RESEARCH ARTICLE

Immigration and establishment of *Trypanosoma cruzi* in Arequipa, Peru

Alexander S. F. Berry[1,2], Renzo Salazar-Sánchez[3], Ricardo Castillo-Neyra[3,4], Katty Borrini-Mayorí[3], Claudia Chipana-Ramos[3], Melina Vargas-Maquera[3], Jenny Ancca-Juarez[3], César Náquira-Velarde[3], Michael Z. Levy[3,4], Dustin Brisson[2]*, for the Chagas Disease Working Group in Arequipa[¶]

1 Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital of Philadelphia, Philadelphia, PA, United States of America, 2 Department of Biology, University of Pennsylvania, Philadelphia, PA, United States of America, 3 Universidad Peruana Cayetano Heredia/University of Pennsylvania Zoonotic Disease Research Laboratory, Arequipa, Peru, 4 Department of Biostatistics, Epidemiology and Informatics, The Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA, United States of America

¶ Membership of the Chagas Disease Working Group in Arequipa is provided in the Acknowledgments.
* dbrisson@sas.upenn.edu

Abstract

Changing environmental conditions, including those caused by human activities, reshape biological communities through both loss of native species and establishment of non-native species in the altered habitats. Dynamic interactions with the abiotic environment impact both immigration and initial establishment of non-native species into these altered habitats. The repeated emergence of disease systems in urban areas worldwide highlights the importance of understanding how dynamic migratory processes affect the current and future distribution and abundance of pathogens in urban environments. In this study, we examine the pattern of invasion of *Trypanosoma cruzi*—the causative agent of human Chagas disease—in the city of Arequipa, Peru. Phylogenetic analyses of 136 *T. cruzi* isolates from Arequipa and other South American locations suggest that only one *T. cruzi* lineage established a population in Arequipa as all *T. cruzi* isolated from vectors in Arequipa form a recent monophyletic group within the broader South American phylogeny. We discuss several hypotheses that may explain the limited number of established *T. cruzi* lineages despite multiple introductions of the parasite.

Introduction

Habitat alterations are transforming biological communities worldwide [1–3]. The current and future geographic distributions of many species in disturbed environments depends upon their interactions with novel biotic and abiotic features during immigration and while establishing a growing population [4]. Although many species fail to establish thriving populations in altered habitats, others are well-suited to migrate to, and prosper in, these novel environments. For example, several populations of plant [5–7], insect [8–11], mammal [12,13], and bird [14,15] species are severely diminished or suffer local extinctions in recently urbanized

environments [1], while several microbial species benefit from the abundance of humans and human-associated hosts or vectors in similar habitats [16]. Although conservation efforts have focused primarily on the impacts of environmental changes on native plant and animal species, establishment or population growth of disease-causing microbial populations can have a strong negative impact on populations of native flora and fauna [17] in addition to their impact on human health and economy [18]. The rate or impacts of invasions of infectious microbes may be mitigated through public health programs based on an understanding of the dynamic processes determining immigration and establishment rates. The regularity at which disease systems are emerging in many urban and urbanizing areas underscores the importance of understanding how disease-causing microbes migrate to, and establish in, urban environments [16], one of the most dramatic examples of habitat alteration [19,20]. In this study, we examine the patterns that govern how *Trypanosoma cruzi*—the causative agent of Chagas disease in humans—invaded and colonized the city of Arequipa, Peru.

Invasion of a new environment by a pathogen occurs in three stages: (1) immigration, or the movement of an individual to the new environment; (2) establishment of a population via reproduction and population growth; and (3) local dispersal [4]. The many studies focusing on the emergence of disease systems have generated a wealth of knowledge concerning factors affecting population growth [21,22] and considerable progress in understanding local dispersal [23]. For example, prior studies concluded that human-created containers increase the abundance of standing water that provide breeding habitats for the mosquitos that spread dengue virus [24]. Relatively few studies, by contrast, have investigated the early stages of invasion, the immigration and establishment processes, due to the practical difficulties of collecting the necessary data in the early stages of establishment of a novel species.

The Chagas disease system in Arequipa, Peru, provides an ideal system in which to study the early invasion processes in urban environments. Arequipa has experienced rapid urbanization and human population growth in the previous half century [25]. The expansion in population sizes and geographic ranges of humans and their domestic animals provide *T. cruzi* and its only known insect vector in Arequipa, *Triatoma infestans*, with abundant and accessible hosts [26–31]. Estimates suggest that *T. cruzi* infection prevalence among humans in Arequipa exceeded 5% [27,29–30]. The population history of the *T. cruzi* currently in Arequipa—including the geographic locations of the migrants that established the current population, the rate at which migrants enter and establish in the area, and the age of each established lineage—have not been investigated. Here, we performed phylogenetic analyses of maxicircle DNA, a non-recombining circular element analogous to mitochondrial DNA, to estimate the number of independent *T. cruzi* lineages established in Arequipa and to estimate the timing of each establishment event. We assessed whether extant *T. cruzi* in Arequipa form a single monophyletic group, indicative of the establishment of a single migrant lineage, or multiple diverse clades, indicative of multiple independent immigration and establishment events.

Results

The maxicircle sequence is ideal for population genetic and phylogenetic analyses because it is conserved among diverse *T. cruzi* lineages and is non-recombining [32], unlike the nuclear genome which undergoes regular recombination [33]. Further, phylogenies based on multiple combinations of *T. cruzi* maxicircle and nuclear genes yield largely similar topologies [34]. While there was substantial maxicircle sequence diversity among samples across South American isolates, almost no diversity was observed among the samples from Arequipa. Over 13% (2055/15367bp) of the maxicircle sites were polymorphic among all 136 samples while only 16 sites were polymorphic (0.1%) among the 123 samples collected within Arequipa (Table 1).

Table 1. Population genetic statistics.

|	Average Pairwise Distance	π	θ	Segregating Sites
All samples (N = 136)	126	$8.18^{*}10^{-3}$	$2.44^{*}10^{-2}$	2055
Arequipa (N = 123)	1.04	$6.80^{*}10^{-5}$	$1.93^{*}10^{-4}$	16
South America (N = 13)	728	$4.73^{*}10^{-2}$	$4.24^{*}10^{-2}$	2022
La Esperanza (N = 3)	643	$4.18^{*}10^{-2}$	$4.18^{*}10^{-2}$	964

Similarly, estimates of diversity among the samples from Arequipa derived from population genetic statistics were substantially lower than the total diversity across all samples (π = $6.8^{*}10^{-5}$ vs $8.18^{*}10^{-3}$; θ = $1.93^{*}10^{-4}$ vs $2.44^{*}10^{-2}$; average pairwise distance 1.04 vs 126). In contrast to the limited genetic diversity within Arequipa, other locales from which multiple isolates were sampled contain substantial genetic diversity among limited samples (N<3; Fig 1).

