J. (1997) Biosystematics of Old World Triatominae *Acta Trop* **63**, 127-40.

109. Galvao, C., Patterson, J. S., Da Silva Rocha, D., Jurberg, J., Carcavallo, R., *et al.* (2002) A new species of Triatominae from Tamil Nadu, India *Med Vet Entomol* **16**, 75-82.

110. Schofield, C. J. (2000) Biosystematics and evolution of the Triatominae *Cad Saude Publica* **16**, 89-92.

111. Ribeiro, J. M., Schneider, M., Isaias, T., Jurberg, J., Galvao, C., *et al.* (1998) Role of salivary antihemostatic components in blood feeding by triatomine bugs (Heteroptera) *J Med Entomol* **35**, 599-610.

112. Patterson, J. S. (2007) ITD (London School of Hygiene and tropical Medicine, London), PhD thesis

113. de Paula, A. S., Diotaiuti, L. & Schofield, C. J. (2005) Testing the sister-group relationship of the Rhodniini and Triatomini (Insecta: Hemiptera: Reduviidae: Triatominae) *Mol Phylogenet Evol* **35**, 712-8.

114. Pereira, M. H., Gontijo, N. F., Guarneri, A. A., Sant'anna, M. R. & Diotaiuti, L. (2006) Competitive displacement in Triatominae: the *Triatoma infestans* success *Trends Parasitol* **22**, 516-20.

115. Noireau, F., Flores, R., Gutierrez, T. & Dujardin, J. P. (1997) Detection of sylvatic dark morphs of *Triatoma infestans* in the Bolivian Chaco *Mem Inst Oswaldo Cruz* **92**, 583-4.

116. Richer, W., Kengne, P., Cortez, M. R., Perrineau, M. M., Cohuet, A., *et al.* (2007) Active dispersal by wild *Triatoma infestans* in the Bolivian Andes *Trop Med Int Health* **12**, 759-64.

117. Harry, M. (1993) Use of the median process of the pygophore in the identification of *Rhodnius nasutus, R. neglectus, R. prolixus* and *R. robustus* (Hemiptera: Reduviidae). *Ann Trop Med Parasitol* **87**, 277-282.

118. Monteiro, F. A., Barrett, T. V., 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.

119. Ponce, C. (2007) Current situation of Chagas disease in Central America *Mem Inst Oswaldo Cruz.* **102**, 41- 44

120. Dumonteil, E., Tripet, F., Ramirez-Sierra, M. J., Payet, V., Lanzaro, G., *et al.* (2007) Assessment of *Triatoma dimidiata* dispersal in the Yucatan Peninsula of Mexico by morphometry and microsatellite markers *Am J Trop Med Hyg* **76**, 930-7.

121. Monroy, M. C., Bustamante, D. M., Rodas, A. G., Enriquez, M. E. & Rosales, R. G. (2003) Habitats, dispersion and invasion of sylvatic *Triatoma dimidiata* (Hemiptera: Reduviidae: Triatominae) in Peten, Guatemala *J Med Entomol* **40**, 800-6.

122. Oscherov, E. B., Damborsky, M. P., Bar, M. E. & Gorla, D. E. (2004) Competition between vectors of Chagas disease, *Triatoma infestans* and *T. sordida:* effects on fecundity and mortality *Med Vet Entomol* **18**, 323-8.

123. Cuba, C. A., Abad-Franch, F., Roldan Rodriguez, J., Vargas Vasquez, F., Pollack Velasquez, L., *et al.* (2002) The triatomines of northern Peru, with emphasis on the ecology and infection by trypanosomes of *Rhodnius ecuadoriensis* (Triatominae) *Mem Inst Oswaldo Cruz* **97**, 175-83.

124. Grijalva, M. J., Palomeque-Rodriguez, F. S., Costales, J. A., Davila, S. & Arcos-Teran, L. (2005) High household infestation rates by synanthropic vectors of Chagas disease in southern Ecuador *J Med Entomol* **42**, 68-74.

125. Valente, V. C., Valente, S. A., Noireau, F., Carrasco, H. J. & Miles, M. A. (1998) Chagas disease in the Amazon Basin: association of *Panstrongylus geniculatus* (Hemiptera: Reduviidae) with domestic pigs *J Med Entomol* **35**, 99-103.

126. Barbosa, S. E., Belisario, C. J., Souza, R. C., Paula, A. S., Linardi, P. M., *et al.* (2006) Biogeography of Brazilian populations of *Panstrongylus megistus* (Hemiptera, Reduviidae, Triatominae) based on molecular marker and paleo-vegetational data *Acta Trop* **99**, 144-54.

127. Damborsky, M. P., Bar, M. E. & Oscherov, E. B. (2001) [Detection of triatomines (Hemiptera: Reduviidae) in domiciliary and extra-domiciliary ecotopes. Corrientes, Argentina.] *Cad Saude Publica* **17**, 843-9.

128. Abad-Franch, F., Aguilar, V. H., Paucar, C. A., Lorosa, E. S. & Noireau, F. (2002) Observations on the domestic ecology of *Rhodnius ecuadoriensis* (Triatominae) *Mem Inst Oswaldo Cruz* **97**, 199-202.

129. Abad-Franch, F., Paucar, A., Carpio, C., Cuba, C. A., Aguilar, H. M., *et al.* (2001) Biogeography of Triatominae (Hemiptera: Reduviidae) in Ecuador: implications for the design of control strategies *Mem Inst Oswaldo Cruz* **96**, 611-20.

130. Zeledon, R., Marin, F., Calvo, N., Lugo, E. & Valle, S. (2006) Distribution and ecological aspects of *Rhodnius pallescens* in Costa Rica and Nicaragua and their epidemiological implications *Mem Inst Oswaldo Cruz* **101**, 75-9.

131. Vazquez-Prokopec, G. M., Cecere, M. C., Canale, D. M., Gurtler, R. E. & Kitron, U. (2005) Spatiotemporal patterns of reinfestation by *Triatoma guasayana* (Hemiptera: Reduviidae) in a rural community of northwestern Argentina *J Med Entomol* **42**, 571-81.

132. Corte, L. A. & Suarez, H. A. (2005) [Triatomines (Reduviidae: Triatominae) in a Chagas disease focus in Talaigua Nuevo (Bolivar, Colombia)] *Biomedica* **25**, 568-74.

133. Rojas de Arias, A., de Guillen, I., Inchausti, A., Samudio, M. & Schmeda-Hirschmann, G. (1993) Prevalence of Chagas' disease in Ayoreo communities of the Paraguayan Chaco *Trop Med Parasitol* **44**, 285-8.

134. Poinar, G., Jr. (2005) *Triatoma dominicana sp.* n. (Hemiptera: Reduviidae: Triatominae), and *Trypanosoma antiquus* sp. n. (Stercoraria: Trypanosomatidae), the first fossil evidence of a triatomine-trypanosomatid vector association *Vector Borne Zoonotic Dis* **5**, 72-81.

135. Abad-Franch, F., Palomeque, F. S., Aguilar, H. M. t. & Miles, M. A. (2005) Field ecology of sylvatic *Rhodnius* populations (Heteroptera, Triatominae): risk factors for palm tree infestation in western Ecuador *Trop Med Int Health* **10**, 1258-66.

136. Miles, M. A. (1979) In *Biology of the Kinetoplastida* (Academic Press, London).

137. Dias-Lima, A. G., Menezes, D., Sherlock, I. & Noireau, F. (2003) Wild habitat and related fauna of *Panstrongylus lutzi* (Reduviidae, Triatominae) *J Med Entomol* **40**, 989-90.

138. D'Alessandro, A., Barreto, P., Saravia, N. & Barreto, M. (1984) Epidemiology of *Trypanosoma cruzi* in the oriental plains of Colombia *Am J Trop Med Hyg* **33**, 1084-95.

139. Patterson, J. S., Schofield, C. J., Dujardin, J. P. & Miles, M. A. (2001) Population morphometric analysis of the tropicopolitan bug *Triatoma rubrofasciata* and relationships with old world species of Triatoma: evidence of New World ancestry *Med Vet Entomol* **15**, 443-51.

140. Hoare, C. A. (1972) *The trypanosomes of mammals* (Blackwell Scientific Publications)

141. Harvey Pough, F. J., C. Heiser, J.B. (1999) *Vertebrate Life* (Prentice Hall Publishing).

142. Janis, C. (1993) Tertiary mammal evolution in the context of changing climates, vegetation, and tectonic events. *Ann Rev Ecol Syst* **24**, 447-57.

143. Krause, D. W., G.V.R., P., von Koenigswald, W., Sahni, A. & Grine, F. E. (1997) Cosmopolitanism among late Cretaceous Gondwanan mammals *Nature* **390,** 504-507.

144. Simpson, G. G. (1980) *Splendid Isolation: the curious history of South American mammals* (Yale University Press)

145. Patterson, B. P., Pascual R. (1972) in *Evolution mammals and Southern Continents* eds. Keast, A. Erk, F.C. Glass, B.P. (State University of New York Press)

146. Novacek, M. J. (1992) Mammalian phylogeny: shaking the tree *Nature* **356,** 121-5.

147. Delsuc, F., Scally, M., Madsen, O., Stanhope, M. J., de Jong, W. W., *et al.* (2002) Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Molecular Biology and Evolution* **19,** 1656-71.

148. Hallstrom, B. M., Kullberg, M., Nilsson, M. A. & Janke, A. (2007) Phylogenomic Data Analyses Provide Evidence that Xenarthra and Afrotheria Are Sister Groups *Mol Biol Evol* **24,** 2059-68.

149. Delsuc, F., Vizcaino, S. F. & Douzery, E. J. (2004) Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans *BMC Evol Biol* **4,** 11.

150. Moller-Krull, M., Delsuc, F., Churakov, G., Marker, C., Superina, M., *et al.* (2007) Retroposed Elements and their Flanking Regions Resolve the Evolutionary History of Xenarthran Mammals (Armadillos, Anteaters and Sloths) *Mol Biol Evol* **24,** 2573-2582.

151. Steiner, C., Tilak, M. K., Douzery, E. J. & Catzeflis, F. M. (2005) New DNA data from a transthyretin nuclear intron suggest an Oligocene to Miocene diversification of living South America opossums (Marsupialia: Didelphidae) *Mol Phylogenet Evol* **35,** 363-79.

152. Marshall, L. & Sempere, T. (1993) *Evolution of the Neotropical land mammal fauna in its geochronologic, stratigraphic and tectonic context* (Yale University Press)

153. Wyss, A. R. F., Norell, J. J. Swisher, M. A. Charrier, R. Novacek, M. J. McKenna, M. C. (1993) South America's earliest rodent and recognition of a new interval of mammalian evolution. *Nature* **365**, 434–37.

154. Huchon, D. & Douzery, E. J. (2001) From the Old World to the New World: a molecular chronicle of the phylogeny and biogeography of hystricognath rodents *Mol Phylogenet Evol* **20**, 238-51.

155. Martin, T. (1994) African Origin of Caviomorph Rodents is Indicated by Incisor Enamel Microstructure *Paleobiology* **20**, 5-13.

156. Eisenberg, J. F. R., K.H. (1999) *Mammals of the Neotropics* (University of Chicago Press)

157. Stevens, J. & Brisse, S. (2004) in *The Trypanosomiases*, eds. Maudlin, I., Holmes, P. & Miles, M. (CABI publishing)

158. Urrea, D. A., Carranza, J. C., Cuba, C. A., Gurgel-Goncalves, R., Guhl, F., *et al.* (2005) Molecular characterisation of *Trypanosoma rangeli* strains isolated from *Rhodnius ecuadoriensis* in Peru, *R. colombiensis* in Colombia and *R. pallescens* in Panama, supports a co-evolutionary association between parasites and vectors *Infect Genet Evol* **5**, 123-9.

159. Stevens, J. R. & Gibson, W. (1999) The molecular evolution of trypanosomes *Parasitol Today* **15**, 432-7.

160. Stevens, J. R., Noyes, H. A., Dover, G. A. & Gibson, W. C. (1999) The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi Parasitology* **118**, 107-16.

161. Simpson, A. G. B., Stevens, J. R. & Lukes, J. (2006) The evolution and diversity of kinetoplastid flagellates *Trends in Parasitology* **22**, 168-174.

162. Hamilton, P. B., Stevens, J. R., Gaunt, M. W., Gidley, J. & Gibson, W. C. (2004) Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA *Int J Parasitol* **34**, 1393-404.

163. Chicharro, C. & Alvar, J. (2003) Lower trypanosomatids in HIV/AIDS patients *Ann Trop Med Parasitol* **97**, 75-8.

164. Miles, M. A., Povoa, M. M., de Souza, A. A., Lainson, R., Shaw, J. J., *et al.* (1981) Chagas's disease in the Amazon Basin: Ii. The distribution of *Trypanosoma cruzi* zymodemes 1 and 3 in Para State, north Brazil *Trans R Soc Trop Med Hyg* **75**, 667-74.

165. 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 Roy Soc Trop Medicine Hyg* **71**, 217-225.

166. Anon, A. (1999) Recommendations from a satellite meeting. *Mem Inst Oswaldo Cruz* **94**, 429-32.

167. Tibayrenc, M. & Miles, M. A. (1983) A genetic comparison between Brazilian and Bolivian zymodemes of *Trypanosoma cruzi Trans R Soc Trop Med Hyg* **77**, 76-83.

168. Tibayrenc, M., Echalar, L., Dujardin, J. P., Poch, O. & Desjeux, P. (1984) The microdistribution of isoenzymic strains of *Trypanosoma cruzi* in southern Bolivia; new isoenzyme profiles and further arguments against Mendelian sexuality *Trans R Soc Trop Med Hyg* **78**, 519-25.

169. Johnson, F. M., Kanapi, C. G., Richardson, R. H., Wheeler, M. R. & Stone, W. S. (1966) An analysis of polymorphisms among isozyme loci in dark and light *Drosophila ananassae* strains from American and Western Samoa *Proc Natl Acad Sci U S A* **56**, 119-25.

170. Steindel, M., Kramer Pacheco, L., Scholl, D., Soares, M., de Moraes, M. H., *et al.* (2007) Characterization of *Trypanosoma cruzi* isolated from humans,

vectors, and animal reservoirs following an outbreak of acute human Chagas disease in Santa Catarina State, Brazil *Diagn Microbiol Infect Dis* **60**, 25-32.

171. Chapman, M., Baggaley, R., Godfrey-Fausset, P., Malpas, T., White, G., *et al.* (1984) *Trypanosoma cruzi* from the Paraguayan Chaco: isoenzyme profiles of strains isolated at Makthlawaiya. *Journal of Protozoology* **31**, 482-486.

172. Ruiz-Garcia, M., Montilla, M., Nicholls, S. & Alvarez, D. (2001) Population genetic analysis of Colombian *Trypanosoma cruzi* isolates revealed by enzyme electrophoretic profiles *Mem Inst Oswaldo Cruz* **96**, 31-51.

173. Jaramillo, N., Moreno, J., Triana, O., Arcos-Burgos, M., Munoz, S., *et al.* (1999) Genetic structure and phylogenetic relationships of Colombian *Trypanosoma cruzi* populations as determined by schizodeme and isoenzyme markers *Am J Trop Med Hyg* **61**, 986-93.

174. Saravia, N. G., Holguin, A. F., Cibulskis, R. E. & D'Alessandro, A. (1987) Divergent isoenzyme profiles of sylvatic and domiciliary *Trypanosoma cruzi* in the eastern plains, piedmont, and highlands of Colombia *Am J Trop Med Hyg* **36**, 59-69.

175. Tibayrenc, M., Neubauer, K., Barnabe, C., Guerrini, F., Skarecky, D., *et al.* (1993) Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis *Proc Natl Acad Sci U S A* **90**, 1335-9.

176. Urdaneta, L., Lal, A., Barnabe, C., Oury, B., Goldman, I., *et al.* (2001) Evidence for clonal propagation in natural isolates of *Plasmodium falciparum* from Venezuela *Proc Natl Acad Sci U S A* **98**, 6725-9.

177. Ferreira Ade, M., Vitor, R. W., Carneiro, A. C., Brandao, G. P. & Melo, M. N. (2004) Genetic variability of Brazilian *Toxoplasma gondii* strains detected by random amplified polymorphic DNA-polymerase chain reaction (RAPD-PCR) and simple sequence repeat anchored-PCR (SSR-PCR) *Infect Genet Evol* **4**, 131-42.

178. Botilde, Y., Laurent, T., Quispe Tintaya, W., Chicharro, C., Canavate, C., et al. (2006) Comparison of molecular markers for strain typing of *Leishmania infantum Infect Genet Evol* **6**, 440-6.

179. Komba, E. K., Kibona, S. N., Ambwene, A. K., Stevens, J. R. & Gibson, W. C. (1997) Genetic diversity among *Trypanosoma brucei rhodesiense* isolates from Tanzania *Parasitology* **115**, 571-9.

180. Carrasco, H. (1997) ITD (London School of Hygiene and Tropical Medicine, London), Vol. PhD, pp. 172.

181. Carrasco, H. J., Frame, I. A., Valente, S. A. & Miles, M. A. (1996) Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi Am J Trop Med Hyg* **54**, 418-24.

182. Stothard, J., Frame, I. & Miles, M. (1999) Genetic diversity and genetic exchange in *Trypanosoma cruzi*: dual drug-resistant "progeny" from episomal transformants *Mem Inst Oswaldo Cruz* **94**, 189-93.

183. Gaunt, M. W., Yeo, M., Frame, I. A., Stothard, J. R., Carrasco, H. J., et al. (2003) Mechanism of genetic exchange in American trypanosomes *Nature* **421**, 936-9.

184. Souto, R. P., Fernandes, O., Macedo, A. M., Campbell, D. A. & Zingales, B. (1996) DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi Mol Biochem Parasitol* **83**, 141-52.

185. Souto, R. P., Vargas, N. & Zingales, B. (1999) *Trypanosoma rangeli*: discrimination from *Trypanosoma cruzi* based on a variable domain from the large subunit ribosomal RNA gene *Exp Parasitol* **91**, 306-14.

186. 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**, 1218-26.

187. Fernandes, O., Santos, S. S., Cupolillo, E., Mendonca, B., Derre, R., et al. (2001) A mini-exon multiplex polymerase chain reaction to distinguish the

major groups of *Trypanosoma cruzi* and *T. rangeli* in the Brazilian Amazon *Trans R Soc Trop Med Hyg* **95**, 97-9.

188. Bosseno, M. F., Telleria, J., Vargas, F., Yaksic, N., Noireau, F., *et al.* (1996) *Trypanosoma cruzi*: study of the distribution of two widespread clonal genotypes in Bolivian Triatoma infestans vectors shows a high frequency of mixed infections *Exp Parasitol* **83**, 275-82.

189. Fernandes, A. P., Nelson, K. & Beverley, S. M. (1993) Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism *Proc Natl Acad Sci U S A* **90**, 11608-12.

190. Marcet, P. L., Duffy, T., Cardinal, M. V., Burgos, J. M., Lauricella, M. A., *et al.* (2006) PCR-based screening and lineage identification of *Trypanosoma cruzi* directly from faecal samples of triatomine bugs from northwestern Argentina *Parasitology* **132**, 57-65.

191. Westenberger, S. J., Barnabe, C., Campbell, D. A. & Sturm, N. R. (2005) Two Hybridization Events Define the Population Structure of *Trypanosoma cruzi Genetics* **171**, 527-543.

