PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

Phylogeography and demographic history of the Chagas disease vector *Rhodnius nasutus* (Hemiptera: Reduviidae) in the Brazilian Caatinga biome

Tatiana Peretolchina[1,2☯], Márcio G. Pavan[1,3☯], Jessica Corrêa-Antônio[1], Rodrigo Gurgel-Gonçalves[4], Marli M. Lima[5], Fernando A. Monteiro[1]*

1 Laboratório de Epidemiologia e Sistemática Molecular, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil, **2** Laboratory of Molecular Systematics, Limnological Institute of the Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia, **3** Laboratório de Mosquitos Transmissores de Hematozoários, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil, **4** Laboratório de Parasitologia Médica e Biologia de Vetores, Faculdade de Medicina, Universidade de Brasília, Brasília, Distrito Federal, Brazil, **5** Laboratório de Ecoepidemiologia da doença de Chagas, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, RJ, Brazil

☯ These authors contributed equally to this work.
* fam@ioc.fiocruz.br

Citation: Peretolchina T, Pavan MG, Corrêa-Antônio J, Gurgel-Gonçalves R, Lima MM, Monteiro FA (2018) Phylogeography and demographic history of the Chagas disease vector *Rhodnius nasutus* (Hemiptera: Reduviidae) in the Brazilian Caatinga biome. PLoS Negl Trop Dis 12 (9): e0006731. https://doi.org/10.1371/journal.pntd.0006731

Editor: Ricardo E. Gürtler, Universidad de Buenos Aires, ARGENTINA

Received: January 1, 2018

Accepted: August 3, 2018

Published: September 24, 2018

Data Availability Statement: GenBank accession numbers of R. nasutus sequences generated for this study are MG734978-MG735124.

Funding: This study was funded by Instituto Nacional de Ciência e Tecnologia em Entomologia Molecular and by the Fundação Oswaldo Cruz of Brazil. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Background

Rhodnius nasutus, a vector of the etiological agent *Trypanosoma cruzi*, is one of the epidemiologically most relevant triatomine species of the Brazilian Caatinga, where it often colonizes rural peridomestic structures such as chicken coops and occasionally invades houses. Historical colonization and determination of its genetic diversity and population structure may provide new information towards the improvement of vector control in the region. In this paper we present thoughtful analyses considering the phylogeography and demographic history of *R. nasutus* in the Caatinga.

Methodology/Principal findings

A total of 157 *R. nasutus* specimens were collected from *Copernicia prunifera* palm trees in eight geographic localities within the Brazilian Caatinga biome, sequenced for 595-bp fragment of the mitochondrial cytochrome b gene (cyt b) and genotyped for eight microsatellite loci. Sixteen haplotypes were detected in the cyt b sequences, two of which were shared among different localities. Molecular diversity indices exhibited low diversity levels and a haplotype network revealed low divergence among *R. nasutus* sequences, with two central haplotypes shared by five of the eight populations analyzed. The demographic model that better represented *R. nasutus* population dynamics was the exponential growth model. Results of the microsatellite data analyses indicated that the entire population is comprised of four highly differentiated groups, with no obvious contemporary geographic barriers that could explain the population substructure detected. A complex pattern of migration was

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

observed, in which a western Caatinga population seems to be the source of emigrants to the eastern populations.

Conclusions/Significance

R. nasutus that inhabit *C. prunifera* palms do not comprise a species complex. The species went through a population expansion at 12–10 ka, during the Holocene, which coincides with end of the largest dry season in South America. It colonized the Caatinga in a process that occurred from west to east in the region. *R. nasutus* is presently facing an important ecological impact caused by the continuous deforestation of *C. prunifera* palms in northeast Brazil. We hypothesize that this ecological disturbance might contribute to an increase in the events of invasion and colonization of human habitations.

Author summary

Chagas disease is endemic to Latin America and the Caribbean and it is estimated that 6–7 million people are infected with the etiological agent *Trypanosoma cruzi*. Although new community-based ecosystem management (ecohealth) initiatives have been implemented, vector control based on insecticide-spraying of households remains one of the most effective strategies to diminish parasite transmission to humans. However, this strategy is not sustainable where native triatomine species are capable of colonizing peridomestic structures and invading human dwellings. The application of molecular markers with the potential of recovering both historical and contemporary information on vector population structure and diversity can improve the understanding of vector dissemination and thus contribute to the development of better disease control strategies. In this study we analyzed *Rhodnius nasutus* populations endemic to the Brazilian Caatinga biome using two sets of markers: a fragment of the mitochondrial gene cytochrome b and eight nuclear microsatellite loci. The information generated and described herein is important as it may contribute to the advancement of our understanding of Chagas disease vector ecology and phylogeography.