The monophyletic group containing all 123 samples derived from *T. infestans* and domestic animals collected in Arequipa coalesce in the very recent past, despite collection sites extending throughout Arequipa and surrounding towns across 7 years (Fig 1). The extremely low genetic diversity among these 123 samples provided insufficient information to resolve any phylogenetic relationships within this monophyletic group. The estimated divergence time suggests that the common ancestor of Arequipan *T. cruzi* lived 8.8-19kya (S1 Fig). However, care should be taken when interpreting divergence time estimates because of the low sequence diversity within the Arequipan clade and because of the uncertainty in divergence time at the root of the phylogeny. One sample derived from an infected human in Arequipa (Fig 1C) belongs to a lineage that is more closely related to samples collected in Rio Grande, Brazil than to the other samples collected in Arequipa (Fig 1). This sample is distinct from all other samples in Arequipa suggesting that *T. cruzi* can immigrate to Arequipa but may not establish in the vector population. The population size of an unsampled *T. cruzi* lineage–if present in the *T. infestans* population–must be at least 42 times smaller than the dominant population in Arequipa to have remained undetected by chance (p<0.05). A maximum likelihood phylogeny that included 18 additional partial maxicircle sequences obtained from NCBI (S1 Table) supports the result that all *T. cruzi* isolates collected in Arequipa form a monophyletic group derived from a recent common ancestor (S2 Fig).

In contrast to the monophyletic ancestry found in Arequipa, genetic diversity was apparent in the samples collected from other regions, despite limited numbers of samples (N<3). Population genetic diversity among the three samples collected La Esperanza, a town of 57 houses in the Cutervo Province of Cajamarca, Peru (π = $4.18^{*}10^{-2}$; θ = $4.18^{*}10^{-2}$), are much larger than those in Arequipa despite the limited number of samples [35]. The genetic relatedness and geographic distance among South American isolates are not correlated (Mantel test among DTU I: r = 0.79, p = 0.083, Mantel test among DTU VI: r = -0.5, p = 0.833; Linear regression, r = 0.22, p = 0.534; Table 2; S3 Fig). For example, while proximal towns La Esperanza and Campo Florido, Peru have closely related *T. cruzi*, isolates from cities around Brazil encompass nearly the total genetic diversity.

Discussion

The invasion of recently altered environments by non-native species impacts the population health of native species as well as human health and economy. Investigation of the dynamic process of immigration and establishment of non-native species into these disturbed habitats has the potential to mitigate the impacts of pests and pathogens that are detrimental to human
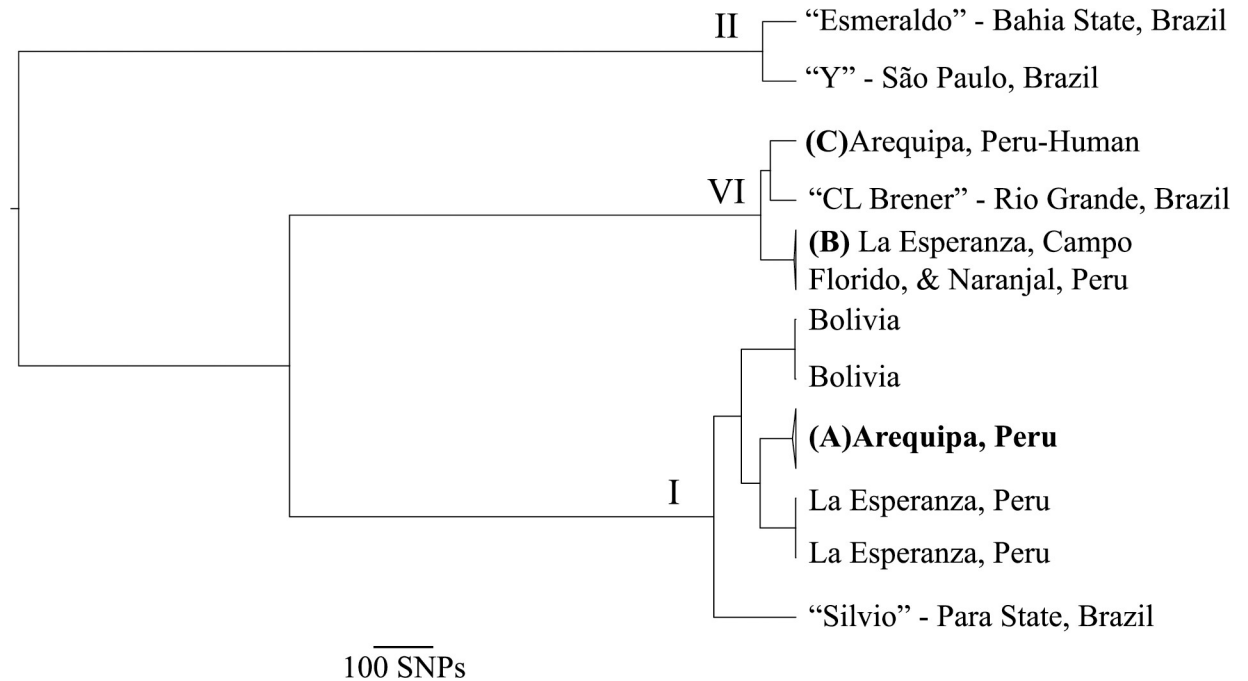
Fig 1. The extant *T. cruzi* population in Arequipa arose from a single, recent introduction. Maximum clade credibility (MCC) tree shows that **(A)** all 123 *T. cruzi* isolated from bugs and domestic mammals in Arequipa form a monophyletic group with a single, recent common ancestor, indicative of a single immigration event in the recent past. Despite substantial genetic diversity among *T. cruzi* throughout South America, those collected in Arequipa show little diversity. **(B)** Three samples collected in Campo Florido and Naranjal, Peru and one sample from La Esperanza, Peru have nearly identical maxicircle sequences and form a monophyletic group. La Esperanza, Peru contains at least two distinct *T. cruzi* lineages, suggesting multiple independent introductions. **(C)** The only *T. cruzi* sample isolated from a human in Arequipa is distinct from all other samples from Arequipa, suggesting that this introduction has not established in the city. All tips represent a single sample except (A) (N = 123) and (B) (N = 4). All nodes have strong support (posterior probability≥0.99). Nodes are collapsed when the samples contained have nearly identical maxicircle sequences. Roman numerals signify DTU.

https://doi.org/10.1371/journal.pone.0221678.g001

Table 2. Distance matrix showing average pairwise SNP distance between samples (bottom triangle) and Euclidean distance between sample collection locations (top triangle).

	(A) Arequipa, Peru	La Esperanza, Peru	(B) La Esperanza, Campo Florido, & Naranjal, Peru	Bolivia	(C) Arequipa, Peru—Human	São Paulo, Brazil	Para State, Brazil	Bahia State, Brazil	Rio Grande, Brazil
(A) Arequipa, Peru		1400km	1400km	700km	0km	2700km	2500km	3200km	2500km
La Esperanza, Peru	95		60km	2000km	1400km	4000km	3000km	4100km	3900km
(B) La Esperanza, Campo Florido, & Naranjal, Peru	970	966		2000km	1400km	4000km	3000km	4100km	3900km
Bolivia	137	137	957		700km	2000km	1800km	2500km	2000km
(C) Arequipa, Peru—Human	975	971	89	967		2700km	2500km	3200km	2500km
São Paulo, Brazil	1044	1045	1033	1045	1058		2100km	1300km	1100km
Para State, Brazil	216	216	984	205	992	1059		1500km	2800km
Bahia State, Brazil	1284	1289	1234	1287	1258	68	1308		2500km
Rio Grande, Brazil	1005	1003	82	1001	77	1052	1026	1284	

(A) 123 samples isolated from bugs, dogs, and guinea pigs in Arequipa are represented here.