192. Rozas, M., De Doncker, S., Adaui, V., Coronado, X., Barnabe, C., *et al.* (2007) Multilocus polymerase chain reaction restriction fragment--length polymorphism genotyping of *Trypanosoma cruzi* (Chagas disease): taxonomic and clinical applications *J Infect Dis* **195**, 1381-8.

193. Gonzalez, A., Lerner, T. J., Huecas, M., Sosa-Pineda, B., Nogueira, N., *et al.* (1985) Apparent generation of a segmented mRNA from two separate tandem gene families in *Trypanosoma cruzi Nucleic Acids Res* **13**, 5789-804.

194. Sturm, N. R., Vargas, N. S., Westenberger, S. J., Zingales, B. & Campbell, D. A. (2003) Evidence for multiple hybrid groups in *Trypanosoma cruzi Int J Parastol* **33**, 269-279.

195. Page, R. D. H., Holmes E.C. (1998) *Molecular Evolution: A phylogenetic approach* (Blackwell Science, Oxford).

196. Briones, M. R. S., Souto, R. P., Stolf, B. S. & Zingales, B. (1999) The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity *Molecular and Biochemical Parasitology* **104**, 219-232.

197. Tibayrenc, M. (1995) Population genetics of parasitic protozoa and other microorganisms. *Advances in Parasitology* **36**.

198. Kawashita, S. Y., Sanson, G. F. O., Fernandes, O., Zingales, B. & Briones, M. R. S. (2001) Maximum-Likelihood Divergence Date Estimates Based on rRNA Gene Sequences Suggest Two Scenarios of *Trypanosoma cruzi* Intraspecific Evolution *Molecular Biology and Evolution* **18**, 2250-2259.

199. Machado, C. A. & Ayala, F. J. (2001) Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi* *Proc Natl Acad Sci U S A* **98**, 7396-401.

200. Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., *et al.* (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms *Proc Natl Acad Sci U S A* **95**, 3140-5.

201. Currie, B. J., Thomas, A. D., Godoy, D., Dance, D. A., Cheng, A. C., *et al.* (2007) Australian and Thai isolates of *Burkholderia pseudomallei* are distinct by multilocus sequence typing; revision of a case of mistaken identity *J Clin Microbiol* **45**, 3828-3829.

202. Hanage, W. P., Kaijalainen, T., Herva, E., Saukkoriipi, A., Syrjanen, R., *et al.* (2005) Using multilocus sequence data to define the pneumococcus *J Bacteriol* **187**, 6223-30.

203. Robles, J. C., Koreen, L., Park, S. & Perlin, D. S. (2004) Multilocus Sequence Typing Is a Reliable Alternative Method to DNA Fingerprinting for Discriminating among Strains of *Candida albicans* *J. Clin. Microbiol.* **42**, 2480-2488.

204.   Chen, K.-W., Chen, Y.-C., Lo, H.-J., Odds, F. C., Wang, T.-H., *et al.* (2006) Multilocus Sequence Typing for Analyses of Clonality of Candida albicans Strains in Taiwan *J. Clin. Microbiol.* **44,** 2172-2178.

205.   Stephens, M. & Donnelly, P. (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data *Am J Hum Genet* **73,** 1162-9.

206.   Hedrick, P. W. (2005) *Genetics of populations* (Jones and Bartlet Publishers, Sudbury, Massachussets).

207.   Webster, M. T., Smith, N. G. & Ellegren, H. (2002) Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments *Proc Natl Acad Sci U S A* **99,** 8748-53.

208.   Einsen, J. A. (1999) in *Microsatellites: evolution and applications,* eds. Goldtsein, D. B. & Schlotterer, C. (Oxford University Press Oxford)

209.   Kimura, M. & Crow, J. F. (1964) The number of alleles that can be maintained In a finite population *Genetics* **49,** 725-738.

210.   Ohta, T. & Kimura, M. (1973) The model of mutation appropriate to estimate the number of electrophoretically detectable alleles In a genetic population. *Genetics Research* **22,** 201-204.

211.   Goldstein, D. B. & Pollock, D. D. (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic interference *J Hered* **88,** 335-42.

212.   Goldstein, D. B., Ruiz Linares, A., Cavalli-Sforza, L. L. & Feldman, M. W. (1995) Genetic absolute dating based on microsatellites and the origin of modern humans *Proc Natl Acad Sci U S A* **92,** 6723-7.

213.   Takezaki, N. & Nei, M. (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA *Genetics* **144,** 389-99.

214.   Fisher, M. C., Koenig, G. L., White, T. J., San-Blas, G., Negroni, R., *et al.* (2001) Biogeographic range expansion into South America by *Coccidioides*

immitis mirrors New World patterns of human migration *Proc Natl Acad Sci U S A* **98**, 4558-62.

215. Oliveira, R. P., Broude, N. E., Macedo, A. M., Cantor, C. R., Smith, C. L., *et al.* (1998) Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites *Proc Natl Acad Sci U S A* **95**, 3776-80.

216. de Freitas, J. M., Augusto-Pinto, L., Pimenta, J. R., Bastos-Rodrigues, L., Goncalves, V. F., *et al.* (2006) Ancestral Genomes, Sex, and the Population Structure of *Trypanosoma cruzi PLoS Pathogens* **2**, e24.

217. Oliveira, R. P., Melo, A. I., Macedo, A. M., Chiari, E. & Pena, S. D. (1999) The population structure of *Trypanosoma cruzi*: expanded analysis of 54 strains using eight polymorphic CA-repeat microsatellites *Mem Inst Oswaldo Cruz* **94**, 65-70.

218. Aufderheide, A. C., Salo, W., Madden, M., Streitz, J., Buikstra, J., *et al.* (2004) A 9,000-year record of Chagas' disease *Proc Natl Acad Sci U S A* **101**, 2034-9.

219. Nicholls, H. (2005) Ancient DNA comes of age. *PLoS Biology* **3**, e56.

220. Iturralde-Vinent, M. A. & MacPhee, R. D. E. (1996) Age and Paleogeographic origin of Dominican amber. *Science* **273**, 1850–1852.

221. Hamilton, P. B., Gibson, W. C. & Stevens, J. R. (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.

222. Cortez, M. R., Pinho, A. P., Cuervo, P., Alfaro, F., Solano, M., *et al.* (2006) *Trypanosoma cruzi* (Kinetoplastida Trypanosomatidae): ecology of the transmission cycle in the wild environment of the Andean valley of Cochabamba, Bolivia *Exp Parasitol* **114**, 305-13.

223. Herrera, L., D'Andrea, P. S., Xavier, S. C., Mangia, R. H., Fernandes, O., *et al.* (2005) *Trypanosoma cruzi* infection in wild mammals of the National Park 'Serra da Capivara' and its surroundings (Piaui, Brazil), an area endemic for Chagas disease *Trans R Soc Trop Med Hyg* **99**, 379-88.

224. Povoa, M. M., de Souza, A. A., Naiff, R. D., Arias, J. R., Naiff, M. F., *et al.* (1984) Chagas' disease in the Amazon basin IV. Host records of *Trypanosoma cruzi* zymodemes in the states of Amazonas and Rondonia, Brazil *Ann Trop Med Parasitol* **78**, 479-87.

225. Lisboa, C. V., Dietz, J., Baker, A. J., Russel, N. N. & Jansen, A. M. (2000) *Trypanosoma cruzi* infection in *Leontopithecus rosalia* at the Reserva Biologica de Poco das Antas, Rio de Janeiro, Brazil *Mem Inst Oswaldo Cruz* **95**, 445-52.

226. Lisboa, C. V., Mangia, R. H., Rubiao, E., de Lima, N. R., das Chagas Xavier, S. C., *et al.* (2004) *Trypanosoma cruzi* transmission in a captive primate unit, Rio de Janeiro, Brazil *Acta Trop* **90**, 97-106.

227. Solari, A., Munoz, S., Venegas, J., Wallace, A., Aguilera, X., *et al.* (1992) Characterization of Chilean, Bolivian, and Argentinian *Trypanosoma cruzi* populations by restriction endonuclease and isoenzyme analysis *Exp Parasitol* **75**, 187-95.

228. Tibayrenc, M. & Ayala, F. J. (1987) *Trypanosoma cruzi* populations: more clonal than sexual *Parasitol Today* **3**, 189-90.

229. Tibayrenc, M., Kjellberg, F. & Ayala, F. (1990) A Clonal Theory of Parasitic Protozoa: The Population Structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas*, and *Trypanosoma* and their Medical and Taxonomical Consequences *Proc Natl Acad Sci USA* **87**, 2414-2418.

230. Tibayrenc, M. & Ayala, F. J. (2002) The clonal theory of parasitic protozoa: 12 years on *Trends Parasitol* **18**, 405-10.

231. Telleria, J., Barnabe, C., Hide, M., Banuls, A.-L. & Tibayrenc, M. (2004) Predominant clonal evolution leads to a close parity between gene expression profiles and subspecific phylogeny in *Trypanosoma cruzi Mol Biochem Parasitol* **137**, 133-141.

232. Lobo, C. A. & Kumar, N. (1998) Sexual differentiation and development in the malaria parasite *Parasitol Today* **14**, 146-50.

233. Ferguson, D. J. P. (2002) *Toxoplasma gondii* and sex: essential or optional extra? *Trends Parasitol* **18**, 351-355.

234. Tait, A., MacLeod, A., Tweedie, A., Masiga, D. & Turner, C. M. R. (2007) Genetic exchange in *Trypanosoma brucei*: Evidence for mating prior to metacyclic stage development *Mol Biochem Parasitol* **151**, 133-136.

235. MacLeod, A., Tweedie, A., McLellan, S., Taylor, S., Cooper, A., *et al.* (2005) Allelic segregation and independent assortment in *T. brucei* crosses: Proof that the genetic system is Mendelian and involves meiosis *Mol Biochem Parasitol* **143**, 12-19.

236. MacLeod, A., Tweedie, A., Welburn, S. C., Maudlin, I., Turner, C. M., *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**, 13442-7.

237. McDaniel, J. & Dvorak, J. (1993) Identification, isolation, and characterization of naturally-occurring *Trypanosoma cruzi* variants. *Mol Biochem Parasitol* **57**, 213-222.

238. Westenberger, S. J., Sturm, N. R. & Campbell, D. A. (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.

239. Stefani, S. & Agodi, A. (2000) Molecular epidemiology of antibiotic resistance *Int J Antimicrob Agents* **13**, 143-53.

240. Widmer, G., Dvorak, J. A. & Miles, M. A. (1987) Temperature modulation of growth rates and glucosephosphate isomerase isozyme activity in *Trypanosoma cruzi Mol Biochem Parasitol* **23**, 55-62.

241. Lima, V. S., Mangia, R. H., Carreira, J. C., Marchewski, R. S. & Jansen, A. M. (1999) *Trypanosoma cruzi*: correlations of biological aspects of the life cycle in mice and triatomines *Mem Inst Oswaldo Cruz* **94**, 397-402.

325

242. Murta, S. M., Gazzinelli, R. T., Brener, Z. & Romanha, A. J. (1998) Molecular characterization of susceptible and naturally resistant strains of *Trypanosoma cruzi* to benznidazole and nifurtimox *Mol Biochem Parasitol* **93**, 203-14.

243. Teixeira, A. R., Nascimento, R. J. & Sturm, N. R. (2006) Evolution and pathology in chagas disease--a review *Mem Inst Oswaldo Cruz* **101**, 463-91.

244. Vago, A. R., Andrade, L. O., Leite, A. A., d'Avila Reis, D., Macedo, A. M., *et al.* (2000) Genetic Characterization of *Trypanosoma cruzi* Directly from Tissues of Patients with Chronic Chagas Disease : Differential Distribution of Genetic Types into Diverse Organs *Am J Path* **156**, 1805-1809.

245. Anez, N., Carrasco, H., Parada, H., Crisante, G., Rojas, A., *et al.* (1999) Myocardial parasite persistence in chronic chagasic patients *Am J Trop Med Hyg* **60**, 726-32.

246. Tarleton, R. L. & Zhang, L. (1999) Chagas Disease Etiology: Autoimmunity or Parasite Persistence? *Parasitology Today* **15**, 94-99.

247. Buscaglia, C. A. & Di Noia, J. M. (2003) *Trypanosoma cruzi* clonal diversity and the epidemiology of Chagas' disease *Microbes and Infection* **5**, 419-427.

248. Breniere, S. F., Carrasco, R., Revollo, S., Aparicio, G., Desjeux, P., *et al.* (1989) Chagas' disease in Bolivia: clinical and epidemiological features and zymodeme variability of *Trypanosoma cruzi* strains isolated from patients *Am J Trop Med Hyg* **41**, 521-9.

249. Anez, N., Crisante, G., da Silva, F. M., 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**, 1319-26.

250. Miles, M. A., Yeo, M. & Gaunt, M. (2003) in *Molecular mechanisms in the pathogenisis of Trypanosoma cruzi*, ed. Kelly, J. M. (Kluwer Academic / Plenum publishers, New York).

251. Valadares, H. M., Pimenta, J. R., de Freitas, J. M., Duffy, T., Bartholomeu, D. C., *et al.* (2008) Genetic profiling of *Trypanosoma cruzi* directly in infected tissues using nested PCR of polymorphic microsatellites *Int J Parasitol* **38**, 839-50.

252. Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993) How Clonal are Bacteria? *Proc Natl Acad Sci USA* **90**, 4384-4388.

253. Wright, S. (1951) The genetical structure of populations *Ann Eugen* **15**, 323-54.

254. Nei, M. (1977) F-statistics and analysis of gene diversity in subdivided populations *Ann. Hum. Genet. Lond.* **41**, 225-233.

255. Nei, M. (1972) Genetic distance between populations. *American Naturalist* **106**, 283-291.

256. Nei, M., Tajima, F. & Tateno, Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data *J Mol Evol* **19**, 153-70.

257. Stephens, J. C., Gilbert, D. A., Yuhki, N. & O'Brien, S. J. (1992) Estimation of heterozygosity for single-probe multilocus DNA fingerprints *Mol Biol Evol* **9**, 729-43.

258. Heitman, J. (2006) Sexual Reproduction and the Evolution of Microbial Pathogens *Current Biology* **16**, R711-R725.

259. Tibayrenc, M. & Ayala, F. J. (1991) Towards a population genetics of microorganisms: The clonal theory of parasitic protozoa *Parasitology Today* **7**, 228-232.

260. Spratt, B. G., Hanage, W. P. & Feil, E. J. (2001) The relative contributions of recombination and point mutation to the diversification of bacterial clones *Curr Opin Microbiol* **4**, 602-6.

261. MacLeod, A., Tait, A. & Turner, C. M. (2001) The population genetics of *Trypanosoma brucei* and the origin of human infectivity *Philos Trans R Soc Lond B Biol Sci* **356**, 1035-44.

262. Ajzenberg, D., Banuls, A. L., Su, C., Dumetre, A., Demar, M., *et al.* (2004) Genetic diversity, clonality and sexuality in *Toxoplasma gondii Int J Parasitol* **34**, 1185-1196.

263. Lehmann, T., Graham, D. H., Dahl, E. R., Bahia-Oliveira, L. M., Gennari, S. M., *et al.* (2004) Variation in the structure of *Toxoplasma gondii* and the roles of selfing, drift, and epistatic selection in maintaining linkage disequilibria *Infect Genet Evol* **4**, 107-14.

264. Balloux, F., Lehmann, L. & de Meeus, T. (2003) The population genetics of clonal and partially clonal diploids *Genetics* **164**, 1635-44.

265. Mark Welch, D. & Meselson, M. (2000) Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange *Science* **288**, 1211-5.

266. Schwenkenbecher, J. M., Wirth, T., Schnur, L. F., Jaffe, C. L., Schallig, H., *et al.* (2006) Microsatellite analysis reveals genetic structure of Leishmania tropica *Int J Parasitol* **36**, 237-46.

267. Kuhls, K., Keilonat, L., Ochsenreither, S., Schaar, M., Schweynoch, C., *et al.* (2007) Multilocus microsatellite typing (MLMT) reveals genetically isolated populations between and within the main endemic regions of visceral leishmaniasis *Microbes and Infection* **9**, 334-343.

268. Birky, C. W., Jr. (1996) Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes *Genetics* **144**, 427-37.

269. Olson, D., Dinerstein, E., Wikramanayake, E., Burgess, N., Powell, G., *et al.* (2001) Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience* **51**, 993-938.

270. Gascon, C., Malcolm, J. R., Patton, J. L., da Silva, M. N., Bogart, J. P., *et al.* (2000) Riverine barriers and the geographic distribution of Amazonian species *Proc Natl Acad Sci U S A* **97**, 13672-7.

271. Anon, A. (1994) Bolivian hemorrhagic fever--El Beni Department, Bolivia, 1994 *MMWR Morb Mortal Wkly Rep* **43**, 943-6.

272. Miles, M. A., de Souza, A. A. & Povoa, M. (1981) Chagas' disease in the Amazon basin III. Ecotopes of ten triatomine bug species (Hemiptera: Reduviidae) from the vicinity of Belem, Para State, Brazil *J Med Entomol* **18**, 266-78.

273. Miles, M. d. S., AA. Povoa, M. (1981) Mammal trapping and nest location in Brazilian forest with an improved spool and line *J Zool* **195**, 331-347.

274. Abad-Franch, F., Noireau, F., Paucar, A., Aguilar, H. M., Carpio, C., *et al.* (2000) The use of live-bait traps for the study of sylvatic *Rhodnius* populations (Hemiptera: Reduviidae) in palm trees *Trans R Soc Trop Med Hyg* **94**, 629-30.

275. Miles, M. A. (1993) Culturing and biological cloning of *Trypanosoma cruzi Methods Mol Biol* **21**, 15-28.

276. Yeo, M., Lewis, M. D., Carrasco, H. J., Acosta, N., Llewellyn, M., *et al.* (2007) Resolution of multiclonal infections of *Trypanosoma cruzi* from naturally infected triatomine bugs and from experimentally infected mice by direct plating on a sensitive solid medium *Int J Parasitol* **37**, 111-20.

277. Yeo, M. (2003) ITD (University of London, London), PhD Thesis

278. Souto, R. P. & Zingales, B. (1993) Sensitive detection and strain classification of *Trypanosoma cruzi* by amplification of a ribosomal RNA sequence *Mol Biochem Parasitol* **62**, 45-52.

279. Grisard, E. C., Campbell, D. A. & Romanha, A. J. (1999) Mini-exon gene sequence polymorphism among *Trypanosoma rangeli* strains isolated from distinct geographical regions *Parasitology* **118**, 375-82.

280. Wu, L., Tang, T., Zhou, R. & Shi, S. (2007) PCR-mediated recombination of the amplification products of the *Hibiscus tiliaceus* cytosolic glyceraldehyde-3-phosphate dehydrogenase gene *J Biochem Mol Biol* **40**, 172-9.

281. Stephens, M., Smith, N. & Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**, 978 - 989.

282. Felsenstein, J. (2004) PHYLIP (Phylogeny Inference Package) v.3.6 (Distributed by the author) Department of Genome Sciences, University of Washington, Seattle.

283. Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood *Mol Biol Evol* **24**, 1586-91.

284. Martin, D., Williamson, C. & Posada, D. (2005).RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260-262.

285. Maynard Smith, J. (1992) Analyzing the mosaic structure of genes. *J Mol Evol* **34**, 126-129.

286. Posada, D. & Crandall, K. (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc Natl Acad Sci USA* **98**, 13757-13762.