(B) 4 samples collected in Campo Florido, Naranjal, and La Esperanza are represented here. Euclidean distances (top triangle) are displayed for Santa Cruz.

(C) The only sample isolated from a human in Arequipa is represented here.

https://doi.org/10.1371/journal.pone.0221678.t002

populations as well as agricultural and native species [17]. The analyses presented here suggest that the population of *T. cruzi* in Arequipa, Peru, descended from a recent invasion consisting of only one evolutionary lineage. This conclusion is supported by the extremely limited genetic diversity observed among *T. cruzi* isolates sampled within and around the city, in contrast with considerable genetic diversity observed regionally and at other locales (Table 1; Fig 1). Several non-exclusive hypotheses that can be experimentally tested in the future may explain these results, including a low immigration rate, that immigration is common but immigrants rarely establish populations as a result of the low transmission rate between hosts and vectors, or that there is a high turnover rate among *T. cruzi* lineages.

Successful invasion of a novel geographic area is a function of the rate of immigration, the temporal duration that a habitat has been suitable for establishment, and the probability that an immigrant can reproduce and establish a population. The influx of humans and associated products and domesticated animals into Arequipa over the last ~60 years due to rapid urbanization and economic growth [25,36] has provided many opportunities for *T. cruzi* immigration. However, many migrants to Arequipa come from areas in which they were not exposed to *T. cruzi*, such as the neighboring regions of Puno and parts of Cusco which are beyond the range of *T. infestans* [37]. Nevertheless, one *T. cruzi* lineage that does not appear to be circulating within the *T. infestans* population was detected in an infected human (Fig 1), suggesting that *T. cruzi* immigration through human movement can occur. These data suggest that multiple *T. cruzi* lineages may have immigrated to Arequipa with all but one failing to transmit sufficiently to establish a population. Future studies will be necessary to identify the source, rate, and potential mechanism of *T. cruzi* immigration, independent of establishment probability, through analyses of the genomic diversity in human infections.

Prior studies across multiple species suggest that the majority of immigrants in most species that reach a novel geographic area fail to establish a population due to both inhospitable local environmental conditions [1,38] and stochasticity [39,40]. Environmental factors that can reduce establishment probabilities include unfavorable abiotic conditions, limited food resources or vectors, or an abundance of predators or competitors. The establishment probability of immigrant *T. cruzi* in an urban environment is likely depressed by a low contact rate between infected humans and vectors [41–44]. For example, many human immigrants moved to locations in the city without established *T. infestans* populations [25,37], which would result in few opportunities for *T. cruzi* transmission from infected immigrant humans to insect vectors, thus curtailing establishment probabilities. Further, the probability of an infected *T. infestans* transmitting *T. cruzi* to novel hosts is low [42,45,46], thus reducing the probability that a recently-immigrated *T. cruzi* lineage will establish a population. Both a low probability of establishing a population due to limited contact between *T. infestans* and infected immigrant humans or due to the limited probability of transmission from vectors to host are consistent with the observation of only a single established *T. cruzi* lineage in vectors in Arequipa.

The observed establishment probability is likely independent of competition among *T. cruzi* lineages. Competitive exclusion—where an existing population prevents the invasion of new immigrants—appears unlikely as the majority of city blocks do not contain *T. cruzi* [25,47] despite substantial vector populations [48]. Under the competitive exclusion hypothesis, one might expect different *T. cruzi* lineages establishing in different areas of the city. Indeed, multiple *T. cruzi* lineages do co-circulate within the same locality [44,49–53], as seen in the samples sequenced here from La Esperanza (Fig 1), and even within the same host [54,55], suggesting that competition is not preventing the establishment of multiple lineages in Arequipa.

The limited genetic diversity within Arequipa could result from a recent replacement of a previously dominant lineage through natural population processes. While the continuous

substitution of a dominant strain through natural selection or drift is common in well-mixed populations, geographic structure within populations tends to result in the persistence of genetic diversity [56]. The absence of samples deriving from a previously established *T. cruzi* lineage in the fragmented urban and inter-district landscapes, much of which contains an active vector population but no *T. cruzi* [25,47,48], is suggestive that no previous lineages dominated this area. Passaging *T. cruzi* through animal hosts in order to isolate *T. cruzi* from infected bugs presents a potential for selection bias against specific genotypes. However, no other *T. cruzi* lineages were detected circulating in the vector population, despite the spatial and temporal range of samples in our dataset (7 years), suggesting that selection bias does not occur unless both guinea pigs and mice completely exclude the same genotypes.

In conclusion, all relevant data suggest that the *T. cruzi* circulating in vector populations prior to the recently-controlled epidemic in the city of Arequipa descended from a single immigrant lineage. While the ancestral lineage that gave rise to all extant *T. cruzi* in Arequipa did not necessarily reside in Arequipa, the data suggest that the common ancestor of all analyzed isolates immigrated recently and founded a population. The single divergent lineage found in a human patient suggests that *T. cruzi* may regularly immigrate to the city but that immigrants rarely establish populations.

Methods

Ethics statement

The Institutional Animal Care and Use Committee (IACUC) of Universidad Peruana Cayetano Heredia reviewed and approved the animal-handling protocol used for this study (identification number 59605). The Universidad Peruana Cayetano Heredia Ethics Committee approved human subject research and approved the collection of bugs from households (identification number 52186). Human subject research participants provided written consent for the collection of bugs from his or her household and for human subject research. The IACUC of Universidad Peruana Cayetano Heredia is registered in the National Institutes of Health at the United States of America with PHS Approved Animal Welfare Assurance Number A5146-01 and adheres to the Animal Welfare Act of 1990 [57].
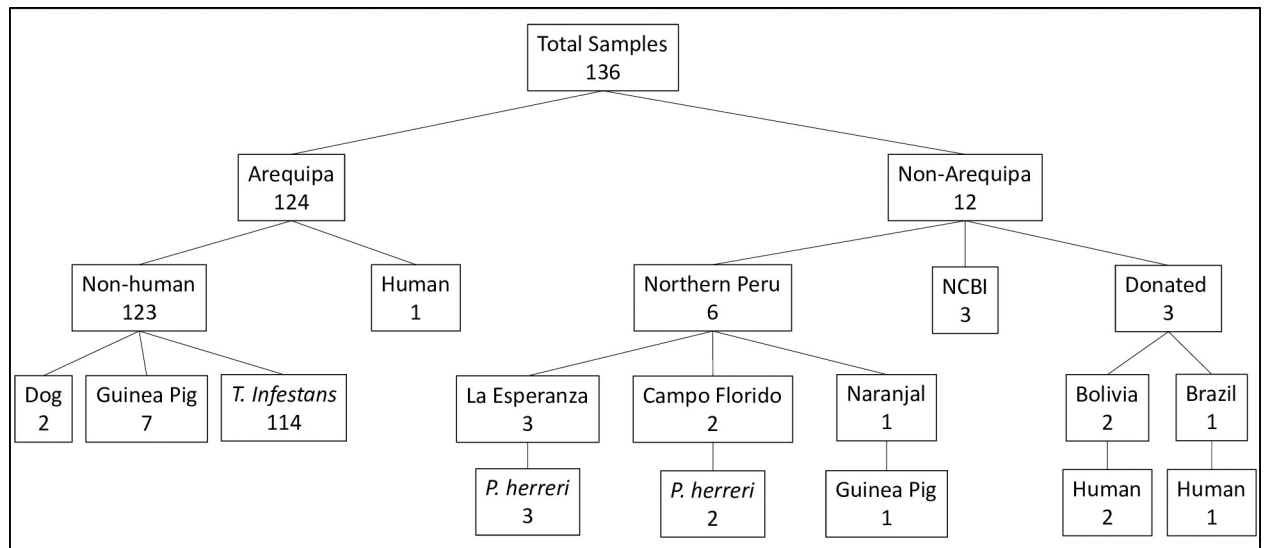
Sample collection

DNA from 133 *T. cruzi* isolates were analyzed to determine phylogenetic relationships (Fig 2; Fig 3; S1 Fig). The majority of samples were isolated from *T. infestans* bugs collected from houses throughout Arequipa (N = 114). *T. cruzi* isolated from bugs provide the most unbiased sample because there is no host selection for a specific genotype as *T. cruzi* from all hosts must pass through this vector in order to infect a new host. Additionally, there is only one vector species in Arequipa and thus there is no vector-specific selection bias. Samples were collected from houses throughout 11 districts in Arequipa from 2008 to 2015, encompassing the majority of the parasite's range within Arequipa. The subset of isolates used in this study (S2 Table) were chosen to maximize the chance of detecting diversity 1) by choosing isolates derived from bugs collected in different houses each year in all but six cases (15 isolates from 6 houses, 1 derived from all other houses) in order to minimize the chance of isolating sibling parasites and 2) by using samples that are spatially and temporally distributed to minimize sampling closely related isolates. Samples were collected as the epidemic was expanding and continued as it was controlled. This sampling scheme represents the optimal chance of detecting minor lineages and capturing diversity within the major lineage. Three of these samples were obtained using xenodiagnosis as described in Chiari & Galvão (1997) [58]. An additional ten samples from Arequipa were isolated from the blood of guinea pigs (N = 7), dogs (N = 2), and
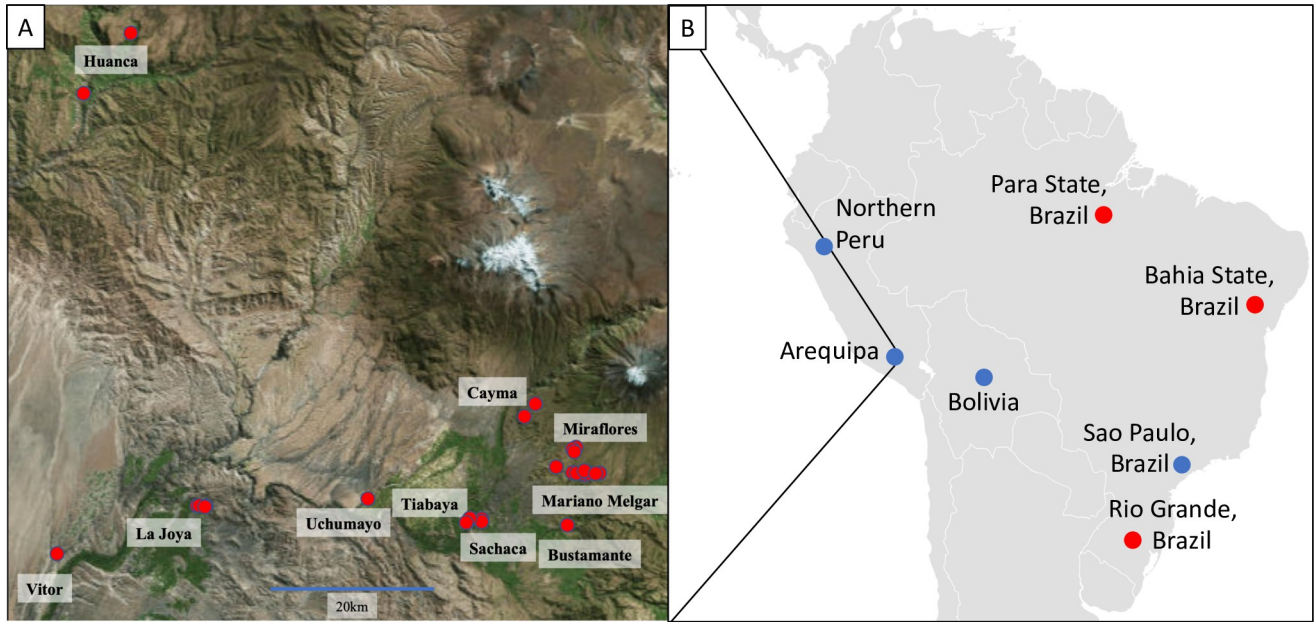
Fig 2. Spatial distribution of samples collected in (A) Arequipa, Peru and (B) South America. (A) The names of ten districts from which *T. cruzi* samples were collected are labeled. Houses from which isolates were collected are represented by red dots. Map of Arequipa was obtained from https://landlook.usgs.gov and QGIS v. 2.18 [75] was used to plot sample locations. (**B**) The sites where isolates were collected are represented by blue dots. Neighboring towns of La Esperanza, Campo Florido, and Naranjal are represented by a single blue dot labeled "Northern Peru". Sequences obtained from NCBI database are represented by red dots. Map of South America was modified from https://commons.wikimedia.org/wiki/Atlas_of_the_world. Figure was modified from Berry *et al.* (2019) [33].

https://doi.org/10.1371/journal.pone.0221678.g002

a human (N = 1). Six isolates were derived from *Panstrongylus lignarius* (N = 5)–also known as *P. herreri* [59]–and one guinea pig (N = 1) collected in the small towns of La Esperanza, Campo Florido, and Naranjal in northern Peru [35]. Cultures of three previously established strains isolated from humans in Bolivia (Bol-SH001 and Bol-DH29) and São Paulo, Brazil

Fig 3. Number of samples collected from each host species per location. Most samples (N = 124) were collected in Arequipa and 6 from small towns in northern Peru. 3 isolates were provided by the Infectious Diseases Research Laboratory at Universidad Peruana Cayetano Heredia. Maxicircle sequences for 3 *T. cruzi* lineages were downloaded from the NCBI database.

https://doi.org/10.1371/journal.pone.0221678.g003

(TC-y) were provided by the Infectious Diseases Research Laboratory at Universidad Peruana Cayetano Heredia (Fig 3).