287. Martin, D., Posada, D., Crandall, K. & Williamson, C. (2005) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints *AIDS Res Hum Retroviruses* **21**, 98-102.

288. Fisher, M. C., Aanensen, D., de Hoog, S. & Vanittanakom, N. (2004) Multilocus microsatellite typing system for *Penicillium marneffei* reveals spatially structured populations *J Clin Microbiol* **42**, 5065-9.

289. Excoffier, L. & Heckel, G. (2006) Computer programs for population genetics data analysis: a survival guide *Nat Rev Genet* **7**, 745-58.

290. Minch, E., Ruíz-Linares, A., Goldstein, D., Feldman, M. & Cavalli-Sforza, L. (1995) MICROSAT - the microsatellite distance program (Stanford University Press, Stanford).

291. Mantel, N. (1967) The detection of disease clustering and a generalized regression approach *Cancer Res* **27**, 209-20.

292. Peakall, R. & Smouse, P. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

293. Jenness, J. (2005) Distance matrix extension for ArcView 3.x, v.2 (Jenness Enterprises)

294. Park, S. D. E. (2001) (University of Dublin, Dublin), Ph.D. thesis.

295. Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

296. Nei, M. (1973) Analysis of gene diversity in subdivided populations *Proc Natl Acad Sci U S A* **70**, 3321-3.

297. Raymond, M. & Rousset, F. (1995) Genepop (Version-1.2) - Population-Genetics Software for Exact Tests and Ecumenicism *J Hered* **86**, 248-249.

298. Rice, W. (1989) Analyzing tables with statistical tests *Evolution* **43**, 223-225.

299. 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**, 163-70.

300. Agapow, P. M. & Burt, A. (2001) Indices of multilocus linkage disequilibrium *Mol Ecol Notes* **1**, 101-102.

301. Porter, H. A. (1990) Testing minimal species boundaries using gene flow statistics: the taxonomy of two hybridizing admiral butterflies (Limenitis: Nymphalidae). *Syst. Zool* **39**, 131-147.

302. Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies *Genetics* **139**, 457-62.

303. Raybould, A., Mogg, R., Aldam, C., Gliddon, C., Thorpe, R., *et al.* (1998) The genetic structure of sea beet (*Beta vulgaris* ssp. *maritima*) populations. III. Detection of isolation by distance at microsatellite loci. *Heredity* **80**, 127-132.

304. Paetkau, D., Calvert, W., Stirling, I. & Strobeck, C. (1995) Microsatellite analysis of population structure in Canadian polar bears *Mol Ecol* **4**, 347-54.

305. Paetkau, D., Slade, R., Burden, M. & Estoup, A. (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power *Mol Ecol* **13**, 55-65.

306. Pritchard, J. K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data *Genetics* **155**, 945-59.

307. Ochsenreither, S., Kuhls, K., Schaar, M., Presber, W. & Schonian, G. (2006) Multilocus microsatellite typing as a new tool for discrimination of Leishmania infantum MON-1 strains *J Clin Microbiol* **44**, 495-503.

308. Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data *Genetics* **131**, 479-91.

309. Houlden, B. A., England, P. R., Taylor, A. C., Greville, W. D. & Sherwin, W. B. (1996) Low genetic variability of the koala *Phascolarctos cinereus* in south-eastern Australia following a severe population bottleneck *Mol Ecol* **5**, 269-81.

310. Weber, D. S., Stewart, B. S., Garza, J. C. & Lehman, N. (2000) An empirical genetic assessment of the severity of the northern elephant seal population bottleneck *Curr Biol* **10**, 1287-90.

311. Rich, S. M., Licht, M. C., Hudson, R. R. & Ayala, F. J. (1998) Malaria's Eve: evidence of a recent population bottleneck throughout the world populations of Plasmodium falciparum *Proc Natl Acad Sci U S A* **95**, 4425-30.

312. Reed, D. L., Smith, V. S., Hammond, S. L., Rogers, A. R. & Clayton, D. H. (2004) Genetic analysis of lice supports direct contact between modern and archaic humans *PLoS Biol* **2**, e340.

313. Reiland, J., Hodge, S. & Noor, M. A. (2002) Strong founder effect in *Drosophila pseudoobscura* colonizing New Zealand from North America *J Hered* **93**, 415-20.

314.   Cornuet, J. M. & Luikart, G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data *Genetics* **144**, 2001-14.

315.   Luikart, G., Allendorf, F. W., Cornuet, J. M. & Sherwin, W. B. (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks *J Hered* **89**, 238-47.

316.   Garza, J. C. & Williamson, E. G. (2001) Detection of reduction in population size using data from microsatellite loci *Mol Ecol* **10**, 305-18.

317.   Spencer, C. C., Neigel, J. E. & Leberg, P. L. (2000) Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks *Mol Ecol* **9**, 1517-28.

318.   Nei, M., Chakraborty, R. & Fuerst, P. A. (1976) Infinite allele model with varying mutation rate *Proc Natl Acad Sci U S A* **73**, 4164-8.

319.   Lukes, J., Mauricio, I. L., Schonian, G., Dujardin, J.-C., Soteriadou, K., *et al.* (2007) Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy *Proc Natl Acad Sci U S A* **104**, 9375-9380.

320.   Fundyga, R. E., Lott, T. J. & Arnold, J. (2002) Population structure of *Candida albicans*, a member of the human flora, as determined by microsatellite loci *Infect Genet Evol* **2**, 57-68.

321.   Marquez, E., Arcos-Burgos, M., Triana, O., Moreno, J. & Jaramillo, N. (1998) Clonal population structure of Colombian sylvatic *Trypanosoma cruzi J Parasitol* **84**, 1143-9.

322.   Imwong, M., Nair, S., Pukrittayakamee, S., Sudimack, D., Williams, J. T., *et al.* (2007) Contrasting genetic structure in *Plasmodium vivax* populations from Asia and South America *Int J Parasitol* **37**, 1013-22.

323. Jacobsen, M. D., Gow, N. A., Maiden, M. C., Shaw, D. J. & Odds, F. C. (2007) Strain typing and determination of population structure of *Candida krusei* by multilocus sequence typing *J Clin Microbiol* **45**, 317-23.

324. http://www.ine.gov.bo/ (2001) Censo nacional.

325. Carmen, M. L. (1995) South-south cooperation programme on environmentally sound socio-economic development in the humid tropics. The Beni Biosphere Reserve, Bolivia.

326. Smith, D. N. & Killeen, T. J. (1995) A comparison of the structure and composition of montane and lowland tropical forest in the Serranía Pilón Lajas, Beni, Bolivia. *Missourri Botanical Garden Poject Reports*.

327. Anon, A. (2007) *Bolivia-floods*. Bureau for Democracy, Conflict, and Humanitarian Assistance (DCHA), USA Government. Fact Sheet 4.

328. Harris, A. F., Matias-Arnez, A. & Hill, N. (2006) Biting time of Anopheles darlingi in the Bolivian Amazon and implications for control of malaria *Trans R Soc Trop Med Hyg* **100**, 45-7.

329. Montes de Oca, I. (1997) *Geografía y recursos naturales de Bolivia* (La Paz, Bolivia)

330. Shean, M. J. (2005) Commodity Intelligence Report, Bolivia (United States Department of Argiculture)

331. British petroleum (2007) Statistcal review of world energy. (www.bp.com)

332. Arias, O. & Robles, M. (2007) in *More than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions.*, ed. Simler, K. (World Bank)

333. Breniere, S. F., Morochi, W., Bosseno, M. F., Ordonez, J., Gutierrez, T., *et al.* (1998) *Trypanosoma cruzi* genotypes associated with domestic Triatoma sordida in Bolivia *Acta Trop* **71**, 269-83.

334. www.fao.org/emergencies/country_information/list/latinamerica/bolivia/en/

335. Dujardin, J. P., Schofield, C. J. & Tibayrenc, M. (1998) Population structure of Andean Triatoma infestans: allozyme frequencies and their epidemiological relevance *Med Vet Entomol* **12**, 20-9.

336. Dujardin, J. P., Forgues, G., Torrez, M., Martinez, E., Cordoba, C., *et al.* (1998) Morphometrics of domestic Panstrongylus rufotuberculatus in Bolivia *Ann Trop Med Parasitol* **92**, 219-28.

337. http://www.barinas.net.ve/infopedia/

338. http://www.inti.gob.ve/index.php?option=com_content&task=view&id=146&Itemid=2

339. Maia Da Silva, F., Junqueira, A. C., Campaner, M., Rodrigues, A. C., Crisante, G., *et al.* (2007) Comparative phylogeography of *Trypanosoma rangeli* and Rhodnius (Hemiptera: Reduviidae) supports a long coexistence of parasite lineages and their sympatric vectors *Mol Ecol* **16**, 3361-73.

340. 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**, 73-9.

341. Lainson, R., Shaw, J. J., Fraiha, H., Miles, M. A. & Draper, C. C. (1979) Chagas's Disease in the Amazon Basin: 1. *Trypanosoma cruzi* infections in silvatic mammals, triatomine bugs and man in the State of Para, north Brazil *Trans R Soc Trop Med Hyg* **73**, 193-204.

342. Solis-Franco, R. R., Romo-Zapata, J. A. & Martinez-Ibarra, J. A. (1997) Wild reservoirs infected by *Trypanosoma cruzi* in the ecological park "El Zapotal", Tuxtla Gutierrez, Chiapas, Mexico *Mem Inst Oswaldo Cruz* **92**, 163-4.

343. Mondolfi, E. P. H., R. (1984) Una nueva subsepecie de Zarigueya de grupo *Didelphis marsupialis Acta Cient Venez* **35**, 407-13.

344.   Patton, J. L. D. S., M.N.F. Malcolm J.R. (2000) Mammals of the Rio Jurua and the evolutionary and ecological diversification of Amazonia *Bull Am Mus Nat Hist* **244**, 1-306.

345.   Olifiers, N., Gentile, R. & Fiszon, J. T. (2005) Relation between small-mammal species composition and anthropic variables in the Brazilian Atlantic Forest. *Braz. J. Biol* **65**, 496-501.

346.   Sunquist, M., Austad, S. & Sunquist, F. (1987) Movement Patterns and Home Range in the Common Opossum (*Didelphis marsupialis*) *J Mammol* **68**, 173-176.

347.   Frutos, S. & van Den Bussche, R. (2002) Genetic Diversity and Gene Flow in Nine-Banded Armadillos in Paraguay *J Mammol* **83**, 815-823.

348.   Huchon, D., Delsuc, F., Catzeflis, F. M. & Douzery, E. J. (1999) Armadillos exhibit less genetic polymorphism in North America than in South America: nuclear and mitochondrial data confirm a founder effect in *Dasypus novemcinctus* (Xenarthra) *Mol Ecol* **8**, 1743-8.

349.   Miles, M., Gaunt, M. & Yeo, M. (2002) in *American trypanosomiasis*, eds. Miles, M. & Tyler, K. (Kluwer publishing)

350.   Legey, A. P., Pinho, A. P., Xavier, S. C., Marchevsky, R., Carreira, J. C., *et al.* (2003) *Trypanosoma cruzi* in marsupial didelphids (*Philander frenata* and *Didelhis marsupialis*): differences in the humoral immune response in natural and experimental infections *Rev Soc Bras Med Trop* **36**, 241-8.

351.   Carreira, J. C., Jansen, A. M., de Nazareth Meirelles, M., Costa e Silva, F. & Lenzi, H. L. (2001) *Trypanosoma cruzi* in the scent glands of *Didelphis marsupialis*: the kinetics of colonization *Exp Parasitol* **97**, 129-40.

352.   Bermudez, H., Balderrama, F. & Torrico, F. (1993) Identification and characterization of wild foci of *Triatoma infestans* in Central Bolivia. *Am J Trop Med Hyg* **49**, 371.

353. Dujardin, J. P., Tibayrenc, M., Venegas, E., Maldonado, L., Desjeux, P., *et al.* (1987) Isozyme evidence of lack of speciation between wild and domestic *Triatoma infestans* (Heteroptera: Reduviidae) in Bolivia *J Med Entomol* **24**, 40-5.

354. Pinho, A., Cuervo, P., Gripp, E., Costa, B., Herrera, H., *et al.* (2003) Isoenzymatic characterization and genetic diversity of *Trypanosoma cruzi* isolates of wild animals from Pantanal, Brazil. *Revista do Instituto de Medicina Tropical de Sao Paulo* **45**, 12.

355. Miles, M. A., Arias, J. R., Valente, S. A., Naiff, R. D., de Souza, A. A., *et al.* (1983) Vertebrate hosts and vectors of *Trypanosoma rangeli* in the Amazon Basin of Brazil *Am J Trop Med Hyg* **32**, 1251-9.

356. Dereure, J., Barnabe, C., Vie, J. C., Madelenat, F. & Raccurt, C. (2001) Trypanosomatidae from wild mammals in the neotropical rainforest of French Guiana *Ann Trop Med Parasitol* **95**, 157-66.

357. Bento, D. N., Farias, L. M., Godoy, M. F. & Araujo, J. F. (1992) [The epidemiology of Chagas' disease in a rural area of the city of Teresina, Piaui, Brazil] *Rev Soc Bras Med Trop* **25**, 51-8.

358. De Stefani Marquez, D., Rodrigues-Ottaiano, C., Monica Oliveira, R., Pedrosa, A. L., Cabrine-Santos, M., *et al.* (2006) Susceptibility of different triatomine species to Trypanosoma rangeli experimental infection *Vector Borne Zoonotic Dis* **6**, 50-6.

359. Matias, A., de la Rive, J. X., Torrez, M. & Dujardin, J. P. (2001) Rhodnius robustus in Bolivia identified by its wings *Mem Inst Oswaldo Cruz* **96**, 947-50.

360. Barretto, M. P. & Albuquerque, R. D. (1969) [Studies on wild reservoirs and vectors of *Trypanosoma cruzi*. 33. Experimental and natural infection of Psammolestes testius Lent and Jurberg, 1965 by T. cruzi] *Rev Inst Med Trop Sao Paulo* **11**, 165-8.

361. Tibayrenc, M. & Le Pont, F. (1984) Étude isoenzymatique d'isolats boliviens de *Trypanosoma cruzi* pratiques chez *Rhodnius pictipes*. Données préliminties sur

la transmissions de la maladie de Chagas dans l'Alto Beni bolivien. *Cah. O.R.S.T.O.M., sér. Ent. méd. et Parasitol.* **22,** 55-57.

362. Miles, M. (1979) *Transmission cycles and the heterogeneity of Trypanosoma cruzi* (Academic Press)

363. Stothard, J. R., Frame, I. A., Carrasco, H. J. & Miles, M. A. (1998) On the molecular taxonomy of *Trypanosoma cruzi* using riboprinting *Parasitology* **117,** 243-7.

364. Feliciangeli, M. D., Sanchez-Martin, M., Marrero, R., Davies, C. & Dujardin, J. P. (2007) Morphometric evidence for a possible role of *Rhodnius prolixus* from palm trees in house re-infestation in the State of Barinas (Venezuela) *Acta Trop* **101,** 169-77.

365. Miles, M. A., Arias, J. R. & de Souza, A. A. (1983) Chagas' disease in the Amazon basin: V. Periurban palms as habitats of *Rhodnius robustus* and Rhodnius pictipes--triatomine vectors of Chagas' disease *Mem Inst Oswaldo Cruz* **78,** 391-8.

366. Pifano, F. (1969) *Algunos aspectos en la ecología y epidemiología de las enfermedades endémicas con focos naturales en el área tropical, especialmente en Venezuela.* (Ministerio de Sanidad & Asistencia Social Ed, Caracas.).

367. Omah-Maharaj, I. (1992) Studies on vectors of *Trypanosoma cruzi* in Trinidad, West Indies *Med Vet Entomol* **6,** 115-20.

368. Feliciangeli, M. D., Dujardin, J. P., Bastrenta, B., Mazzarri, M., Villegas, J., *et al.* (2002) Is Rhodnius robustus (Hemiptera: Reduviidae) responsible for Chagas disease transmission in Western Venezuela? *Trop Med Int Health* **7,** 280-7.

369. Robello, C., Gamarro, F., Castanys, S. & Alvarez-Valin, F. (2000) Evolutionary relationships in *Trypanosoma cruzi*: molecular phylogenetics supports the existence of a new major lineage of strains *Gene* **246,** 331-338.

370. Steindel, M., Toma, H. K., Ishida, M. M., Murta, S. M., de Carvalho Pinto, C. J., *et al.* (1995) Biological and isoenzymatic characterization of *Trypanosoma cruzi*

strains isolated from sylvatic reservoirs and vectors from the state of Santa Catarina, Southern Brazil *Acta Trop* **60**, 167-77.

371. Travi, B. L., Jaramillo, C., Montoya, J., Segura, I., Zea, A., *et al.* (1994) Didelphis marsupialis, an important reservoir of *Trypanosoma* (*Schizotrypanum*) *cruzi* and *Leishmania* (Leishmania) *chagasi* in Colombia *Am J Trop Med Hyg* **50**, 557-65.

372. Rozas, M., Botto-Mahan, C., Coronado, X., Ortiz, S., Cattan, P. E., *et al.* (2007) Coexistence of *Trypanosoma cruzi* genotypes in wild and periodomestic mammals in Chile *Am J Trop Med Hyg* **77**, 647-53.

373. Bargues, M. D., Marcilla, A., Ramsey, J. M., Dujardin, J. P., Schofield, C. J., *et al.* (2000) Nuclear rDNA-based molecular clock of the evolution of triatominae (Hemiptera: reduviidae), vectors of Chagas disease *Mem Inst Oswaldo Cruz* **95**, 567-73.

374. Willis, K. J. & Whittaker, R. J. (2000) Perspectives: paleoecology. The refugial debate *Science* **287**, 1406-7.

375. Lewis, M. (2008) ITD (London School of Hygiene and Tropical Medicine, London), PhD thesis

376. www.genedb.org.

377. Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599.

378. Nei, M. & Li, W. H. (1976) The transient distribution of allele frequencies under mutation pressure *Genet Res* **28**, 205-14.

379. Broutin, H., Tarrieu, F., Tibayrenc, M., Oury, B. & Barnabe, C. (2006) Phylogenetic analysis of the glucose-6-phosphate isomerase gene in *Trypanosoma cruzi Experimental Parasitology* **113**, 1-7.

380. Macedo, A. M., Pimenta, J. R., Aguiar, R. S., Melo, A. I., Chiari, E., *et al.* (2001) Usefulness of microsatellite typing in population genetic studies of *Trypanosoma cruzi Mem Inst Oswaldo Cruz* **96**, 407-13.

381. Mauricio, I. L., Yeo, M., Baghaei, M., Doto, D., Pratlong, F., *et al.* (2006) Towards multilocus sequence typing of the *Leishmania donovani* complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD) *Int J Parasitol* **36**, 757-69.

382. Grigg, M. E., Bonnefoy, S., Hehl, A. B., Suzuki, Y. & Boothroyd, J. C. (2001) Success and Virulence in Toxoplasma as the Result of Sexual Recombination Between Two Distinct Ancestries *Science* **294**, 161-165.

383. Grynberg, P., Fontes, C. J., Hughes, A. L. & Braga, E. M. (2008) Polymorphism at the Apical Membrane Antigen-1 locus Reflects the World Population History of *Plasmodium vivax BMC Evol Biol* **8**, 123.

384. Herrera, C., Bargues, M. D., Fajardo, A., Montilla, M., Triana, O., *et al.* (2007) Identifying four *Trypanosoma cruzi* I isolate haplotypes from different geographic regions in Colombia *Infection, Genetics and Evolution* **7**, 535-539.

385. Salazar, A., Schijman, A. G. & Triana-Chavez, O. (2006) High variability of Colombian *Trypanosoma cruzi* lineage I stocks as revealed by low-stringency single primer-PCR minicircle signatures *Acta Tropica* **100**, 110-118.

386. Lewicka, K., Breniere-Campana, S. F., Barnabe, C., Dedet, J. P. & Tibayrenc, M. (1995) An isoenzyme survey of *Trypanosoma cruzi* genetic variability in sylvatic cycles from French Guiana *Exp Parasitol* **81**, 20-8.

387. Higo, H., Miura, S., Horio, M., Mimori, T., Hamano, S., *et al.* (2004) Genotypic variation among lineages of *Trypanosoma cruzi* and its geographic aspects *Parasitology International* **53**, 337-344.

388. Morgan, J. A., Vredenburg, V. T., Rachowicz, L. J., Knapp, R. A., Stice, M. J., *et al.* (2007) Population genetics of the frog-killing fungus *Batrachochytrium dendrobatidis Proc Natl Acad Sci U S A* **104**, 13845-50.

389. Machado, R. L., Povoa, M. M., Calvosa, V. S., Ferreira, M. U., Rossit, A. R., *et al.* (2004) Genetic structure of Plasmodium falciparum populations in the Brazilian Amazon region *J Infect Dis* **190**, 1547-55.

390. Thiele, E. A., Sorensen, R. E., Gazzinelli, A. & Minchella, D. J. (2008) Genetic diversity and population structuring of Schistosoma mansoni in a Brazilian village *Int J Parasitol* **38**, 389-99.

391. da Silva, M. N. & Patton, J. L. (1998) Molecular phylogeography and the evolution and conservation of Amazonian mammals *Mol Ecol* **7**, 475-86.

392. Fitzpatrick, S., Feliciangeli, M. D., Sanchez-Martin, M. J., Monteiro, F. A. & Miles, M. A. (2008) Molecular Genetics Reveal That Silvatic *Rhodnius prolixus* Do Colonise Rural Houses *PLoS Negl Trop Dis* **2**, e210.

393. Herber, O. & Kroeger, A. (2003) Pyrethroid-impregnated curtains for Chagas' disease control in Venezuela *Acta Trop* **88**, 33-8.

394. Miles, M. A., Apt, B. W., Widmer, G., Povoa, M. M. & Schofield, C. J. (1984) Isozyme heterogeneity and numerical taxonomy of *Trypanosoma cruzi* stocks from Chile *Trans R Soc Trop Med Hyg* **78**, 526-35.

395. Ruiz-Sanchez, R., Leon, M. P., Matta, V., Reyes, P. A., Lopez, R., *et al.* (2005) *Trypanosoma cruzi* isolates from Mexican and Guatemalan acute and chronic chagasic cardiopathy patients belong to *Trypanosoma cruzi* I *Mem Inst Oswaldo Cruz* **100**, 281-3.

396. Diosque, P., Barnabe, C., Padilla, A. M., Marco, J. D., Cardozo, R. M., *et al.* (2003) Multilocus enzyme electrophoresis analysis of *Trypanosoma cruzi* isolates from a geographically restricted endemic area for Chagas' disease in Argentina *Int J Parasitol* **33**, 997-1003.

397. Diosque, P., Padilla, A. M., Cimino, R. O., Cardozo, R. M., Negrette, O. S., *et al.* (2004) Chagas disease in rural areas of Chaco Province, Argentina: epidemiologic survey in humans, reservoirs, and vectors *Am J Trop Med Hyg* **71**, 590-3.

398. Macedo, A. M., Machado, C. R., Oliveira, R. P. & Pena, S. D. (2004) *Trypanosoma cruzi*: genetic structure of populations and relevance of genetic variability to the pathogenesis of chagas disease *Memórias do Instituto Oswaldo Cruz* **99** 1-12

399. Ruiz-Garcia, M., Montilla, M., Nicholls, S. O., Angarita, L. & Alvarez, D. (2000) Genetic relationships and spatial genetic structure among clonal stocks of *Trypanosoma cruzi* in Colombia *Heredity* **85 Pt 4**, 318-27.

400. Breniere, S. F., Braquemond, P., Solari, A., Agnese, J. F. & Tibayrenc, M. (1991) An isoenzyme study of naturally occurring clones of *Trypanosoma cruzi* isolated from both sides of the West Andes highland *Trans R Soc Trop Med Hyg* **85**, 62-6.

401. Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., *et al.* (2008) Worldwide human relationships inferred from genome-wide patterns of variation *Science* **319**, 1100-4.

402. Tibayrenc, M., Ward, P., Moya, A. & Ayala, F. J. (1986) Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure *Proc Natl Acad Sci U S A* **83**, 115-9.

403. Coronado, X., Zulantay, I., Albrecht, H., Rozas, M., Apt, W., *et al.* (2006) Variation in *Trypanosoma cruzi* clonal composition detected in blood patients and xenodiagnosis triatomines: implications in the molecular epidemiology of Chile *Am J Trop Med Hyg* **74**, 1008-12.

404. Burgos, J. M., Begher, S., Silva, H. M., Bisio, M., Duffy, T., *et al.* (2008) Molecular identification of *Trypanosoma cruzi* I tropism for central nervous system in Chagas reactivation due to AIDS *Am J Trop Med Hyg* **78**, 294-7.

405. Taylor, J. E. & Rudenko, G. (2006) Switching trypanosome coats: what's in the wardrobe? *Trends Genet* **22**, 614-20.

406. Pearce, R. J., Drakeley, C., Chandramohan, D., Mosha, F. & Roper, C. (2003) Molecular determination of point mutation haplotypes in the dihydrofolate

reductase and dihydropteroate synthase of *Plasmodium falciparum* in three districts of northern Tanzania *Antimicrob Agents Chemother* **47**, 1347-54.

407. Mota, J., Chacon, J. C., Gutierrez-Cabrera, A. E., Sanchez-Cordero, V., Wirtz, R. A., *et al.* (2007) Identification of blood meal source and infection with *Trypanosoma cruzi* of Chagas disease vectors using a multiplex cytochrome b polymerase chain reaction assay *Vector Borne Zoonotic Dis* **7**, 617-27.

408. Nix, N., Hernandez, B., Mendoza, C. & Kelin, R. (1995) Knowledge, attitudes and practices (KAP) survey for Chagas disease in an endemic area of Guatemala. *Am J Trop Med Hyg* **53**, S187.

409. Bogreau, H., Renaud, F., Bouchiba, H., Durand, P., Assi, S. B., *et al.* (2006) Genetic diversity and structure of African *Plasmodium falciparum* populations in urban and rural areas *Am J Trop Med Hyg* **74**, 953-9.

410. Yoshida, N. (2008) *Trypanosoma cruzi* infection by oral route: how the interplay between parasite and host components modulates infectivity *Parasitol Int* **57**, 105-9.

411. Martins, E. & Bonato, V. (2004) On the diet of *Gracilianus microtarsus* (Marsupiala, Didelphidae) in an Atlantic Rainforest fragment in southeastern Brazil *Mammalian Biology* **1**, 58-60.

412. Wilkinson, S. R., Taylor, M. C., Touitha, S., Mauricio, I. L., Meyer, D. J., *et al.* (2002) TcGPXII, a glutathione-dependent *Trypanosoma cruzi* peroxidase with substrate specificity restricted to fatty acid and phospholipid hydroperoxides, is localized to the endoplasmic reticulum *Biochem J* **364**, 787-94.

413. Machado, C. A. & Ayala, F. J. (2002) Sequence variation in the dihydrofolate reductase-thymidylate synthase (DHFR-TS) and trypanothione reductase (TR) genes of *Trypanosoma cruzi Mol Biochem Parasitol* **121**, 33-47.

414. Andersson, B., Aslund, L., Tammi, M., Tran, A. N., Hoheisel, J. D., *et al.* (1998) Complete sequence of a 93.4-kb contig from chromosome 3 of *Trypanosoma cruzi* containing a strand-switch region *Genome Res* **8**, 809-16.

415.   Banks, M., Eichert, W. & Olsen, J. (2003) Which genetic loci have greater population assignment power? *Bioinformatics Application Note* **19**, 1436-1438.

Publications arising

ELSEVIER

Origins of Chagas disease: *Didelphis* species are natural hosts of *Trypanosoma cruzi* I and armadillos hosts of *Trypanosoma cruzi* II, including hybrids

Matthew Yeo[a,*], Nidia Acosta[b], Martin Llewellyn[a], Humberto Sánchez[c], Susie Adamson[a], Graham A.J. Miles[a], Elsa López[b], Nilsa González[b], James S. Patterson[a], Michael W. Gaunt[a], Antonieta Rojas de Arias[b], Michael A. Miles[a]

[a]*Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London, UK*
[b]*Departamento de Medicina Tropical, Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción, Paraguay*
[c]*Museo Nacional de Historia Natural del Paraguay, San Lorenzo, Paraguay*

Abstract

Trypanosoma cruzi, the causative agent of Chagas disease, has at least two principal intraspecific subdivisions, *T. cruzi* I (TCI) and *T. cruzi* II (TCII), the latter containing up to five subgroups (a–e). Whilst it is known that TCI predominates from the Amazon basin northwards and TCII to the South, where the disease is considered to be clinically more severe, the precise clinical and evolutionary significance of these divisions remains enigmatic. Here, we present compelling evidence of an association between TCI and opossums (*Didelphis*), and TCII and armadillos, on the basis of key new findings from the Paraguayan Chaco region, together with a comprehensive analysis of historical data. We suggest that the distinct arboreal and terrestrial ecologies, respectively, of these mammal hosts provide a persuasive explanation for the extant *T. cruzi* intraspecific diversity in South America, and for separate origins of Chagas disease in northern South America and in the southern cone countries.
© 2004 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

Keywords: Armadillos; *Didelphis*; Edentates; Marsupials; *Trypanosoma cruzi*; Triatomines

1. Introduction

American trypanosomiasis, or Chagas disease, is the most important parasitic infection in Latin America. More than 10 million people carry the protozoan agent, *Trypanosoma cruzi*. The disease is a complex zoonosis, with mammals as natural reservoir hosts (Miles et al., 2003). Transmission is primarily by contact with the contaminated faeces of domiciliated blood sucking triatomine bugs.

Phenotypic and genotypic characters have been used to divide the species *T. cruzi* into at least two principal divisions

(Brisse et al., 2001; Campbell et al., 2004). Following international consensus, these divisions have been named *T. cruzi* I (TCI) and *T. cruzi* II (TCII) (Anon, 1999). *Trypanosoma cruzi* II is divided into up to five subgroups, IIa–e. By DNA sequence analysis, TCI consists of a single relatively homogeneous clade, whereas TCII emerges as two or three distinct phylogenetic clades (equivalent to IIa–c), with two hybrid lineages (IId, IIc) that have haplotypes split across the IIb and IIc clades (Brisse et al., 2001; Machado and Ayala, 2001, 2002; Gaunt et al., 2003). Further DNA sequencing has supported the idea that TCIIa and IIc have some affinities with TCI and indicated that they have hybrid characteristics, derived from TCIIb and TCI. In contrast, TCIIb and TCI appear to be less complex lineages (Sturm et al., 2003; Campbell et al., 2004). This present concept of *T. cruzi* divisions may change as many more *T. cruzi* isolates

* Corresponding author. Address: Room 331, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. Tel.: +44 20 7927 2814; fax: +44 20 7636 8739.
E-mail address: matthew.yeo@lshtm.ac.uk (M. Yeo).

are examined, and with a wider range of methods. Estimated dates for the divergence of TCI and TCII have varied between 3 million years ago (mya) and 88 mya (Briones et al., 1999; Machado and Ayala, 2001). Phylogenetic analyses and production of hybrids in the laboratory have shown that *T. cruzi* has both an ancient and active capacity for genetic exchange (Machado and Ayala, 2001; Gaunt et al., 2003).

Trypanosoma cruzi I was originally described from silvatic cycles in Brazil. This has led to the misconception that TCI is always associated with silvatic transmission cycles. This is not the case. TCI predominates from the Amazon basin northwards, where it is also the main cause of Chagas disease in endemic areas such as Venezuela (Miles et al., 1981). In contrast, TCII predominates in domestic transmission cycles in southern cone countries of South America (Argentina, Brazil, Bolivia, Chile, Paraguay, and Uruguay) (Chapman et al., 1984). Differences in pathology between TCI and TCII infections remain enigmatic. In the southern cone region (TCII) chagasic mega syndromes are common, whereas they are considered virtually absent north of the Amazon (TCI) (Miles et al., 1981). The implication is that TCII is more pathogenic than TCI (Di Noia et al., 2002). Limited comparisons in a C3H/HeN mouse model support the notion that certain TCII strains may shed more *trans*-sialidase, induce more thymus lesions and produce higher mortality than TCI strains (Risso et al., 2004).

Trypanosoma cruzi I has been consistently isolated from the common marsupial opossum, *Didelphis* species, north and south of the Amazon Basin. We therefore proposed that TCI has an evolutionary history associated with *Didelphis*, and possibly the triatomine tribe Rhodniini and the palm tree ecotope (Gaunt and Miles, 2000). Virtually all *Rhodnius* species are primarily associated with palm trees; exceptions are *Rhodnius domesticus*, from bromeliads and *Rhodnius paraensis*, known only from a single Amazonian tree hole. Interestingly, forms similar to infective stages in the triatomine vector are isolated from anal gland secretions of *Didelphis*, suggesting a supplementary non-vectorial route of transmission in this particular host (Carreira et al., 2001).

In contrast, the natural hosts of TCII are not clearly resolved (Jansen et al., 1999). Previous silvatic records are sparse (Breniere et al., 1998; Lisboa et al., 2000; Pinho et al., 2000; Barnabe et al., 2001). The isolation of *T. cruzi* IIa (formerly designated Z3; Povoa et al., 1984; Mendonca et al., 2002) from six armadillos and the associated triatomine vector *Panstrongylus geniculatus* in Amazon forest led us to speculate that armadillos might preferentially harbour TCII (Gaunt and Miles, 2000). Early mammalian fauna of South America included both placentals of the order Xenarthra (armadillos, sloths and anteaters) and the marsupials, all of which were present 65 mya (Webb and Marshall, 1982), whilst rodents, primates and bats arrived much later, 40 mya. Armadillos are still particularly abundant in central and southern regions of South America, where several genera are extant despite intensive hunting (Wetzel, 1982). The ancient South American ancestry of

armadillos and opossums, and their terrestrial and arboreal niches, respectively, suggested to us that they may have long-standing associations with different *T. cruzi* lineages.

Here, we characterise new *T. cruzi* isolates from extensive mammal captures in Paraguay. Additionally, we analyse all published host records of *T. cruzi* lineages to resolve host associations of the TCI and TCII genotypes. Together these data provide the most comprehensive analysis yet undertaken on the silvatic host associations of *T. cruzi*. We test the proposal that armadillos are natural hosts of TCII and didelphids the natural hosts of TCI.

2. Materials and methods

In view of the paucity of *T. cruzi* isolates from armadillos, we undertook extensive sampling and analysis of armadillos, other mammals and triatomines in remote areas of Paraguay, a country which, like Bolivia, is known to have remarkable *T. cruzi* lineage diversity. To test further the strength of association between TCI and *Didelphis* and TCII and armadillos we also analysed all *T. cruzi* host association data from the last 30 years.

2.1. Isolation and characterisation of Paraguayan strains

Study sites were Boquerón and Alto Paraguay in the western Chaco, a semi-arid region of xerophytic forest (23° S 60° W), and the department of San Pedro, at the southern most point of the Chaco, 300 km north of Asuncion (24° 15'S 56° 30'W). Animals, captured in live traps or by local hunters, were examined by xenodiagnosis and culture of blood, using 'through-the-cap' inoculation into diphasic media, as previously described (Miles, 1992). Animals were subsequently released unharmed.

Triatoma infestans was collected from houses. Peridomestic and silvatic bugs were searched for only in Boquerón, in the Chaco. Isolation of *T. cruzi* from triatomine bugs was performed using direct culture of triatomine faeces, as described by Miles (1992). *Trypanosoma cruzi* clones were also isolated by direct culture of infected triatomine faeces, on to improved solid medium plates, developed from the method described by Mondragon et al. (1999) (Yeo M., unpublished results).

Trypanosoma cruzi isolates were characterised by PCR amplification and sizing of products from the mini-exon gene, 18S rDNA and the D7 divergent domain of 24Sα rDNA (Brisse et al., 2001). Amplification of the non-transcribed spacer of the mini-exon gene used the pooled primers, TC (5'-CCC CCC TCC CAG GCC ACA CTG), TC1 (5'-GTG TCC GCC ACC TCC TTC GGG CC) and TC2 (5'-CCT GCA GGC ACA CGT GTG TGT G), with 27 cycles (30 s at 94 °C, 30 s at 55 °C, 30 s at 72 °C) followed by a final elongation of 5 min at 72 °C. Amplification of the size variable domain of the 18S rRNA used primers V1 (5'-CAA GCG GCT GGG TGG TTA TTC CA) and V2

Table 1
Amplification products in base pairs and lineage identification

	TCI	TCIIa	TCIIb	TCIIc	TCIId	TCIIe	Subgroup resolution (TCII)
Mini-exon	350	400[a]	300	250[b]	300	300	TCIIa,c
18S rDNA[c]	160	155	165	165	165	165	TCIIa
24Sα rDNA	110	120	125	110	110/125	125	TCIIa,c,d

[a] With no amplification at 300 or 350 bp regions.
[b] With no amplification at 300 bp.
[c] Unlike Brisse et al. (2001) both TCIIb and IIe gave a 165 bp amplicon.

(5-'TTG AGG GAA GGC ATG ACA CAT GT), with 30 cycles (1 min 94 °C, 1 min 50 °C, 1 min 72 °C) followed by elongation for 5 min at 72 °C. Amplification of the D7 divergent domain of the 24Sα RNA used primers D71 (5'-AAG GTG CGT CGA CAG TGT GG) and D72 (5'-TTT TCA GAA TGG CCG AAC AGT) with 30 cycles (1 min at 94 °C, 1 min at 60 °C, 1 min at 72 °C) and 5 min elongation at 72 °C. All amplifications were performed in a final volume of 20 μl containing 16 mM $(NH_4)_2SO_4$, 67 mM Tris–HCl (pH 8.8), 0.01% Tween-20, 1.5 mM $MgCl_2$, 2 mM of each dNTP, 20 pM of each primer and 0.5 units of Taq polymerase (Bioline, UK). Miniexon PCR products were separated in 1.5% agarose gels, 18S and 24Sα rRNA products in 3% agarose gels.

Sizing of these products distinguishes TCI and all of the TCII subgroups except TCIIb and IIe (Table 1). TCIIb and IIe were distinguished by isoenzyme phenotyping using glucosephosphate isomerase (GPI), phosphoglucomutase (PGM) and 6-phosphogluconate dehydrogenase (6-PGDH) as described by Chapman et al. (1984).