T. cruzi was isolated from vertebrates (N = 8) using an adaptation of an artificial feeding system that was originally described in Harington (1960) [60]. Briefly, each blood sample was collected with citrate-phosphate-dextrose, transferred into a small plastic jar, and covered with a latex membrane fitted tightly with a rubber band. The jars were placed into an incubator and gradually heated to 35° C. Once the temperature was reached, the jars were inverted to allow uninfected *T. infestans* to feed through the membrane for 15 minutes. *T. cruzi* from the eight laboratory-infected *T. infestans*, all 114 naturally-infected *T. infestans*, and five naturally-infected *P. lignarius* were passaged through guinea pigs or mice in order to avoid isolating other microbes present in the vector, as described in Castillo-Neyra *et al.* (2016) [61]. Feces from infected vectors were injected into guinea pigs or mice and *T. cruzi* was isolated from the blood of each experimentally-infected mammal. *T. cruzi* were directly isolated in LIT culture media from the blood samples of three naturally-infected guinea pigs collected in Arequipa without passage through *T. infestans*.

Reference sequences of three *T. cruzi* isolates obtained from NCBI database were used in subsequent analyses: Silvio, isolated from a human in Para State, Brazil; Esmeraldo, isolated from a human in Bahia State, Brazil; and CL Brener, isolated from a human in Rio Grande, Brazil [62] (Fig 3).

Sequencing

DNA from all laboratory cultures was extracted using Qiagen DNEasy DNA Purification Kit. 150bp single-end read libraries were prepared using TruSeq Nano kit and sequenced to an average depth of >50X using Illumina's NextSeq500. Six *T. cruzi* isolates were prepared in duplicate, and one in triplicate, to allow estimation of sequencing error. Low quality bases were trimmed from raw reads using trimmomatic-0.32 [63].

Sequence assembly

Bowtie2 [64] was used to assemble maxicircle sequences to the most closely related reference sequence, Silvio (gi|225217165|gb|FJ203996.1), obtained from NCBI [62]. Duplicate reads were removed from the assembly using Picard's MarkDuplicates functionality [65]. The assembly had an average depth of >600X across all maxicircles. Maxicircle consensus sequences were determined using VarScan [66], ensuring highly-confident base calls by requiring a 60% match to call each SNP.

Maxicircle alignment

The *T. cruzi* maxicircle contains several conserved genes spanning more than 15kb. All assembled maxicircle sequences and the reference were aligned to the Silvio partial maxicircle sequence (gi|225217165|gb|FJ203996.1|), Esmeraldo strain complete maxicircle (gi|85718082|gb|DQ343646.1), and the CL Brener complete maxicircle (gi|85718081|gb|DQ343645.1) downloaded from the NCBI database. The sequences were aligned using MUSCLE as implemented in MEGA7 [67]. The ends were trimmed so that all sequences started and ended on the same nucleotide, resulting in a final alignment of 15357bp.

Phylogenetic analyses

The non-recombining *T. cruzi* maxicircle is ideal for coalescent analyses because such analyses are based on the assumption that there is no recombination. Phylogenetic analyses of the

15357bp maxicircle sequence from all samples and reference strains were performed using BEAST 1.8.4 [68]. Phylogenetic analyses assumed a model of sequence evolution in which the rates of A→T, C→G, and G→T are equal (123343) with γ-distributed rate heterogeneity. An Extended Bayesian Skyline tree prior [69] with constant evolutionary rates across lineages (strict clock) was chosen based on BEAST Model Test implemented in BEAST2 [70]. Starting with a UPGMA tree and running one Markov chain Monte-Carlo chain for each of five independent runs of 20 million iterations sampling every 2000 iterations ensured sufficient mixing after a 10% burn-in (ESS values >200 in Tracer v1.6.0) [71]. Tree files were combined using LogCombiner1.8.4, excluding a 10% burn-in for each. A Maximum Clade Credibility tree was generated from the combined tree file using TreeAnnotator 1.8.4 and FigTree v1.4.2 was used to visualize tree files (available at http://beast.bio.ed.ac.uk). Phylogenetic analyses were performed using the BEAGLE library to increase computation speed [72,73].

Eighteen partial *T. cruzi* maxicircle sequences (>4.5kb) available on NCBI (S1 Table) were aligned to the samples sequenced here and the reference sequences. Gaps were removed, yielding a final alignment of 2551bp. These eighteen sequences were chosen because they provided the longest overlapping regions of the maxicircle and provide sufficient information for phylogenetic analyses. A Maximum Likelihood estimation was performed using MEGA7 with a Tamura-Nei model and 1000 bootstrap replicates (S2 Fig).

Statistical Analyses

Metrics of population genetic variation, π and θ, were calculated using MEGA7 [67]. Assuming the sample from 123 infected *T. infestans* is representative of the *T. cruzi* population in vectors in Arequipa, the probability that a distinct lineage could be co-circulating but not detected by chance can be calculated using a binomial distribution. The minor lineage must constitute less than 2.41% of the total population in order for there to be a statistically significant chance that a distinct lineage was not detected in any of 123 sampled *T. cruzi*. Mantel tests and linear regression were used to test for a correlation between geographic and genetic distance. Mantel tests were performed using the *vegan* package in R [74]. Two Mantel tests were performed: one testing for correlation among samples belonging to DTU I and another among samples belonging to DTU VI. Statistics were based on Spearman's correlation using 100,000 permutations. A linear regression based on Spearman's correlation was performed using the *rcorr* function in R (S3 Fig). Linear regression was performed on all pairwise distances between samples belonging to the same DTU.

Supporting information

S1 Fig. Most recent common ancestor to extant *T. cruzi* population in Arequipa lived as recently as 8.8kya. Estimates for divergence timing are displayed as 95% confidence intervals at each node. Estimates are based on rooting the phylogeny at 3mya. Sample collection locations are shown at the tips. The DTU each sample belongs to is labeled with its corresponding Roman Numeral in superscript. Due to the low diversity in maxicircle sequence among Arequipan isolates and the uncertainty in dating the tree root, care should be taken when interpreting divergence time estimates.
(TIF)

S2 Fig. The 123 Arequipan isolates form a monophyletic group even after adding all 18 DTU I partial maxicircle sequences greater than 4.5 kb obtained from NCBI. Short maxicircle sequences (<4.5kb) were not included in the analyses due to limited overlap with each other and the limited number of informative SNPs. The total alignment length was 2551 base

pairs. Clades and branches that include samples sequenced here (Fig 1) are in bold font. Bootstrap support greater than 0.8 is labeled on nodes.
(TIF)

S3 Fig. Euclidean distance is not correlated with genetic distance between maxicircles.
Only within-DTU distance data was used to perform the linear regression (red line).
(TIF)

S1 Table. 18 additional partial maxicircle sequences obtained from NCBI.
(DOCX)

S2 Table. Sample collection data.
(DOCX)

Author Contributions

Conceptualization: Alexander S. F. Berry, Dustin Brisson.

Data curation: Alexander S. F. Berry.

Formal analysis: Alexander S. F. Berry.

Funding acquisition: Michael Z. Levy, Dustin Brisson.