2.2. Analysis of historical data

Trypanosoma cruzi genotype and host association data were collated from publications and from our own unpublished records. Resolution of host associations was limited to the major lineages TCI and TCII, because of the small numbers of isolates of each subgroup of TCII. Statistical analysis used a standard χ^2 test of independence (StatCalc, Epi Info version 501b). Geographical mapping of the distribution of TCI and TCII to investigate sympatry was performed using GIS software (Arcview, ERSI). A complete list of source references is available from the authors on request.

3. Results

Results of the characterisation of recent T. cruzi isolates from Paraguayan mammals and triatomines are summarised in Table 2. A representative example of the results of genotyping by amplification of 24Sα is shown in Fig. 1. Genotypic characters accorded with those previously described in the literature (Brisse et al., 2001), except that, in our hands, both TCIIb and IIe gave 18S rDNA amplicons with a size of 165 bp. Representative isoenzyme data, as used to distinguish TCIIb and IIe, are shown in Fig. 2.

Strikingly, isolates from armadillos in Paraguay were exclusively identified as TCII, along with one isolate from the terrestrial marsupial, Monodelphis domestica.

Table 2
Genotypes of Trypanosoma cruzi II from Paraguayan hosts and vectors

Species	No (T. cruzi infected)[a]	T. cruzi II subgroup[b]
Dasypus novemcinctus (nine-banded armadillo)	38 (17)	TCIIc (15), TCIId (1)
Euphractus sexcinctus (six-banded armadillo)	23 (4)	TCIIb (1), TCIIc (2), TCIId (1)
Tolypeutes matacus (three-banded armadillo)	16 (0)	–
Chaetophractus spp.(hairy armadillo)	28 (1)	TCIIc (1)
Galea musteloides (common yellow-toothed cavy)	26 (0)	–
Calomys laucha (small vesper mouse)	5 (0)	–
Graomys griseoflavus (grey leaf-eared mouse)	3 (0)	–
Monodelphis domestica (short-tailed opossum)	3 (1)	TCIIc (1)
Didelphis albiventris (white-eared opossum)	3 (0)	–
Conepatus chinga (skunk)	1 (0)	–
Triatoma infestans (peridomestic)	45 (5)	TCIIb (2)
Triatoma infestans (domestic)	31[c]	TCIIb (4), TCIId (8), TCIIe (13), TCIIb/TCIIe (4)[d]
Triatoma sordida (silvatic)	30 (0)	–
Triatoma guasayana (silvatic)	283[e] (0)	–

[a] By xenodiagnosis and/or blood culture.
[b] Six isolates not characterised.
[c] 31 infected bugs selected from San Pedro, Chaco and Paraguari regions.
[d] Bugs with mixed infections.
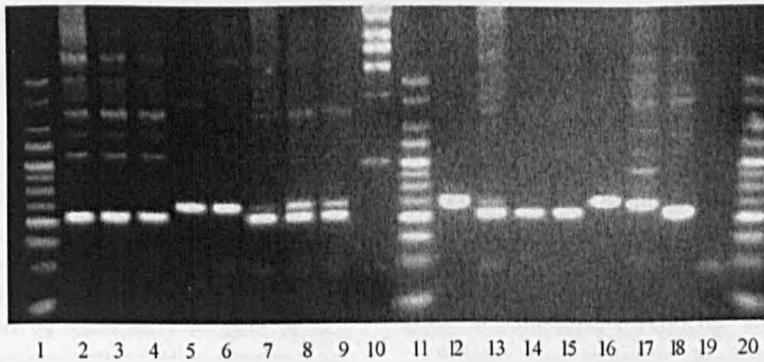[e] 43 of 47 nymphs of Triatoma sordida or Triatoma guasayana also uninfected and a further 168 T. guasayana not dissected.

Fig. 1. Amplification of the 24Sα rDNA divergent domain; lanes: 1, 0.5 kb ladder; 2–4, *Dasypus novemcinctus* TCIIc; 5, TCIIb reference; 6, *Euphractus sexcinctus* TCIIb; 7–8, domestic *Triatoma infestans* TCIId; 9, *D. novemcinctus* TCIId; 10, 1 kb ladder; 11, 0.5 kb ladder; 12–18, reference strains, TCIIe, TCIId, TCIIc, TCIIc, TCIIb, TCIIa and TCI, respectively; 19 control; 20, 0.5 kb ladder; origins at top.

Prevalence of infection in *Dasypus novemcinctus* was high. Furthermore, three subgroups of TCII were found in armadillos. These included the putative hybrid strain IId and the proposed parent-like strains IIb and IIc (Table 2).

We recorded three subgroups of TCII from domestic *T. infestans* in Paraguay, namely IIb, IId and the second hybrid strain IIe. In addition, we have previously found

TCIIc in domestic *T. infestans* (Chapman et al., 1984). Four domestic *T. infestans* carried multiclonal infections of TCIIe and IIb, detected by direct plating of clonal colonies from infected triatomine faeces and characterisation of individual clones, as illustrated in Fig. 3. *Trypanosoma cruzi* IIe clones were more abundant than IIb among the colonies from each multiclonal infection (Yeo M., unpublished) suggesting greater survival of the hybrid on this medium or its greater abundance in infected *T. infestans* (Finley and Dvorak, 1987).

Two colonies of peridomestic *T. infestans* were found. The colonies were within 300 m of a nearby indigenous community, in fallen trees among low scrub, associated with unidentified burrows and rodent faeces. Amphibians
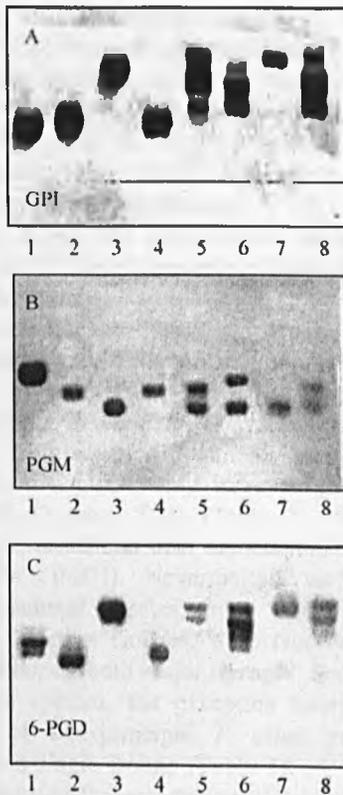
Fig. 2. Isoenzyme profiles of reference strains, illustrating the phenotypic characters distinguishing TCI and TCII and separating the subgroups of TCII. A, glucose phosphate isomerase. B, phosphoglucomutase. C, 6-phosphogluconate dehydrogenase; lanes, left to right: TCI (X10/1), TCIIa (CANIII), TCIIb (Esm clone2), TCIIc (X109/2), TCIId (92.80 clone1); TCIIe (CL Brener); *Euphractus sexcinctus* TCIIb and *Dasypus novemcinctus* TCIId; origins at bottom.
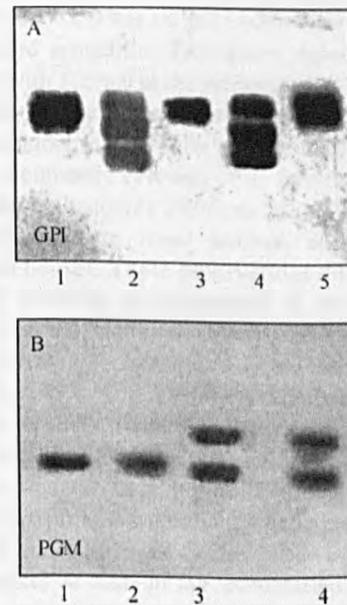
Fig. 3. Isoenzyme profiles of clonal *Trypanosoma cruzi* colonies, which were isolated by plating the faeces of naturally infected triatomines on to blood agar plates. It is apparent that multiclonal *T. cruzi* infections occur in single bugs. A, glucose phosphate isomerase; lanes 1, 3, 5, TCIIb, 2, 4 TCIIe B, phosphoglucomutase; lanes 1–2, TCIIb, 3, 4 TCIIe; origins at bottom.

and reptiles were also present. A total of 35 adult *T. infestans* and 10 nymphs were collected from the two sites. These peridomestic *T. infestans* were infected with TCIIb (Table 2).

Triatoma guasayana and *Triatoma sordida* co-infested the two peridomestic localities mentioned above, and these two triatomine species were also found at several other sites, under tree bark or in fallen trees, and attracted to light traps. *Triatoma guasayana* was particularly abundant but not recorded as infected with *T. cruzi* (Table 2). Although *T. guasayana* is known to fly seasonally (Gajete et al., 1996), contrary to common perceptions of triatomine bug behaviour, this aggressive species behaved in a surprising manner reminiscent of tsetse flies or mosquitos, flying, hovering and attacking humans in large numbers at night and during the day. For example, 35 *T. guasayana* were caught on three observers at night on a forest road beside a vehicle, field workers were also attacked in the Chaco forest at midday, and bugs were captured feeding on the authors in the vehicle both at night and during the day. *Psammolestes coreodes* (predominantly), *T. sordida*, *Triatoma platensis*, *T. guasayana* and *Triatoma melanosoma* were found in bird nests but, as to be expected, none of these specimens were infected with *T. cruzi* (González et al., 2000).

The results of the historical survey of all *T. cruzi* genotypes recorded from silvatic mammals, incorporating the above results from Paraguay, are shown in Table 3. *Trypanosoma cruzi* isolates from a total of 33 genera have been characterised, from the southern states of the USA to northern Argentina. These include isolates from marsupials, edentates, rodents, bats, and primates.

We applied a test for independence to examine the significance of *T. cruzi* genotype associations with particular host species. Distinct host affinities of the major *T. cruzi* lineages were apparent. Thus, 392 didelphid isolates were TCI and only seven were TCII. Furthermore, when marsupial genera were examined together (*Didelphis, Philander, Metachirus, Monodelphis, Marmosa* and *Caluromys*) 417 animals carried TCI and only 19 TCII. In contrast, only three armadillo isolates were TCI and 32 were TCII (Table 3). The test for independence revealed that both associations were highly significant ($P = <0.001$). Nevertheless, we noted that seven other mammal species, from which there were relatively few *T. cruzi* isolates, were recorded as being capable of harbouring both major lineages. For each of six of those seven species, the exception being *Philander frenata*, one of the principal *T. cruzi* lineages was represented by a single isolate (Table 3).

We also analysed the association of principal *T. cruzi* lineage with the niche of the mammal, whether considered to be arboreal or terrestrial, as given in Table 3. This analysis was also significant. For TCI, 443 isolates were from arboreal mammals, and 17 from terrestrial mammals. For TCII, we excluded golden lion tamarins from the analysis, because of the link with a putative epizootic

outbreak, as explained below. Of the remaining TCII isolates, 30 were from arboreal mammals and 52 from terrestrial mammals ($P = <0.001$). The marsupials *M. domestica* (three isolates) and *Monodelphis brevicaudata* (one isolate), which are exclusively terrestrial, were infected exclusively with TCII. Interestingly, 14 of the 30 TCII isolates from mammals considered to be arboreal were from the racoon, *Procyon lotor*, which is described as foraging both arboreally and terrestrially and as opportunistically sheltering in hollow trees or in burrows, such as those of armadillos (Eisenberg, 1989; Table 3).

Finally, we looked at the geographical origins of the isolates in Table 3, as depicted in Fig. 4 (maps a–d) to determine if there was a geographical separation between the distribution of TCI and TCII. This analysis reveals that, when the subgroups of TCI are ignored, the lineages TCI and TCII are not allopatric in their geographical distributions. However, we note that, as yet: only TCI and TCIIa have been reported in endemic areas north of the Amazon basin and in the USA; TCIIc is present in the Amazon basin and the southern cone countries, and TCIIb, IId and IIe are confined to the southern cone countries (data not shown). The association of TCI with the arboreal habitat and TCII with the terrestrial habitat was also apparent from GIS mapping (Fig. 4, maps e–h).

4. Discussion

The discovery of TCIIc, IIb and the corresponding hybrid lineage IId in armadillos in the Paraguayan Chaco, supports the hypothesis that TCII has an association with armadillos. The three-banded armadillo, *Tolypeutes matacus* was not found infected with *T. cruzi* in the present study. This species does not dig burrows suitable for triatomine infestation but simply shelters under leaf litter on the forest floor.

The two peridomestic colonies of *T. infestans* appear to be the dark morph phenotype (Noireau et al., 2000), similar to that found infesting local houses, and may have originated from houses. These observations do not indicate that silvatic *T. infestans* is widespread in the Paraguayan Chaco. However, the niche we describe is similar to that reported for silvatic *T. infestans* in the Bolivian Chaco (Noireau et al., 2000). This questions whether *T. infestans* has originated entirely from a habitat with caviomorph rodents at Cochabamba in Bolivia and spread from there throughout the southern cone region (Dujardin, 1987).

Previously, we proved that extant genetic recombination, via an unusual mechanism of fusion followed by genome erosion, can occur at least in the mammalian stage of the *T. cruzi* life cycle (Gaunt et al., 2003). This mechanism resembles whole genome duplication, known to accelerate evolution in *Saccharomyces cerevisae* (Kellis et al., 2004). The discovery of TCII putative parents and corresponding hybrids in armadillos sheds light on the possible origins of the virulent hybrid lineages IId and IIe, which predominate

Table 3
Hosts and lineages of *Trypanosoma cruzi* from historical studies

Genus	Species	Name	TCI	TCII subgroup					
				TCIIa	TCIIb	TCIIc	TCIId	TCIIe	TCII undefined
Akodon	Unspecified	Grass mouse	1						
Bradypus	*torquatus*	Sloth	1						
Caluromys	*lanatus*	**Woolly opossum**	1						
Cavia	*porcellus*	Cavy					3		1
Cavia	Unspecified	Cavy	3						
Coendou	Unspecified[a]	**Porcupine**	5						
Chaetophractus	Unspecified	Hairy armadillo				1[b]			
Cyclopes	*didactylus*	**Silky anteater**	1						
Dasyprocta	*fuliginosa*	Agouti		2					
Dasyprocta	Unspecified	Agouti	2						
Dasypus	*novemcinctus*	Nine-banded armadillo	1	6		18(15[b])	1[b]		
Dasypus	Unspecified	Nine-banded armadillo	2	2					
Didelphis	*marsupialis*	**Common opossum**	310	1	1			1	2
Didelphis	*azarae*	**Common opossum**	5						
Didelphis	*albiventris*	**Common opossum**	72		2				
Didelphis	*virginiana*	**Common opossum**	2						
Didelphis	Unspecified	**common opossum**	3						
Echymis	*chrysurus*	**Arboreal spiny rat**	1						
Echymis	*dasythrix*	**Arboreal spiny rat**	1						
Euphractus	*sexcinctus*	Six-banded armadillo		1[b]	2[b]	1[b]			
Holochilus	*Brasiliensis*[a]	Water rat	1						
Leontopithecus	*rosalia*[c]	**Golden lion tamarin**							39
Molossus	*molossus*	**Bat**	3	1					
Macaca	*mulatta*	**Rhesus monkey**		1					
Marmosa	*cinerea*	**Mouse opossum**	1						
Mephitidae	Undefined	Skunk							
Metachirus	*nudicaudatus*	Brn. four-eyed opossum	2						
Monodelphis	*brevicaudata*	Short-tailed opossum		4					
Monodelphis	*domestica*	Short-tailed opossum				1[b]			2
Nasua	*nasua*	**Coati**		1			2		
Nectomys	*squamipes*	Water rat	1						
Octodon	*degus*	Brush-tailed rat (degu)	1				6		
Octodontomys	Unspecified	Rat (degu)					2		
Oryzomys	*capito*	Rice rat	2						
Philander	*frenata*[a]	Gr. four-eyed opossum	6						4
Philander	*opossum*[a]	Gr. four-eyed opossum	15		1				
Procyon	*lotor*[a]	Racoon	1	14					
Proechimys	*semispinosus*	Terrestrial spiny rat	3	1					
Rattus	*rattus*	Black rat	7	1					
Saguinus	*midas*	**Red-handed tamarin**	1						
Sciurus	Unspecified	**Squirrel**	1	1					
Saimiri	*sciureus*	**Squirrel monkey**	1						
Tylomys	*mirae*	Climbing rat	3						

Bold italics indicates arboreal species.
[a] Species with terrestrial or aquatic habits but frequently nesting off the ground.
[b] Paraguayan isolates.
[c] See comments in text and Lisboa et al. (2004).

in domestic transmission cycles in Bolivia, Chile and Paraguay. We now propose that armadillos, harbouring co-infections, might prove to be the intermediate host that facilitated the evolution of hybrid lineages. An unproven hypothesis (Widmer et al., 1987) is that hybrids provide phenotypic advantage, for example by enhanced metabolic versatility over the wide range of environmental temperatures where they are prevalent, or in hosts, such as armadillos, subject to fluctuating body temperatures (30–36 °C; Arruda and Arruda, 1999).

Didelphis is rare in arid regions of the Paraguayan Chaco, so we were unable to undertake direct sympatric comparisons of *T. cruzi* lineages in armadillos and *Didelphis* in that precise locality. Three specimens of *Didelphis* captured in San Pedro were not infected with *T. cruzi*. However, consistent with the fact that armadillos are natural hosts of TCII, *Didelphis*, including examples from nearby sites in Argentina show significant bias for infection with TCI; rodents from Bolivia and Chile were also infected with TCI (Widmer et al., 1987; Diosque et al., 2003; Table 3).
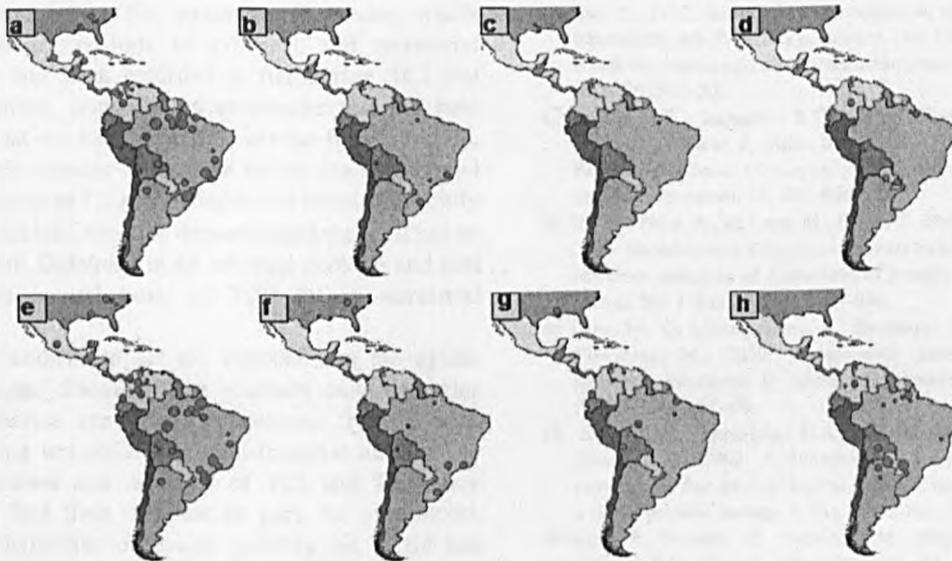
Fig. 4. GIS mapping to compare the known geographical distribution of the *Trypanosoma cruzi* lineages TCI and TCII, firstly in *Didelphis* as compared to armadillos, and secondly in arboreal mammals as compared to terrestrial mammals. TCI is associated with *Didelphis* and the arboreal habitat and TCII with armadillos and the terrestrial habitat. Note, however, that host and habitat associations are not absolute, and that the subgroups of TCII are pooled in this analysis, see text for comment on disparate geographical distributions of TCII subgroups: a, geographical distribution of TCI isolates from *Didelphis* b, distribution of TCII isolates from *Didelphis* c, TCI isolates from armadillos d, TCII isolates from armadillos e, TCI isolates from arboreal mammals f, TCII isolates from arboreal mammals (in this case including *Leontopithecus rosalia*, see text) g, TCI isolates from terrestrial mammals h, TCII isolates from terrestrial mammals; point sizes are proportional to the numbers of isolates at each location.

We originally speculated that TCII might predominate in terrestrial rodents, in part because they may share refuges with armadillos (Gaunt and Miles, 2000). Of 23 isolates from arboreal rodents, 21 were TCI and only two were TCII, supporting the partition of TCI and TCII between arboreal and terrestrial habitats. For isolates from terrestrial rodents 11 were TCII, however, nine were TCI (Table 3). There are, as yet, too few genotyped isolates to explore the possible host associations of the subgroups of TCII (Table 3).

The predominance of TCI in endemic areas of Chagas disease North of the Amazon is readily explained by the facility with which the triatomine bug *Rhodnius prolixus* has invaded human dwellings from palm trees, or been introduced into houses when their roofs are constructed from *Rhodnius* infested palm fronds.

It is less clear how TCII moved from terrestrial silvatic cycles into domestic transmission cycles in the southern cone countries. The subterranean vector associated with armadillos in Paraguay remains to be conclusively described. Large numbers of silvatic bugs (*T. guasayana* and *T. sordida*) were identified in the terrestrial but not the subterranean ecotope and neither species was infected with *T. cruzi*. One likely vector for transmission of TCII in silvatic ecotopes with armadillos is *P. geniculatus*, which is very widespread in South America and has been associated with burrowing animals. *Panstrongylus geniculatus* adults are attracted by light to dwellings, and may occasionally transmit *T. cruzi* to inhabitants but they rarely form domestic or peridomestic colonies. The common hunting and use of armadillos for food may have played a role in

establishing domestic transmission cycles. Whereas *Dasypus* does not survive easily for long in captivity, *Euphractus* is apparently easier to keep in a domestic environment. Both *Dasypus* and *Euphractus* may be kept for some time in houses before they are killed and eaten. The fact that they have been a principal food source, and still are among indigenous peoples, also provides an opportunity for transmission of TCII into people. *Dasypus* was the first silvatic host of *T. cruzi* to be described, by Carlos Chagas in Brazil (Chagas, 1912).

The association of TCI with an arboreal mammalian niche and TCII with a terrestrial niche is notable. One anomaly is the occurrence of TCII in *P. lotor* in the USA, already mentioned (above). A second anomaly is the finding of TCII in golden lion tamarins in the Brazilian Atlantic forest. Thirty eight tamarins were infected with TCII (Lisboa et al., 2000), whereas other sympatric arboreal mammals (marsupials and a sloth), were infected with TCI. However, the tamarin host records appear to be the result of an epizootic outbreak, because some animals were re-introduced to the wild from a captive population, known to be infected with *T. cruzi* II, and held in a triatomine-infested enclosure (Lisboa et al., 2004).

Independent transmission cycles occur in sympatric ecotopes (Pinho et al., 2000). Competitive exclusion, due to differential susceptibilities of host and vector species to distinct *T. cruzi* lineages, may reinforce the host and vector associations (de Lana et al., 2000; da Silveira et al., 2000). However, both principal *T. cruzi* lineages sometimes occur

in single host species. For example, *Philander*, which spends significant periods in arboreal and terrestrial environments, has been recorded as harbouring TCI and II. It is, therefore, important to emphasise that the host associations that we have described are far from absolute, and much work remains to be done before the history and adaptive evolution of *T. cruzi* lineages will be unequivocally known. Nevertheless, we have demonstrated that TCI has an association with *Didelphis* in an arboreal ecotope and that armadillos are natural hosts of TCII in the terrestrial ecotope.

Our observations support the concept that the agents of severe Chagas disease in the southern cone countries of South America are primarily derived from silvatic cycles involving terrestrial and semi-fossorial niches. The origin, distribution and diversity of TCI and TCII may thus, for the first time, at least in part, be understood. Furthermore, hybridisation events yielding the TCIId and IIe present in southern cone domestic cycles may have taken place in these silvatic terrestrial transmission cycles.

References

Anon, 1999. Recommendations from a satellite meeting. Mem. Inst. Oswaldo Cruz 94, 429–432.

Arruda, O.S., Arruda, M.S.P., 1999. Study on the median sacral artery ramification and the body temperature of the armadillo (*Dasypus novemcinctus*). Rev. Chil. Anat. 17, 147–151.

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.

Breniere, S.F., Morochi, W., Bosseno, M.F., Ordonez, J., Gutierrez, T., Vargas, F., Yaksic, N., Noireau, F., 1998. *Trypanosoma cruzi* genotypes associated with domestic *Triatoma sordida* in Bolivia. Acta Trop. 71, 269–283.

Briones, M.R., Souto, R.P., Stolf, B.S., Zingales, B., 1999. The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. Mol. Biochem. Parasitol. 104, 219–232.

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.

Campbell, D.A., Westenberger, S.J., Sturm, N.R., 2004. The determinants of Chagas disease: connecting parasite and genetics. Curr. Mol. Med. 4, 549–562.

Carreira, J.C., Jansen, A.M., de Nazareth Meirelles, M., Costa e Silva, F., Lenzi, H.L., 2001. *Trypanosoma cruzi* in the scent glands of *Didelphis marsupialis*: the kinetics of colonization. Exp. Parasitol. 97, 129–140.

Chagas, C., 1912. Sobre um trypanosome do tatu, *Tatusia novemcincta*, transmitido pela *Triatoma geniculata* Latr. (1811). Possibilidade do ser o tatu um depositario do *Trypanosoma cruzi* no mundo exterior. Bras. Med. 26, 305–306.

Chapman, M.D., Baggaley, R.C., Godfrey-Fausset, P.F., Malpas, T.J., White, G., Canese, J., Miles, M.A., 1984. *Trypanosoma cruzi* from the Paraguayan Chaco: isoenzyme profiles of strains isolated at Makthlawaiya. J. Protozool. 31, 482–486.

da Silveira Pinto, A., de Lana, M., Britto, C., Bastrenta, B., Tibayrenc, M., 2000. Experimental *Trypanosoma cruzi* biclonal infection in *Triatoma infestans*: detection of distinct clonal genotypes using kinetoplast DNA probes. Int. J. Parasitol. 30, 843–848.

de Lana, M., da Silviera Pinto, A., Bastrenta, B., Barnabe, C., Noel, S., Tibayrenc, M., 2000. *Trypanosoma cruzi*: infectivity of clonal genotype infections in acute and chronic phases in mice. Exp. Parasitol. 96, 61–66.

Di Noia, J.M., Buscaglia, C.A., De Marchi, C.R., Almeida, I.C., Frasch, A.C., 2002. A *Trypanosoma cruzi* small surface molecule provides the first immunological evidence that Chagas disease is due to a single parasite lineage. J. Exp. Med. 195, 401–413.

Diosque, P., Barnabe, C., Padilla, A.M., Marco, J.D., Cardozo, R.M., Limino, R.O., Nasser, J.R., Tibayrenc, M., Basombrio, M.A., 2003. Multilocus enzyme electrophoresis analysis of *Trypanosoma cruzi* isolates from a geographically restricted endemic area for Chagas disease in Argentina. Int. J. Parasitol. 33, 997–1003.

Dujardin, J.P., Tibayrenc, M., Venegas, E., Maldonado, L., Desjeux, P., Ayala, F.J., 1987. Isozyme evidence of lack of speciation between wild and domestic *Triatoma infestans* (Heteroptera: Reduviidae) in Bolivia. J. Med. Entomol. 24, 40–45.

Eisenberg, J.F., 1989. Mammals of the Tropics, Volume 1: the Northern Tropics. University of Chicago Press, IL, USA.

Finley, R.W., Dvorak, J.A., 1987. *Trypanosoma cruzi*: analysis of the population dynamics of heterogeneous mixtures. J. Protozool. 34, 409–415.

Gajete, P.P., Bottiazzi, M.V., Pietrokovsky, S.M., Wisnivesky-Colli, C., 1996. Potential colonization of the peridomicile by *Triatoma guasayana* (Hemiptera: Reduviidae) in Santiago del Estero. Argentina J. Med. Entomol. 33, 635–639.

Gaunt, M.W., Miles, M.A., 2000. The ecotopes and evolution of triatomine bugs (Triatominae) and their associated trypanosomes. Mem. Inst. Oswaldo Cruz 95, 557–565.

Gaunt, M.W., Yeo, M., Frame, I.A., Stothard, J.R., Carrasco, H.J., Taylor, M.C., Mena, S.S., Veazey, P., Miles, G.A., Acosta, N., de Arias, A.R., Miles, M.A., 2003. Mechanism of genetic exchange in American trypanosomes. Nature 421, 936–939.

González, N., Vega, C., Rolón, M., Rojas de Arias, A., 2000. Triatiomineos y outros arthropodos en nidos de aves de comunidades indigenas y menonitas del Chaco Paraguayo. XVth Int. Congr. Trop. Med. Mal. 2, 86.

Jansen, A.M., Santos de Pinho, A.P., Lisboa, C.V., Cupolillo, E., Mangia, R.H., Fernandes, O., 1999. The sylvatic cycle of *Trypanosoma cruzi*: a still unsolved puzzle. Mem. Inst. Oswaldo Cruz 94 (Suppl. 1), 203–204.

Kellis, M., Birren, B.W., Lander, S.L., 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature (London) 428, 617–624.

Lisboa, C.V., Dietz, J., Baker, A.J., Russel, N.N., Jansen, A.M., 2000. *Trypanosoma cruzi* infection in *Leontopithecus rosalia* at the Reserva Biologica de Poco das Antas, Rio de Janeiro, Brazil. Mem. Inst. Oswaldo Cruz 95, 445–452.

Lisboa, C.V., Mangia, R.H., Rubiao, E., de Lima, N.R., das Chagas Xavier, S.C., Picinatti, A., Ferreira, L.F., Fernandes, O., Jansen, A.M., 2004. *Trypanosoma cruzi* transmission in a captive primate unit, Rio de Janeiro, Brazil. Acta Trop. 90, 97–106.

Machado, C.A., Ayala, F.J., 2001. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of *Trypanosoma cruzi*. Proc. Natl Acad. Sci. USA 98, 7396–7401.

Machado, C.A., Ayala, F.J., 2002. Sequence variation in the dihydrofolate reductase-thymidylate synthase (DHFR-TS) and trypanothione reductase (TR) genes of *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 121, 33–47.

Mendonca, M.B., Nehme, N.S., Santos, S.S., Cupolillo, E., Vargas, N., Junqueira, A., Naiff, R.D., Barrett, T.V., Coura, J.R., Zingales, B., Fernandes, O., 2002. Two main clusters within *Trypanosoma cruzi* zymodeme 3 are defined by distinct regions of the ribosomal RNA cistron. Parasitology 124, 177–184.

Mondragon, A., Wilkinson, S.R., Taylor, M.C., Kelly, J.M., 1999. Optimization of conditions for growth of wild-type and genetically transformed *Trypanosoma cruzi* on agarose plates. Parasitology 118, 461–467.

Miles, M.A., 1992. in: Hyde, E.H. (Ed.), Protocols in Molecular Parasitology. Humanae Press, Clifton, UK, pp. 15–28.

Miles, M.A., Cedillos, R.A., Povoa, M.M., de Souza, A.A., Prata, A., Macedo, V., 1981. Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas disease?. Lancet 1, 1338–1340.

Miles, M.A., Feliciangeli, M.D., de Arias, A.R., 2003. American trypanosomiasis (Chagas disease) and the role of molecular epidemiology in guiding control strategies. Br. Med. J. 326, 1444–1448.

Noireau, F., Bastrenta, B., Catala, S., Dujardin, J.P., Panzera, F., Torres, M., Perez, R., Galvao, C., Jurberg, J., 2000. Sylvatic population of *Triatoma infestans* from the Bolivian Chaco: from field collection to characterization. Mem. Inst. Oswaldo. Cruz 95 (Suppl. 1), 119–122.

Pinho, A.P., Cupolillo, E., Mangia, R.H., Fernandes, O., Jansen, A.M., 2000. *Trypanosoma cruzi* in the sylvatic environment: distinct transmission cycles involving two sympatric marsupials. Trans. R. Soc. Trop. Med. Hyg. 94, 509–514.

Povoa, M.M., de Souza, A.A., Naiff, R.D., Arias, J.R., Naiff, M.F., Biancardi, C.B., Miles, M.A., 1984. Chagas disease in the Amazon basin IV. Host records of *Trypanosoma cruzi* zymodemes in the states of Amazonas and Rondonia, Brazil. Ann. Trop. Med. Parasitol. 78, 479–487.

Risso, M.G., Garbarino, G.B., Mocetti, E., Campetella, O., Gonzalez Cappa, S.M., Buscaglia, C.A., Leguizamon, M.S., 2004. Differential expression of a virulence factor, the *trans*-sialidase, by the main *Trypanosoma cruzi* phylogenetic lineages. J. Infect. Dis. 189, 2250–2259.

Sturm, N.R., Vargas, N.S., Westenberger, S.J., Zingales, B., Campbell, D.A., 2003. Evidence for multiple hybrid groups in *Trypanosoma cruzi*. Int. J. Parasitol. 33, 269–279.

Webb, S.D., Marshall, L.G., 1982. Mammalian biology in South America, in: Mares, M.A., Genoways, H.H. (Eds.), Pymatuning Symposia in Ecology 6. Special Publication Series. University of Pittsburgh, Pittsburgh, PA, pp. 39–52.

Wetzel, R.M., 1982. Systematics, distribution, ecology, and conservation of South American edentates, in: Mares, M.A., Genoways, H.H. (Eds.), Mammalian Biology in South America, Pymatuning Symposia in Ecology 6. Special Publication Series. University of Pittsburgh, Pittsburgh, PA, pp. 345–375.

Widmer, G., Dvorak, J.A., Miles, M.A., 1987. Temperature modulation of growth rates and glucosephosphate isomerase isozyme activity in *Trypanosoma cruzi*. Mol. Biochem. Parasitol. 23, 55–62.

Resolution of multiclonal infections of *Trypanosoma cruzi* from naturally infected triatomine bugs and from experimentally infected mice by direct plating on a sensitive solid medium

Matthew Yeo [a,*], Michael D. Lewis [a], Hernan J. Carrasco [b], Nidia Acosta [c], Martin Llewellyn [a], Sebastião Aldo da Silva Valente [d], Vera de Costa Valente [d], Antonieta Rojas de Arias [c], Michael A. Miles [a]

[a] *Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK*
[b] *Instituto de Medicina Tropical, Facultad de Medicina, Universidad Central de Venezuela, Caracas, Venezuela*
[c] *Departamento de Medicina Tropical, Instituto de Investigaciones en Ciencias de la Salud, Universidad Nacional de Asunción, Paraguay*
[d] *Instituto Evandro Chagas, Belém, Pará State, Brazil*

Abstract

The isolation of biological clones of *Trypanosoma cruzi* by microscopically dispensing individual organisms or by serial dilution is laborious and time consuming. The inability to resolve mixed *T. cruzi* infections, from vectors and hosts, and to isolate clones of slow growing genotypes by efficient plating on solid media, has hindered characterisation studies and downstream applications. We have devised and validated a sensitive, solid medium plating technique for rapid in vitro isolation of clones representative of all the recognised *T. cruzi* lineages (TCI, TCIIa–e), including the slow growing strain CANIII (TCIIa) and *Trypanosoma rangeli*, with high plating efficiencies. Furthermore, the method is effective for the isolation of clones directly from silvatic triatomine bugs and from experimentally infected mice harbouring mixed infections, allowing resolution of multiclonal infections from varied sources.
© 2006 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

Keywords: Cloning; Culture; Lineage; Multiclonal; *Trypanosoma cruzi*; *Trypanosoma rangeli*

1. Introduction

Historically, methods for the isolation of *Trypanosoma cruzi* clones from triatomine bugs and mammals are time consuming and often ineffective. Usual methods involve the isolation of single organisms by direct microscopy or by limiting dilution (Miles, 1974).

Efforts to clone organisms on solid media have been described for several parasitic protozoa, such as *Blastocistys hominis* (Tan et al., 2000), *Entamoeba histolytica*, *Giardia lamblia* (Gillin and Diamond, 1978), the insect flagellates, as well as the trypanosomatids, including *Trypanosoma brucei* (Carruthers and Cross, 1992) and *Leishmania* (Gambarelli and Dumon, 1988). Variants of the technique have been applied to isolation of *T. cruzi* mutants, screening drug susceptibility, detection of mixed infections, screening genetically transfected clones (Santos et al., 2000), and other transfection studies (Mondragon et al., 1999).

Clonal isolation of *T. cruzi* on solid medium has been cited in a limited number of publications (Goldberg and Chiari, 1980; Mondragon et al., 1999; Santos et al., 2000; Podlipaev and Naumov, 2000). Goldberg and Chiari

---

* Corresponding author. Tel.: +44 20 7927 2814; fax: +44 20 7636 8739.
E-mail address: Matthew.yeo@lshtm.ac.uk (M. Yeo).

(1980) originally demonstrated clonal growth and isolation on agar plates. Typically, epimastigotes are spread over the surface of agar plates which are then incubated (Goldberg and Chiari, 1980). Various nutrient combinations have been tested, including LIT growth medium (liver infusion-tryptose), BLAB medium (0.75% agar supplemented with 48.4% LIT medium, 48.4% brain–heart infusion, and 2.45% defibrinated human blood), and RPMI 1640 (GibcoBRL) based medium (Mondragon et al., 1999).

The plating efficiencies (percentage of cells inoculated that form a colony) of previous methods varied widely, ranging from 5%, for the laboratory strains CL Brener and Y (Goldberg and Chiari, 1980), to 87% for an uncharacterised stock (Gomes et al., 1991). Colonies usually became visible after 20–35 days, reaching a diameter of approximately 1 mm. The primary morphological forms reported were epimastigotes. Plating efficiencies were affected by a number of factors including solidification time and temperature (Ng-Ying-Kin and Yaphe, 1972), agar sulphate content (Santos et al., 2000), inoculum size, and growth phase (Mondragon et al., 1999). However, Goldberg and Chiari (1980) found no correlation between growth phase and plating efficiency.