Investigation: Alexander S. F. Berry.

Methodology: Alexander S. F. Berry, Renzo Salazar-Sánchez, Katty Borrini-Mayorí, Claudia Chipana-Ramos, Melina Vargas-Maquera, Jenny Ancca-Juarez, Michael Z. Levy, Dustin Brisson.

Project administration: Dustin Brisson.

Resources: Ricardo Castillo-Neyra, Dustin Brisson.

Software: Alexander S. F. Berry.

Supervision: Renzo Salazar-Sánchez, Ricardo Castillo-Neyra, César Náquira-Velarde, Michael Z. Levy, Dustin Brisson.

Validation: Alexander S. F. Berry.

Visualization: Alexander S. F. Berry.

Writing – original draft: Alexander S. F. Berry.

Writing – review & editing: Alexander S. F. Berry, Renzo Salazar-Sánchez, Ricardo Castillo-Neyra, Katty Borrini-Mayorí, Claudia Chipana-Ramos, Melina Vargas-Maquera, Jenny Ancca-Juarez, César Náquira-Velarde, Michael Z. Levy, Dustin Brisson.

References

1. McKinney ML. Urbanization, biodiversity, and conservation. BioScience. 2002; 52: 883–890.

2. Foley JA, DeFries R, Asner GP, Barford C, Bonan G, Carpenter SR, et al. Global consequences of land use. Science. American Association for the Advancement of Science; 2005; 309: 570–574.

3. Seto KC, Güneralp B, Hutyra LR. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. Proceedings of the National Academy of Sciences. National Academy Sciences; 2012; 109: 16083–16088.

4. Theoharides KA, Dukes JS. Plant invasion across space and time: factors affecting nonindigenous species success during four stages of invasion. New Phytol. 2007; 176: 256–273. https://doi.org/10.1111/j.1469-8137.2007.02207.x PMID: 17822399

5. Kowarik I. On the Role of Alien Species in Urban Flora and Vegetation. Plant Invasions-General Aspects and Special Problems. 1995;: 85–103.

6. Hahs AK, McDonnell MJ, McCarthy MA, Vesk PA, Corlett RT, Norton BA, et al. A global synthesis of plant extinction rates in urban areas. Ecol Lett. Wiley Online Library; 2009; 12: 1165–1173. https://doi.org/10.1111/j.1461-0248.2009.01372.x PMID: 19723284

7. Aronson MF, La Sorte FA, Nilon CH, Katti M, Goddard MA, Lepczyk CA, et al. A global analysis of the impacts of urbanization on bird and plant diversity reveals key anthropogenic drivers. Proc R Soc B. The Royal Society; 2014; 281: 20133330. https://doi.org/10.1098/rspb.2013.3330 PMID: 24523278

8. Denys C, Schmidt H. Insect communities on experimental mugwort (*Artemisia vulgaris L.*) plots along an urban gradient. Oecologia. 1998; 113: 269–277. https://doi.org/10.1007/s004420050378 PMID: 28308207

9. McIntyre NE. Ecology of Urban Arthropods: A Review and a Call to Action. an. 2000; 93: 825–835.

10. Deguines N, Julliard R, De Flores M, Fontaine C. Functional homogenization of flower visitor communities with urbanization. Ecology and evolution. Wiley Online Library; 2016; 6: 1967–1976. https://doi.org/10.1002/ece3.2009 PMID: 27066219

11. New TR. Promoting and developing insect conservation in Australia's urban environments. Austral Entomology. Wiley Online Library; 2018.

12. Mackin-Rogalska R, Pinowski J, Solon J, Wojcik Z. Changes in vegetation, avifauna, and small mammals in a suburban habitat. Polish Ecological Studies. 1988; 14: 293–330.

13. Pekin BK, Pijanowski BC. Global land use intensity and the endangerment status of mammal species. Diversity and Distributions. Wiley Online Library; 2012; 18: 909–918.

14. Blair RB. Birds and butterflies along urban gradients in two ecoregions of the United States: Is urbanization creating a homogeneous fauna? McKinney ML, Lockwood JL, editors. Biotic Homogenization. 2001; 33–56.

15. Strohbach MW, Haase D, Kabisch N. Birds and the city: urban biodiversity, land use, and socioeconomics. Ecology and Society. JSTOR; 2009; 14.

16. Alirol E, Getaz L, Stoll B, Chappuis F, Loutan L. Urbanisation and infectious diseases in a globalised world. The Lancet Infectious Diseases. Elsevier; 2011; 11: 131–141. https://doi.org/10.1016/S1473-3099(10)70223-1 PMID: 21272793

17. Crowl TA, Crist TO, Parmenter RR, Belovsky G, Lugo AE. The spread of invasive species and infectious disease as drivers of ecosystem change. Frontiers in Ecology and the Environment. Wiley-Blackwell; 2008; 6: 238–246.

18. Vitousek PM, D Antonio CM, Loope LL, Westbrooks R. Biological invasions as global environmental change. American Scientist. Research Triangle Park; 1996; 84: 468–478.

19. Seto KC, Fragkias M, Güneralp B, Reilly MK. A meta-analysis of global urban land expansion. PLoS ONE. Public Library of Science; 2011; 6: e23777. https://doi.org/10.1371/journal.pone.0023777 PMID: 21876770

20. Beninde J, Veith M, Hochkirch A. Biodiversity in cities needs space: a meta-analysis of factors determining intra-urban biodiversity variation. Ecol Lett. Wiley/Blackwell (10.1111); 2015; 18: 581–592. https://doi.org/10.1111/ele.12427 PMID: 25865805

21. Deplazes P, Hegglin D, Gloor S, Romig T. Wilderness in the city: the urbanization of Echinococcus multilocularis. Trends in Parasitology. 2004; 20: 77–84. PMID: 14747021

22. Bradley CA, Altizer S. Urbanization and the ecology of wildlife diseases. Trends in Ecology & Evolution. Elsevier; 2007; 22: 95–102. https://doi.org/10.1016/j.tree.2006.11.001 PMID: 17113678

23. Harrington LC, Scott TW, Lerdthusnee K, Coleman RC, Costero A, Clark GG, et al. Dispersal of the dengue vector Aedes aegypti within and between rural communities. The American Journal of Tropical Medicine and Hygiene. 2005; 72: 209–220. https://doi.org/10.4269/ajtmh.2005.72.209 PMID: 15741559

24. Adalja AA, Sell TK, Bouri N, Franco C. Lessons learned during dengue outbreaks in the United States, 2001–2011. Emerg Infect Dis. 2012; 18: 1–7. https://doi.org/10.3201/eid1801.110533

25. Levy MZ, Barbu CM, Castillo Neyra R, Quispe-Machaca VR, Ancca Juárez J, Escalante-Mejia P, et al. Urbanization, land tenure security and vector-borne Chagas disease. Proc Biol Sci. The Royal Society; 2014; 281: 20141003–20141003. https://doi.org/10.1098/rspb.2014.1003 PMID: 24990681