Obstacles to cloning *T. cruzi*, in comparison with *T. brucei* and *Leishmania*, which form discrete colonies in 5–7 days, result from the relatively slow rates of *T. cruzi* growth (Mondragon et al., 1999). Gels eventually desiccate and inhibit growth, although this can be partially solved by the use of a humidified incubator (Mondragon et al., 1999). *Trypanosoma cruzi* growth is also inhibited by basic pH. In rapidly dividing cultures organic acids produced by glycolysis reduce inhibition of growth. The use of a 4% CO_2 incubator can facilitate maintenance of an appropriate pH range. Additionally, the well-documented genetic heterogeneity of *T. cruzi* is problematic. *Trypanosoma cruzi* is divided into at least two major lineages, *T. cruzi* I (TCI) and *T. cruzi* II (TCII), with *T. cruzi* II containing up to five groups, TCIIa–e. TCIId and TCIIe are relatively recent aneuploid hybrids of TCIIb and TCIIc. TCIIa and TCIIc may themselves be more ancient hybrids, in which diploidy has been restored, with TCI and TCIIb the likely progenitors (Brisse et al., 2001; Machado and Ayala, 2001; Gaunt et al., 2003; Sturm et al., 2003; Westernberger et al., 2005). Different strains of *T. cruzi* have been proven to vary dramatically in their respective doubling times during in vitro growth, ranging from 6 days (CAN III), to 22 h (CL Brener) or 16 h (X10/6) (Finley and Dvorak, 1987; Mondragon et al., 1999). Only one previous attempt has been made to grow colonies of strains representative of each of the two major lineages TCI and TCII and a slow growing strain of TCIIa. Mondragon et al. (1999) were unsuccessful in clonal propagation of the slow growing CANIII. They concluded that modifications to the method would be necessary to grow strains with such long doubling times.

Multiple *T. cruzi* genotypes are known to occur sympatrically in some geographical regions (Chapman et al., 1984; Yeo et al., 2005). Furthermore, it is known that in some

field locations multiple genotypes exist in individual bugs (Bosseno et al., 1996), mammals, and humans (Solari et al., 2001). Xenodiagnosis, a technique frequently used in Latin America, or direct inoculation of mice with infected triatomine faeces can lead to selective overgrowth of one genotype, potentially biasing research on epidemiology, host/vector associations, and strain-specific pathologies. Indirect methods of identifying mixed infections, for example using kinetoplast DNA probes, although of value, depend on unequivocal specificity and homogeneity of target sequences, despite hybridisation events, and do not yield *T. cruzi* isolates for further investigation. Additionally, in some field locations up to 50% of isolates from domestic triatomine bugs do not grow in LIT medium (Rojas de Arias, unpublished data) and are lost, confirming the need for a single method to both sustain growth and produce clones for all lineages of *T. cruzi*.

The experimental rationale here was to develop a sensitive plating technique to (i) enable the effective growth and isolation of individual clones from representative strains, including those that are slow growing, (ii) facilitate growth without a CO_2 incubator, and (iii) allow resolution of mixed infections directly from triatomine bug faeces and mammals. We thus describe a sensitive new tool for the acquisition of genetically distinct clones of *T. cruzi*, encompassing all the known genetic groups.

2. Materials and methods

2.1. Solid phase plating for the isolation of clonal colonies of T. cruzi

Direct isolation of clonal colonies of *T. cruzi* required the preparation of highly sensitive nutrient-fortified agar plates. The objective was to stimulate growth and inhibit cell mobility while avoiding contamination. An experiment was initially designed to compare plating efficiencies of different strains representing the two major lineages TCI and TCII (X10/1 and Esm cl2, respectively) and the slow growing strain CANIII (TCIIa). Strains were grown on different nutrient combinations to optimise sensitivity for clonal isolation (Table 1). Additionally, plating efficiencies were compared between plates maintained in a CO_2 enriched atmosphere (4%) and those grown in airtight, sealed Petri dishes. The optimum nutrient and incubation conditions were then applied to an expanded range of cultured *T. cruzi* (Table 2) and to direct resolution of *T. cruzi* infections from naturally infected silvatic triatomine bugs and from experimentally infected laboratory mice. The method was also applied to the related trypanosome *Trypanosoma rangeli*. Control plates for each of the experiments were prepared in parallel using saline as the inoculum.

Plates containing growth media were prepared in two stages. Initially, an agar base underlay fortified with nutrients was prepared. Second, a low melting point agarose (LMP) overlay was added containing the inoculum and in some cases further nutrients.

Table 1
Growth comparisons for three *Trypanosoma cruzi* reference strains

Plate group	Plate type[a]	Incubation conditions	X10/1 (TCI)		ESM cl2 (TCII)		CANIII (TCIIA)	
			First visible colony (days)	Plate effic (%)	First visible colony (days)	Plate effic (%)	First visible colony (days)	Plate effic (%)
A	RPMI underlay + saline overlay	CO_2	11	54	22	22	32	5
		Sealed	18	34	27	23	32	2
B	RPMI underlay + RPMI overlay	CO_2	11	58	19	38	29	22
		Sealed	19	36	24	42	26	8
C	Blood agar (10% whole blood) + saline overlay	CO_2	9	64	19	36	31	21
		Sealed	18	55	21	27	36	14
D	Blood agar (10% FCS) + saline overlay	CO_2	18	4	N/A	0	N/A	0
		Sealed	26	12	N/A	0	N/A	0
E	Blood agar (10% lysed whole blood) + saline overlay	CO_2	9	52	24	6	N/A	0
		Sealed	20	40	27	9	N/A	0
F	Blood agar (10% lysed and centrifuged blood) + saline overlay	CO_2	10	52	19	9	N/A	0
		Sealed	19	43	N/A	0	N/A	0
G	Blood agar (10% whole blood) + RPMI overlay	CO_2	9	75	19	32	27	26
		Sealed	18	61	22	21	26	17
H	Blood agar (10% FCS) + RPMI overlay	CO_2	13	12	27	0	N/A	0
		Sealed	24	2	N/A	0	N/A	0
I	Blood agar (10% lysed blood) + RPMI overlay	CO_2	10	62	20	4	29	0
		Sealed	19	14	32	7	N/A	0
J	Blood agar (10% lysed and centrifuged blood) + RPMI overlay	CO_2	9	63	24	11	33	1.6
		Sealed	19	44	35	12	N/A	0

[a] Plate types were performed in triplicate for both CO_2 and sealed plates; first visible colonies (>0.2 mm); N/A denotes no growth occurred. Plating efficiency is defined as the number of cells that grow into colonies per 100 cells inoculated expressed as a percentage.

Table 2
Plating an expanded range of *Trypanosoma cruzi* genotypes

Isolate	Group	Days to harvesting	No. of colonies expanded
B187	TCI	30	5
Chile C22	TCI	30	6
Chile Wall	TCI	15	6
ERA	TCIIa	23	4
X10610	TCIIa	23	3
Pot 7a	TCIIb	37	5
Pot 7b	TCIIb	37	3
M5631	TCIIc	13	4
JA2	TCIIc	15	5
Arma 13	TCIIc	15	6
Arma 18	TCIIc	30	6
Chaco 2	TCIId	30	6
Para 5	TCIId	30	2
Para 6	TCIId	30	5
Para 7	TCIId	30	3
Para 1	TCIId	30	2
Para 4	TCIId	30	3
Chaco 21	TCIIe	60	10
Chaco 24	TCIIe	60	10
Trypanosoma rangeli	–	13	4

Plates were type C, maintained in 4% CO_2 in a humidified incubator at 28 °C.

2.2. Preparation of the underlay

Five types of underlay were used (Table 1), of which four were variants of blood agar medium consisting of Difco Bacto agar base (Difco, E. Molesley, Surrey UK) 1.4% (w/v), trypticase peptone (BBL, Cowley, Oxford, UK), 0.5% (w/v), NaCl 0.6% (w/v), purified agar (Oxoid L28) 0.5% (w/v); the mixture was autoclaved (121 °C, 15 min) and allowed to cool to 50 °C. Blood agar base (70133, Sigma) can be used as an alternative to Difco Bacto agar base. Aseptically defibrinated rabbit blood or its derivatives (Table 1) was added to a final concentration of 10%; 5-fluorocytosine and gentamycin were each added to a final concentration of 100 µg/ml. 10.8 ml of the solution was dispensed rapidly into 9-cm Petri dishes (Sterilin) and allowed to set for 3–5 min.

The fifth underlay consisted of RPMI 1640 medium (GIBCO BRL, Paisley, Scotland) supplemented with 0.5% (w/v) trypticase (BBL), 0.5% (w/v) Hepes, 0.03 M haemin, 10% (v/v) FCS (heat-inactivated), 2 mM sodium glutamate, 2 mM sodium pyruvate, with gentamycin (150 µg/ml), and 5-fluorocytosine (150 µg/ml). In this case the underlay was prepared by mixing 2.2 ml of molten 3% bacto agar (oxoid L28) and 10.8 ml of supplemented RPMI before dispensing into Petri dishes.

Plates were sealed with parafilm (American National) and stored at 4 °C in readiness for the overlay. Prior to the addition of the overlay the plates were warmed to ambient temperature for 15 min.

2.3. Preparation of the overlay

Two overlay preparations were used (Table 1) containing either 0.9% sterile saline or supplemented RPMI, with antibiotics (100 µg/ml 5-fluorocytosine and 100 µg/ml gentamycin). Final overlays consisted of 2.4 ml of saline or supplemented RPMI added to 0.6 ml of 3% LMP. Components were maintained at 40 °C, source material (maximum volume 150) containing *T. cruzi* was added and the suspension mixed and poured over pre-prepared underlay. Plates were allowed to set for approximately 5 min before incubation.

2.4. Testing growth conditions using reference strains

Plates were inoculated with 1000 epimastigotes from axenic culture. The appropriate volume of inoculum was calculated by haemocytometer (Scientific Laboratory Supplies). Separate plates were inoculated with reference strains, X10/1 (TCI), Esm cl2 (TCIIb) and CANIII (TCIIa) using mid-log-growth phase cultures. Medium combinations and incubation conditions are summarised in Table 1. Each plate type (A–J) was used in triplicate for each nutrient combination and maintained in sealed, air tight petri dishes preventing gaseous exchange (sealed plates). Identical plates, also in triplicate, were prepared and sealed with Micropore tape (Unichem) enabling gaseous exchange, in 4% CO_2 in a humidified incubator. The method supporting the most rapid growth and highest plating efficiencies, was applied to an expanded set of cultured isolates representing TCI and TCIIa–e (Table 2) and *T. rangeli*, to assess if a single nutrient type would enable growth of all strains.

2.5. Trypanosoma cruzi from silvatic triatomine bugs

Triatomine bugs were immersed for 10 min in White's solution (Miles, 1993; HgCl₂ 0.025 g, NaCl 0.65 g, HCl (sp. gr. 1. 18) 0.125 ml, ethanol (abs) 25 ml, H₂0 75 ml) and rinsed in 0.9% sterile saline solution containing 300 µg/ml gentamycin and 300 µg/ml 5-fluorocytosine. Bugs were dried with a sterile mediwipe. Gut contents were dissected by drawing out the rectum and hindgut into a drop of saline solution (containing 300 µg/ml gentamycin, 300 µg/ml 5-fluorocytosine) on a sterile microscope slide and homogenising with a blunt microspatula. The majority of the contents were then transferred into a sterile eppendorf containing up to 1 ml of sterile PBS (150 µg/ml gentamycin, 150 µg/ml 5-fluorocytosine) in readiness for plating. Material remaining on the slide was examined microscopically for the presence of trypanosomes. The concentration of cells in the eppendorf was determined using a haemocytometer.

2.6. Isolation of trypanosomes from peripheral blood

Single and mixed infections were established in mice using two previously characterised *T. cruzi* strains, Esm cl2 (TCIIb) and ARMA 12 (TCIIc), as governed by licence from the relevant UK authorities. Mice were inoculated simultaneously with 10^4 metacyclic trypomastigotes of each

strain, derived from stationary phase blood agar cultures. Whole blood plating involved the incorporation of 10–150 µl of whole blood with different parasitemias directly into the LMP layer (Table 3). Direct plating of plasma was also performed, whereby whole blood was left to settle for 1 h (in vials containing EDTA anticoagulant, Monovette, Sarstedt, Germany) or centrifuged at low speed (40g, 5 min). Vials were then incubated at 37 °C for 45 min allowing the motile trypomastigotes to become dispersed in the plasma. Cell counts were performed and aliquots (10–150 µl) added to the LMP overlay solution maintained at 40 °C. The suspension was homogenised and dispensed rapidly onto pre-prepared underlay.

2.7. Harvesting colonies

Single clonal colonies were picked off plates, using a sterile pipette tip, and transferred into either blood agar slopes or 24-well cell culture plates (Nunclon) containing 1 ml of nutrient fortified RPMI. Cultures were incubated at 28 °C until log phase growth was observed. DNA was then extracted (DNeasy, Qiagen) according to the manufacturers instructions.

2.8. Characterisation

2.8.1. Isoenzyme electrophoresis

Thin layer starch gel electrophoresis (Miles et al., 1981) and cellulose acetate electrophoresis (Lanham et al., 1981), were performed on clones expanded in culture. Enzyme systems employed were aspartate aminotransferase (E.C.2.6.1.1, ASAT); glucose phosphate dehydrogenase (E.C.1.1.1.49, G6PD); glucose phosphate isomerase (E.C.5.3.1.9, GPI); phosphoglucomutase (E.C.2.7.5.1, PGM), and 6-phosphogluconate dehydrogenase (E.C.1.1.1.44, 6PGD).

2.8.2. Random amplification of polymorphic DNA (RAPD)

Reaction conditions were based on those described by Carrasco et al., 1996) using primers A1 5'-TCACGA TGCA, A2 5'-GAAACGGGTG, A4 5'-AATCGGGCTG, A6 5'-GTGATCGCAG, H1 5'-CGCGCCCGCT, L4 5'-G TGGATGCGA, L5 5'-AAGAGCCCGT. All amplifications were performed in a Hybaid thermal reactor (Hybaid Ltd., Middlesex, United Kingdom). Cycles were as follows: two cycles at 95 °C for 5 min, 30 °C for 2 min and 72 °C for 1 min, 32 cycles at 95 °C for 1 min, 40 °C for 30 s and 72 °C for 2 min; with a final cycle with extension at 72 °C for 5 min. Reaction products were electrophoresed on a 2% agarose gel (Sigma Chemical Co., St Louis, Mo) in 1x TAE buffer (0.04 M Tris–acetate, 0.001 M EDTA), stained with ethidium bromide, and visualised on a transilluminator, under UV light.

2.8.3. PCR amplification of the non-transcribed spacer of the mini-exon gene

Characterisation of the mini-exon was according to Souto et al. (1996) using a pool of three primers TC (5'-CCC CCC TCC CAG GCC ACA CTG), TCI (5'-GTG TCC GCCACC TCC TTC GGG CC), and TCII (5'-CCT GCA GGC ACA CGT GTG TGT G). Amplification was with 27 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s, followed by final extension at 72 °C for 5 min. PCR products were separated by electrophoresis in 1.5% agarose gels for 90 min at 90 V in 0.5x TBE buffer (Sigma–Aldrich), stained with ethidium bromide, and visualised under ultraviolet light.

2.8.4. PCR amplification of the D7 divergent domain of the 24Sα rRNA

Characterisation of the 24Sα rRNA was performed following the method described by Souto et al. (1996) using primers D71 (5'-AAG GTG CGT CGA CAG TGT) and

Table 3
Plating *Trypanosoma cruzi* from experimentally infected mice

Plate number	Strain inoculated	Time to first visible colony (days)	Inoculum volume (µl)	Parasitiemia (cells/ml)	No. of cells plated	Plating efficiency at 30 days (%)
1	TCIIc[a]	33	25	50	1.25	0
2	TCIIc[a]	28	50	50	2.5	0
3	TCIIc[b]	26	100	50	5	0
4	TCIIc[a]	32	100	50	5	0
5	TCIIc[b]	35	150	50	7.5	0
6	TCIIc[a]	25	150	50	7.5	0
7	TCIIc + TCIIb[a]	26	25	250	6.25	0
8	TCIIc + TCIIb[a]	32	50	250	12.5	0
9	TCIIc + TCIIb[a]	25	100	250	25	8
10	TCIIc + TCIIb[a]	32	10	6000	60	8.3
11	TCIIc + TCIIb[a]	30	20	6000	120	10.8
12	TCIIc + TCIIb[a]	28	30	6000	180	0
13	TCIIc + TCIIb[b]	32	50	6000	300	6.6
14	TCIIc + TCIIb[a]	33	100	6000	600	24

Plates were type C, maintained in 4% CO_2 in a humidified incubator at 28 °C.
[a] Whole blood inoculum.
[b] Plasma inoculum.

D72 (5′-TTT TCA GAA TGG CCG AAC AGT). Cycles were as follows: 30 cycles of 94 °C for 60 s, 60 °C for 60 s, 72 °C for 60 s, followed by a final elongation step of 5 min at 72 °C. Amplified products were separated by electrophoresis at 80 V over 2 h in 3% agarose gels (Sigma) with 0.5× TBE buffer, stained with ethidium bromide, and visualised under UV light.

3. Results

3.1. Growth of reference strains

The three major factors determining colony growth and plating efficiency were the *T. cruzi* strain, the nutrient type and the incubation conditions. No *T. cruzi* colony growth occurred on un-inoculated control plates and these were excluded from analysis.

Reference strains differed dramatically in the time to first colony appearance and plating efficiency. When maintained in a 4% CO_2 atmosphere, strain X10/1 grew on all plate combinations, with a mean of 10.9 days to first visible colonies. Esm cl2 grew on nine of 10 plate combinations (mean 21.4 days) and CANIII grew on only six of 10 plate combinations (mean 30.1 days), as shown in Table 1. X10/1 colonies were visible in 9 days on four nutrient types (C, E, G, and J) all of which contained a Bacto agar base underlay with either saline or RPMI overlay. Plates with an RPMI underlay were slightly less effective. Plates containing FCS gave significantly lower plating efficiencies or failed to support growth. Overall plating efficiencies for CANIII were low compared with the other two reference strains (Table 1).

In comparison with 4% CO_2, incubation in sealed plates reduced colony growth, with greater time to first colony appearance for X10/1 and Esm cl2 (Table 1). Plating efficiencies were also significantly higher for plates maintained in 4% CO_2. Highest plating efficiencies were achieved with strain X10/1 (75%, plate type G) followed by Esm cl2 (42%, plate type B) and CANIII (26%, plate type G) (Figs. 1 and 2). In sealed plates (without supplemented CO_2), nutrient types A, B, C, and G supported growth of all reference strains.

In summary, plates consisting of a Bacto agar base underlay fortified with whole blood and with a saline or RPMI overlay sustained most rapid growth and higher plating efficiencies (plate types C and G). Plates incorporating an RPMI base also supported high plating efficiencies (A and B). In contrast, plates with underlays incorporating 10% FCS (D and H) had the fewest colonies, with those containing lysed blood (E, F, I, and J) giving intermediate levels of growth.

To further validate the applicability of the technique for growing a wide range of *T. cruzi* strains, we plated an expanded set of isolates encompassing all the known genetic groups (Table 2). These uncloned isolates have previously been genetically characterised in our laboratory and while none exhibited mixed profiles, the possibility of undetected heterogeneity could not be ruled out. The most sen-
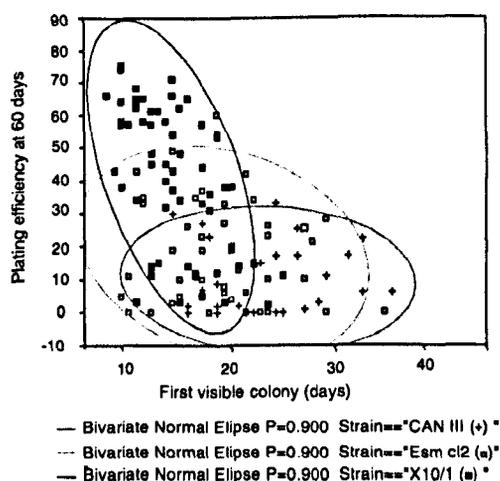

Fig. 1. Bivariate analysis: plating efficiency by first visible colony for the reference strains CAN III (+), Esm cl2 (□) and X10/1 (■). Data are derived from all nutrient plate types (A–J).

sitive nutrient combination, identified previously from the growth of reference strains, was used, namely Difco blood agar base supplemented with 10% defibrinated rabbit blood with a saline overlay (plate type C). Plates were sealed in order to assess the applicability of the technique to situations where CO_2 incubation is unavailable. Some isolates were plated directly from thawed stabilates, while others were plated from axenic cultures in the log growth phase. Without exception, all isolates produced clonal colonies, the development of which could be monitored under a microscope, allowing transfer of clones to liquid media after 13–60 days; *T. rangeli* clones could be harvested after 13 days.