26. Levy MZ, Bowman NM, Kawai V, Waller LA, Cornejo del Carpio JG, Benzaquen EC, et al. Periurban *Trypanosoma cruzi*–infected *Triatoma infestans*, Arequipa, Peru. Emerg Infect Dis. Centers for Disease Control and Prevention; 2006; 12: 1345–1352. https://doi.org/10.3201/eid1209.051662 PMID: 17073082

27. Levy MZ, Kawai V, Bowman NM, Waller LA, Cabrera L, Pinedo-Cancino VV, et al. Targeted screening strategies to detect *Trypanosoma cruzi* infection in children. PLOS Neglected Tropical Diseases. Public Library of Science; 2007; 1: e103. https://doi.org/10.1371/journal.pntd.0000103 PMID: 18160979

28. Levy MZ, Small DS, Vilhena DA, Bowman NM, Kawai V, Cornejo del Carpio JG, et al. Retracing microepidemics of Chagas disease using epicenter regression. Meyers LA, editor. PLoS Comput Biol. Public Library of Science; 2011; 7: e1002146. https://doi.org/10.1371/journal.pcbi.1002146 PMID: 21935346

29. Bowman NM, Kawai V, Levy MZ, Cornejo del Carpio JG, Cabrera L, Delgado F, et al. Chagas disease transmission in periurban communities of Arequipa, Peru. CLIN INFECT DIS. 2008; 46: 1822–1828. https://doi.org/10.1086/588299 PMID: 18462104

30. Hunter GC, Borrini-Mayorí K, Ancca Juárez J, Castillo Neyra R, Verastegui MR, Malaga Chavez FS, et al. A field trial of alternative targeted screening strategies for Chagas disease in Arequipa, Peru. PLOS Neglected Tropical Diseases. Public Library of Science; 2012; 6: e1468. https://doi.org/10.1371/journal.pntd.0001468 PMID: 22253939

31. Foley EA, Khatchikian CE, Hwang J, Ancca Juárez J, Borrini-Mayorí K, Quispe-Machaca VR, et al. Population structure of the Chagas disease vector, *Triatoma infestans*, at the urban-rural interface. Molecular Ecology. 2013; 22: 5162–5171. https://doi.org/10.1111/mec.12471 PMID: 24103030

32. Gaunt MW, Yeo M, Frame IA, Stothard JR, Carrasco HJ, Taylor MC, et al. Mechanism of genetic exchange in American trypanosomes. Nature. Macmillian Magazines Ltd. SN; 2003; 421: 936–939. https://doi.org/10.1038/nature01438 PMID: 12606999

33. Berry AS, Salazar Sánchez R, Castillo Neyra R, Borrini-Mayorí K, Chipana-Ramos C, Vargas-Maquera M, et al. Sexual reproduction in a natural *Trypanosoma cruzi* population. PLOS Neglected Tropical Diseases. Public Library of Science; 2019; 13: e0007392. https://doi.org/10.1371/journal.pntd.0007392 PMID: 31107905

34. Lima L, Espinosa-Álvarez O, Ortiz PA, Trejo-Varón JA, Carranza JC, Pinto CM, et al. Genetic diversity of *Trypanosoma cruzi* in bats, and multilocus phylogenetic and phylogeographical analyses supporting Tcbat as an independent DTU (discrete typing unit). Acta Tropica. Elsevier; 2015; 151: 166–177. https://doi.org/10.1016/j.actatropica.2015.07.015 PMID: 26200788

35. Alroy KA, Huang C, Gilman RH, Quispe-Machaca VR, Marks MA, Ancca Juárez J, et al. Prevalence and transmission of *Trypanosoma cruzi* in people of rural communities of the high jungle of Northern Peru. PLOS Neglected Tropical Diseases. Public Library of Science; 2015; 9: e0003779. https://doi.org/10.1371/journal.pntd.0003779 PMID: 26000770

36. Pedersen D, Tremblay J, Errázuriz C, Gamarra J. The sequelae of political violence: assessing trauma, suffering and dislocation in the Peruvian highlands. Social science & medicine. Elsevier; 2008; 67: 205–217.

37. Bayer AM, Hunter GC, Gilman RH, Cornejo del Carpio JG, Naquira C, Bern C, et al. Chagas disease, migration and community settlement patterns in Arequipa, Peru. PLOS Neglected Tropical Diseases. Public Library of Science; 2009; 3: e567. https://doi.org/10.1371/journal.pntd.0000567 PMID: 20016830

38. Thomas CD. Translocation of species, climate change, and the end of trying to recreate past ecological communities. Trends in Ecology & Evolution. Elsevier Ltd; 2011; 26: 216–221. https://doi.org/10.1016/j.tree.2011.02.006 PMID: 21411178

39. Pitelka LF, Group PMW. Plant migration and climate change: a more realistic portrait of plant migration is essential to predicting biological responses to global warming in a world drastically altered by human activity. American Scientist. JSTOR; 1997;: 464–473.

40. Lockwood JL, Cassey P, Blackburn T. The role of propagule pressure in explaining species invasions. Trends in Ecology & Evolution. 2005; 20: 223–228.

41. Cohen JE, Cecere MC, Lauricella MA, Chuit R, Segura EL. Influence of humans and domestic animals on the household prevalence of *Trypanosoma cruzi* in *Triatoma infestans* populations in northwest Argentina. The American Journal of Tropical Medicine and Hygiene. ASTMH; 1998; 58: 748–758. https://doi.org/10.4269/ajtmh.1998.58.748 PMID: 9660458

42. Cohen JE, Gürtler RE. Modeling household transmission of American trypanosomiasis. Science. American Association for the Advancement of Science; 2001; 293: 694–698. https://doi.org/10.1126/science.1060638 PMID: 11474111

43. Gürtler RE, Cecere MC, Lauricella MA, Cardinal MV, Kitron U, Cohen JE. Domestic dogs and cats as sources of *Trypanosoma cruzi* infection in rural northwestern Argentina. Parasitology. Cambridge University Press; 2007; 134: 69–82. https://doi.org/10.1017/S0031182006001259 PMID: 17032467

44. Llewellyn MS, Miles MA, Carrasco HJ, Lewis MD, Yeo M, Vargas J, et al. Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. Holmes EC, editor. PLOS Pathogens. Public Library of Science; 2009; 5: e1000410. https://doi.org/10.1371/journal.ppat.1000410 PMID: 19412340

45. Rabinovich JE, Wisnivesky-Colli C, Solarz ND. Probability of transmission of Chagas disease by *Triatoma infestans* (Hemiptera: Reduviidae) in an endemic area of Santiago del Estero, Argentina. Bulletin of the World Health Organization. World Health Organization; 1990; 68: 737. PMID: 2127382

46. Nouvellet P, Dumonteil E, Gourbière S. The improbable transmission of *Trypanosoma cruzi* to human: the missing link in the dynamics and control of Chagas disease. PLOS Neglected Tropical Diseases. Public Library of Science; 2013; 7: e2505. https://doi.org/10.1371/journal.pntd.0002505 PMID: 24244766

47. Levy MZ, Tustin A, Castillo Neyra R, Mabud TS, Levy K, Barbu CM, et al. Bottlenecks in domestic animal populations can facilitate the emergence of *Trypanosoma cruzi*, the aetiological agent of Chagas disease. Proc R Soc B. The Royal Society; 2015; 282: 20142807.