3.2. Trypanosoma cruzi from naturally infected triatomine bugs

A total of 195 triatomine bugs were used in the analysis, 184 from Paraguay, five from Brazil and six from Venezuela. In all cases clones were obtained using plate type C (Table 1).

The 184 triatomine bugs from Paraguay were captured from three regions (Chaco, San Pedro, Paraguari). Triatomine species were *Triatoma sordida* (59), *Triatoma guasayana* (51), and *Triatoma infestans* (74). Microscopy revealed that 29 of the bugs were infected, all *T. infestans* (27 domestic and two from outside houses). *Trypanosoma cruzi* isolates were characterised by a combination of mini exon PCR, 24Sα rRNA PCR and the enzymes GPI, PGM and 6-PGD (Yeo et al., 2005). Sixteen of 29 bugs from the Chaco and San Pedro carried TCIIe; five bugs from the Chaco region and San Pedro carried TCIIb and eight bugs from the Chaco and Paraguari were infected with TCIId. Clonal colonies were obtained from 16 of the 29 bugs and characterisation of colonies revealed that four of the 16 bugs harboured mixed infections of TCIIb with TCIId. All four mixed infections were in domestic *T. infestans*,
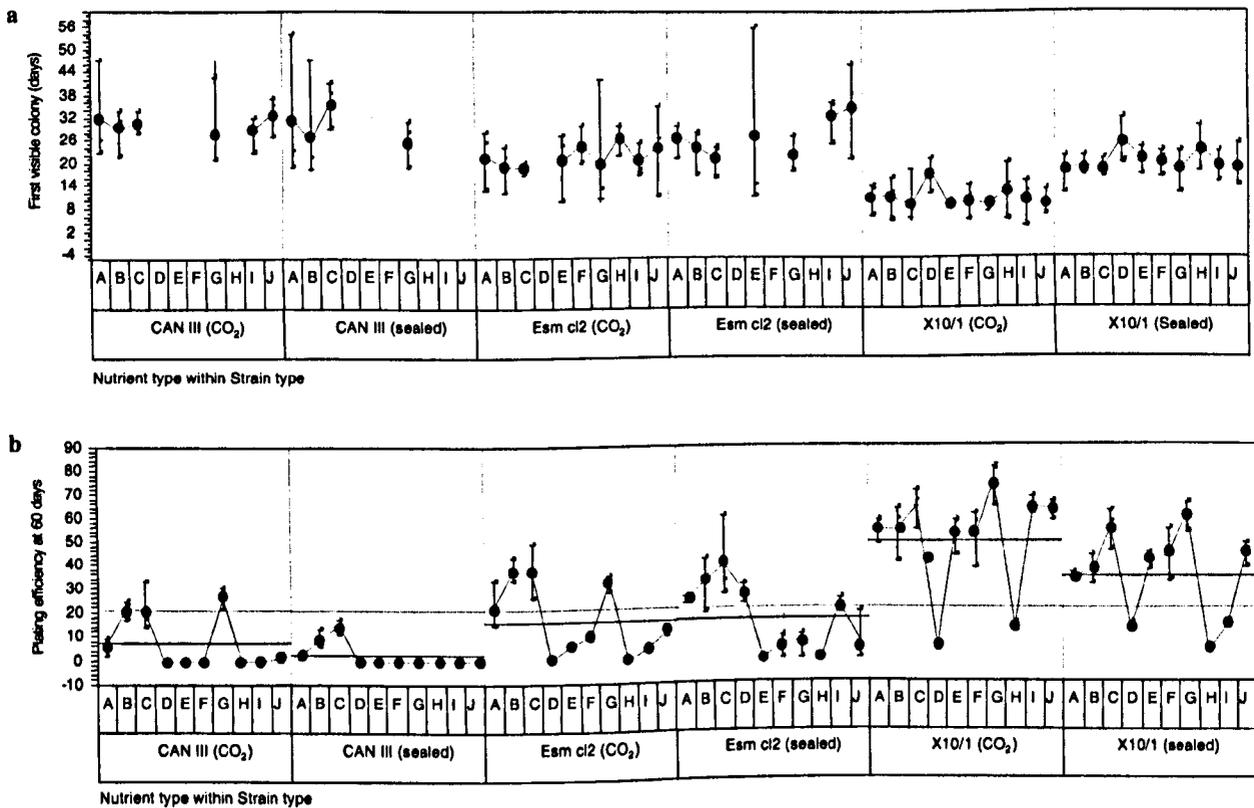
Fig. 2. (a) Variability chart showing first visible colony (days) for each nutrient plate type (A–J), strain and incubation conditions. Vertical black lines on the Y axis join standard deviation points. Columns with no data indicate that the nutrient type did not support growth. (b) Variability chart showing plating efficiency at 60 days for each nutrient plate type (A–J), strain and incubation conditions. The horizontal line represents the group mean. Vertical black lines on the Y axis join standard deviation points.

three from the Chaco and one from San Pedro. A total of 23 clones were isolated from the four triatomines, seven were characterised as TCIIb and 16 characterised as TCIIe suggesting overgrowth of TCIIe in this vector. In support of this, isolates acquired directly after dissection, but prior to cloning, were passaged in liquid culture medium and only gave TCIIe genotypes.

Twenty clonal colonies were obtained from each of five silvatic *Rhodnius robustus*, collected in Pará State, Brazil. The colonies from each of the bugs were characterised via mini-exon PCR and with the enzymes ASAT, PGM, GPI, 6PGD and G6PD (Chapman et al., 1984). All colony profiles corresponded with the TCI reference strain, X10/1. RAPD analysis with seven primers (A1, A2, A4, A6, H1, L4, and L5) confirmed TCI. However, primer A6 showed that different colonies from a single bug had distinct RAPD genotypes: one of 20 colonies had the minority genotype (Fig. 3). On replating a further 20 colonies from cryopreserved bug faeces, three of 20 had the minority genotype. Thus, one of the five *R. robustus* contained a mixed infection of two distinct genotypes.

Evidence of mixed *T. cruzi/T. rangeli* infections was observed in two of six specimens of *Rhodnius prolixus* (four from palm trees, two domestic) from Portuguesa State, Venezuela. In total, 78 clonal colonies were harvested from

four plates. Two bugs (both from the palm tree ecotope) showed mixed infections of *T. cruzi* and *T. rangeli* distinguished by RAPD analysis. Of 12 colonies harvested from one plate, seven were *T. cruzi* and five were *T. rangeli*; eight harvested from the second plate were *T. cruzi* and six were *T. rangeli*. RAPD, mini-exon and 24α rDNA analysis showed the *T. cruzi* clones to be TCI. Overall, 35% of plates became contaminated when plating from triatomine bugs.

3.3. Trypanosomes from infected mice

Single infection with TCIIc and mixed infections of *T. cruzi* TCIIb with TCIIc were established in mice, 21 days prior to plating of blood or plasma (Table 3). Varying volumes of *T. cruzi* infected whole blood or plasma were incorporated into blood agar plates with a saline overlay (plate type C, Table 1). *Trypanosoma cruzi* inocula in more than 100 μl of whole blood impeded subsequent microscopic visualisation of colony growth. However, after the third week the majority of erythrocytes had lost their structural integrity, enabling easier visualisation of colonies. Five of 14 plates showed plating efficiencies (>1%) between 26 and 33 days. Colonies were typically 0.2–1.5 mm in diameter after 35 days, with central masses of amastigotes and
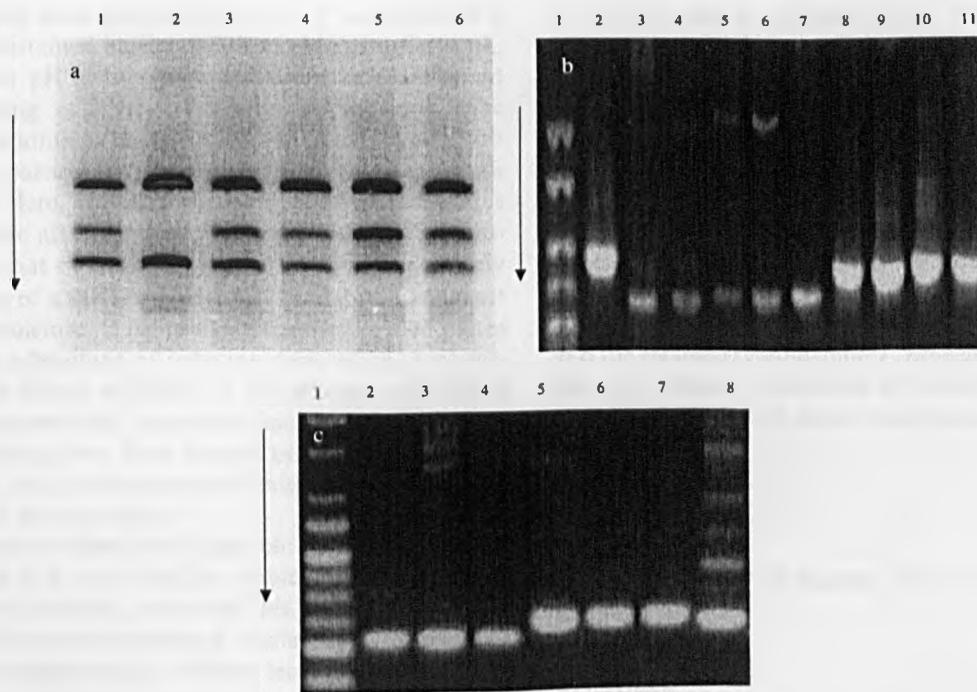
Fig. 3. (a) Random amplification of polymorphic DNA (RAPD) profiles (primer A6), for six clonal colonies, showing a mixed infection in a single triatomine bug. Lane 2 contains the minority genotype. Isoenzyme analysis, mini-exon PCR, and six other RAPD primers indicated a mixed TCI infection (see text body). (b) Mini exon PCR profiles of clonal colonies obtained from a mixed infection in a single mouse: lanes 2 and 3, reference strains TCIIb and TCIIc, respectively; lanes 4–7 clonal colonies of TCIIc (250 bp); lanes 8–11 clonal colonies of TCIIb (300 bp). No single colony had a mixed profile. (c) 24Sα rRNA PCR profile of clonal colonies obtained from a mixed infection in a single mouse: lanes 2 and 8, reference strains TCIIc and TCIIb; lanes 3–4, clonal colonies of TCIIc (110 bp); lanes 5–7, clonal colonies of TCIIb (125 bp). No single colony had a mixed profile.

peripheral epimastigotes. Plates inoculated with less than 25 trypomastigotes showed no viable colonies after 30 days. Although a single colony became visible in each of plates 1–8, they were not viable. Plates prepared with higher numbers of parasites demonstrated higher plating efficiencies (Table 3).

After 38 days, 73% of colonies picked from various plates were successfully expanded in RPMI based liquid medium. Fifteen colonies from each of plates 13 and 14 were characterised by mini-exon and 24Sα rRNA PCR (Fig. 3). Of 15 colonies from plate 13, eight were TCIIb and seven TCIIc. Twelve colonies from plate 14 were TCIIb and three colonies were TCIIc. None of the colonies gave a mixed profile.

4. Discussion

Previous cloning techniques have resulted in plating efficiencies from epimastigote cultures in the region of 40%, with colony isolation possible in 21 days and colonies remaining viable for up to 40 days (Mondragon et al., 1999). Our results demonstrate significant improvements, with viable colony isolation possible in as little as 9 days, plating efficiencies up to 75% and colony viability up to 60 days, dependent on the genotypic group of the *T. cruzi* strain. TCIIa grew on only four of the 10 nutrient combinations both in CO_2 and sealed plates (plate types A, B,

C, and G). Furthermore, the improved method allows recovery of colonies without a CO_2 incubator and it supports growth of *T. rangeli* (plate type C). The fidelity of clones was shown by genotyping colonies.

Large differences were apparent in the time to first colony appearance. The primary influencing factors were *T. cruzi* strain, plate type and incubation conditions. Reference strain TCI showed prolific growth on most plate types, whereas the TCIIa reference strain grew on only four of the 10 nutrient combinations. Such differences in growth kinetics, for strains and clones, have previously been noted during cultivation in liquid media (Finley and Dvorak, 1987; Lauria-Pires et al., 1997). The isolation of discrete colonies thus has obvious benefits if bugs and mammals carry mixed infections differing in growth rate.

Generally the combination of a blood agar base containing 10% defibrinated rabbit blood with a saline or RPMI overlay supported the fastest time to first colony appearance, highest number of colonies and highest plating efficiencies, with many colonies remaining viable after 60 days (plate type C). Similar results were obtained using RPMI as an overlay (G). Plates incorporating RPMI in the underlay and overlay also supported high plating efficiencies (A and B) using three reference strains. Importantly, all strains tested could be cultured in sealed plates, although growth was less rapid. Time required for CO_2 to accumulate in sealed plates, as a by-product of

respiration, may have reduced growth of *T. cruzi*, which is inhibited by sustained basic conditions (Mondragon et al., 1999). A lower pH at time of inoculation therefore might improve plating efficiency. Previous publications have reported desiccation to be a primary factor in reducing colony viability, particularly for slow growing strains, including CANIII. Here, colonies on many plate combinations remained viable after 60 days. Increased viability, in comparison with that of Mondragon et al. (1999) is probably due to our use of a thicker underlay, which formed a larger reservoir of moisture. Despite slower growth, sealed plates also have the advantage of reducing desiccation, accounting for higher colony viabilities in comparison with plates closed with porous tape. Incubator humidification is essential for unsealed plates. Poor growth on plates incorporating FCS may be a consequence of limited haemin, utilized by *T. cruzi* as an iron source.

Lysed blood or lysed centrifuged blood was incorporated into plates as a substitute for whole blood to improve visualisation of colonies. However, this diminished growth of colonies. On plates containing whole blood, colonies of 0.2 mm were readily visible, enabling isolation in as little as 9 days. By microscopy, colonies may be detected and removed much earlier. RPMI underlays, although occasionally less efficient, provide better transparency.

The overall fungal contamination rate was 35% when plating triatomine faeces. The method proved effective at isolating clonal colonies from naturally infected triatomine bugs and infections with mixed genotypes could be resolved. Incorporation of improved anti-fungal agents may reduce the contamination rate. Four of 16 *T. infestans*, one of five *R. robustus* and two of six *R. prolixus* carried mixed infections, the latter co-infections of *T. cruzi* and *T. rangeli*. Thus, culture of *T. cruzi* from triatomines and mammals without plating may distort perceptions of the levels of mixed infections. Indeed, when parental isolates of four *T. infestans* possessing mixed infections were subject to prolonged culture, TCIIe became the dominant *in vitro* genotype. Competitive growth in culture, without cloning, might obscure slow growing strains and give misleading vector and geographic distributions of *T. cruzi* genotypes.

Attempts made to isolate clones from sylvatic mammals in South America were unsuccessful using whole blood or centrifuged blood from seven animals positive by xenodiagnosis (data not shown). Low parasitaemia, of less than four organisms per plate, was an obvious explanation. The triple centrifugation method was inadequate for concentrating trypomastigotes and no haematocrit centrifuge was available. Wittner et al. (1982) observed that <100 organisms per plate resulted in low plating efficiencies. Santos et al. (2000) reported no effect on sensitivity or growth by the addition of preconditioned growth media to agarose plates. Depletion of glucose has been implicated in stimulating differentiation (Tyler and Engman, 2000) but has not been shown to trigger accelerated growth. Similarly, a minimum of 25 trypomastigotes per plate was required

to obtain clonal colonies from mice carrying mixed infections and plating efficiencies (maximum 24%) were lower than when using other sources of *T. cruzi*. Epimastigotes are susceptible to lysis by normal or immune sera in a complement-dependent reaction (Fernandez-Presas et al., 2001) and complement in the mouse blood may cause epimastigote death during differentiation.

This study demonstrates that clonal isolation of multiple infections from single vectors and mammal hosts is feasible. Importantly, the method described enables clonal isolation of all known *T. cruzi* genotypes (TCI and TCIIa–e) and the related trypanosome *T. rangeli* (using plate type C), allowing efficient production of clones from varied sources and the resolution of mixed infections, essential for downstream analysis.

References

Bosseno, M.F., Telleria, J., Vargas, F., Yaksic, N., Noireau, F., Morin, A., Breniere, S.F., 1996. *Trypanosoma cruzi*: study of the distribution of two widespread clonal genotypes in Bolivian *Triatoma infestans* vectors shows a high frequency of mixed infections. Exp. Parasitol. 83, 275–282.

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.

Carrasco, H.J., Frame, I.A., Valente, S.A., Miles, M.A., 1996. Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi*. Am. J. Trop. Med. Hyg. 54, 418–424.

Carruthers, V.B., Cross, G.A., 1992. High-efficiency clonal growth of bloodstream- and insect-form *Trypanosoma brucei* on agarose plates. Proc. Natl. Acad. Sci. USA 89, 8818–8821.

Chapman, M.D., Baggaley, R.C., Godfrey-Fausset, P.F., Malpas, T.J., White, G., Canese, J., Miles, M.A., 1984. *Trypanosoma cruzi* from the Paraguayan Chaco: isoenzyme profiles of strains isolated at Makthlawaiya. J. Protozool. 31, 482–486.

Fernandez-Presas, A.M., Zavala, J.T., Fauser, I.B., Merchant, M.T., Guerrero, L.R., Willms, K., 2001. Ultrastructural damage of *Trypanosoma cruzi* epimastigotes exposed to decomplemented immune sera. Parasitol. Res. 87, 619–625.

Finley, R.W., Dvorak, J.A., 1987. *Trypanosoma cruzi*: analysis of the population dynamics of heterogeneous mixtures. J. Protozool. 34, 409–415.

Gambarelli, F., Dumon, H., 1988. A simple method for cloning *Leishmania* spp.. Ann. Parasitol. Hum. Comp. 63, 160–162.

Gaunt, M.W., Yeo, M., Frame, I.A., Stothard, J.R., Carrasco, H.J., Taylor, M.C., Mena, S.S., Veazey, P., Miles, G.A., Acosta, N., de Arias, A.R., Miles, M.A., 2003. Mechanism of genetic exchange in American trypanosomes. Nature 421, 936–939.

Gillin, F.D., Diamond, L.S., 1978. Clonal growth of *Entamoeba histolytica* and other species of *Entamoeba* in agar. J. Protozool. 25, 539–543.

Goldberg, S.S., Chiari, E., 1980. Growth and isolation of single colonies of *Trypanosoma cruzi* on solid medium. J. Parasitol. 66, 677–679.

Gomes, M.L., Araujo, S.M., Chiari, E., 1991. *Trypanosoma cruzi*: growth of clones on solid medium using culture and blood forms. Mem. Inst. Oswaldo. Cruz 86, 131–132.