48. Billig E. Detecting and controlling insect vectors in urban environments: novel Bayesian methods for complex spatial data. Publicly Accessible Penn Dissertations. 2017. Available: https://repository.upenn.edu/edissertations/2193

49. Oliveira RP, Broude NE, Macedo AM, Cantor CR, Smith CL, Pena SDJ. Probing the genetic population structure of *Trypanosoma cruzi* with polymorphic microsatellites. Proceedings of the National Academy of Sciences of the United States of America. The National Academy of Sciences; 1998; 95: 3776–3780. https://doi.org/10.1073/pnas.95.7.3776 PMID: 9520443

50. Higo H, Miura S, Horio M, Mimori T, Hamano S, Agatsuma T, et al. Genotypic variation among lineages of *Trypanosoma cruzi* and its geographic aspects. Parasitology International. 2004; 53: 337–344. https://doi.org/10.1016/j.parint.2004.06.001 PMID: 15464443

51. Llewellyn MS, Lewis MD, Acosta N, Yeo M, Carrasco HJ, Segovia M, et al. *Trypanosoma cruzi* IIc: phylogenetic and phylogeographic insights from sequence and microsatellite analysis and potential impact on emergent Chagas disease. PLOS Neglected Tropical Diseases. Public Library of Science; 2009; 3: e510. https://doi.org/10.1371/journal.pntd.0000510 PMID: 19721699

52. Curtis-Robles R, Snowden KF, Dominguez B, Dinges L, Rodgers S, Mays G, et al. Epidemiology and molecular typing of *Trypanosoma cruzi* in naturally-infected hound dogs and associated triatomine vectors in Texas, USA. PLOS Neglected Tropical Diseases. 2017; 11: e0005298–18. https://doi.org/10.1371/journal.pntd.0005298 PMID: 28095511

53. Hodo CL, Wilkerson GK, Birkner EC, Gray SB, Hamer SA. *Trypanosoma cruzi* transmission among captive nonhuman primates, wildlife, and vectors. EcoHealth. Springer; 2018;: 1–11.

54. Perez CJ, Lymbery AJ, Thompson RCA. Chagas disease: the challenge of polyparasitism? Trends in Parasitology. Elsevier Ltd; 2014; 30: 176–182. https://doi.org/10.1016/j.pt.2014.01.008 PMID: 24581558

55. Dumonteil E, Ramirez-Sierra M-J, Pérez-Carrillo S, Teh-Poot C, Herrera C, Gourbiére S, et al. Detailed ecological associations of triatomines revealed by metabarcoding and next- generation sequencing: implications for triatomine behavior and *Trypanosoma cruzi* transmission cycles. Scientific Reports. Springer US; 2018; 8: 1–13. https://doi.org/10.1038/s41598-017-17765-5

56. Kerr B, Riley MA, Feldman MW, Bohannan BJ. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. Nature Publishing Group. Nature Publishing Group; 2002; 418: 171.

57. Castillo Neyra R, Barbu CM, Salazar R, Borrini K, Naquira C, Levy MZ. Host-seeking behavior and dispersal of *Triatoma infestans*, a vector of Chagas disease, under semi-field conditions. PLOS Neglected Tropical Diseases. Public Library of Science; 2015; 9: e3433. https://doi.org/10.1371/journal.pntd.0003433 PMID: 25569228

58. Chiari E, Galvão L. Diagnóstico parasitológico da doença de Chagas. Clínica e terapêutica da doença de Chagas: uma abordagem prática para o clínico geral Rio de Janeiro: Fiocruz. 1997; 85–97.

59. Cáceres AG, Troyes L, Gonzáles-Pérez A, Llontop E, Bonilla C, Heredia N, et al. Enfermedad de Chagas en la región nororiental del Perú. I. *Triatominos* (Hemiptera, Reduviidae) presentes en Cajamarca y Amazonas. Revista Peruana de Medicina Experimental y Salud Publica. Instituto Nacional de Salud; 2002; 19: 17–23.

60. Harington JS. A simple apparatus for the artificial feeding of *Rhodnius prolixus* (Hemiptera, Reduviidae). Parasitology. Cambridge University Press; 1960; 50: 273–277. https://doi.org/10.1017/s0031182000025361 PMID: 14399808

61. Castillo Neyra R, Borrini-Mayorí K, Salazar Sánchez R, Ancca Suarez J, Xie S, Náquira Velarde C, et al. Heterogeneous infectiousness in guinea pigs experimentally infected with *Trypanosoma cruzi*. Parasitology International. 2016; 65: 50–54. https://doi.org/10.1016/j.parint.2015.09.009 PMID: 26432777

62. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, et al. The NCBI BioSystems database. Nucleic Acids Research. Oxford University Press; 2010; 38: D492–D496. https://doi.org/10.1093/nar/gkp858 PMID: 19854944

63. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. Oxford University Press; 2014; 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

64. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9: 357–359. https://doi.org/10.1038/nmeth.1923 PMID: 22388286

65. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 2010; 20: 1297–1303. https://doi.org/10.1101/gr.107524.110 PMID: 20644199

66. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. Oxford University Press; 2009; 25: 2283–2285. https://doi.org/10.1093/bioinformatics/btp373 PMID: 19542151

67. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution. 2016; 33: 1870–1874. https://doi.org/10.1093/molbev/msw054 PMID: 27004904

68. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution. Oxford University Press; 2012; 29: 1969–1973. https://doi.org/10.1093/molbev/mss075 PMID: 22367748

69. Heled J, Drummond AJ. Bayesian inference of population size history from multiple loci. BMC Evol Biol. 2008; 8: 289–15. https://doi.org/10.1186/1471-2148-8-289 PMID: 18947398

70. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, et al. BEAST 2: A software platform for Bayesian evolutionary analysis. Prlic A, editor. PLoS Comput Biol. Public Library of Science; 2014; 10: e1003537. https://doi.org/10.1371/journal.pcbi.1003537 PMID: 24722319

71. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA, Susko E. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology. 2018; 22: 341–syy032 VL–IS–. https://doi.org/10.1093/sysbio/syy032 PMID: 29718447

72. Suchard MA, Rambaut A. Many-core algorithms for statistical phylogenetics. Bioinformatics. Oxford University Press; 2009; 25: 1370–1376. https://doi.org/10.1093/bioinformatics/btp244 PMID: 19369496

73. Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, et al. BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. Systematic Biology. Oxford University Press; 2012; 61: 170–173. https://doi.org/10.1093/sysbio/syr100 PMID: 21963610

74. Oksanen J, Blanchet FG, Kindt R, Legendre P, O'hara RB, Simpson GL, et al. Vegan: community ecology package. R package version 1.17–4. http://cran r-project org> Acesso em. 2010; 23: 2010.

75. QGIS Development Team. QGIS Geographic Information System [Internet]. 2nd ed. Open Source Geospatial Foundation Project. Available: http://qgis.osgeo.org.