Introduction

Chagas disease is caused by the protozoan *Trypanosoma cruzi* (Kinetoplastida, Trypanosomatidae) and transmitted primarily through the feces of infected triatomine bugs (Hemiptera, Reduviidae) [1]. Endemic to Latin America and the Caribbean, it is estimated that approximately 6–7 million people are infected worldwide [2].

Although curable when treated early with antiparasitics in the acute phase, there is no vaccine available and treatment of chronic patients only reduces serum parasite detection, but not cardiac clinical complications [3]. Therefore, control programs have placed their efforts on the elimination of domestic vectors, by spraying insecticides indoors [4]. A recurrent problem seems to be the constant re-infestation of insecticide-treated households by abundant native triatomine species, which consists of a major challenge for the consolidation of Chagas disease control efforts [5, 6]. This dreadful scenario is often reported in Brazil where native *T. cruzi*-infected triatomines continuously invade houses in both rural and urban areas maintaining the risk of transmission [7, 8].

The Caatinga and the Cerrado biomes together, harbor most of the triatomine species diversity in Brazil [9]. Of the 68 triatomine species known to occur in the country, 10 have successfully adapted to the harsh droughts imposed by the Caatinga. *Rhodnius nasutus* Stål, 1859 is one of these species, frequently reported in bird and mammal nests on the crown of *Copernicia prunifera* (Carnaúba) palm trees in the vicinity of human habitations. It is often naturally infected with the *T. cruzi* parasite and is capable of invading rural and urban houses and colonizing peridomestic structures [10, 11]. It can also occur in *Attalea speciosa*, *Mauritia flexuosa*, *Syagrus oleracea* and *Acrocomia intumescens* palm trees [12, 13]. Interestingly, *R. nasutus* has also been reported to occur in *Licania rigida* (Oiticica) dicotyledon trees typical of northeast Brazil [10, 14].

Determination of *R. nasutus* population structure in the Caatinga biome and estimation of present gene flow may contribute to a better understanding of the species' historical colonization process as well as provide new information towards the improvement of vector control in the region. In this paper we present the phylogeography and demographic history of this Chagas disease vector in the Caatinga inferred with mtDNA (cyt b) and nuclear (microsatellites) markers.

Materials and methods

Insect collection

Two-hundred and twenty-eight *Rhodnius* specimens were collected after the inspection of 53 *Copernicia prunifera* palms in eight geographic localities within the Brazilian Caatinga biome (Table 1). The two closest collection sites are located 40 km apart (Altos and Campo Maior), while the most distant are 670 km apart (Parnaíba and Serra Talhada). The Caatinga biome covers nearly 85,000 km^2 of the northeast region, belongs to the seasonally dry tropical forests phytogeographic unit [15], and is a species-rich xeric environment [16, 17].

Insect field captures were approved by the *Instituto Chico Mendes de Conservação da Biodiversidade* (ICMBio) and were carried out following Gurgel-Gonçalves et al. [18]. Briefly, after reaching up to the palm tree crowns with a ladder, leaves and fibers were pruned and placed into plastic bags. The collected material was then lowered to the ground and scattered over a piece of white cloth to facilitate bug detection. *Rhodnius nasutus* specimens were identified morphologically based on Lent and Wygodzinsky [19] and confirmed by molecular taxonomy

Table 1. Sampling localities, number of *Rhodnius*-infested *Copernicia prunifera* palm trees and sample sizes of *R. nasutus* used in each molecular analysis.

Sampling locations	Geographical coordinates	Palms			Triatomines			Year	Code
		N_P	Infestation (%)	Habitat[1]	N_R	N^*	$N^{\dagger}$		
1. Altos, Piauí	05°02′S 42°28′W	5	80	Sylvatic	12	4	2	2010	ALT
2. Campo Maior, Piauí	04°49′S 42°10′W	8	50	Sylvatic	21	16	20	2010	CAM
3. Parnaíba, Piauí	02°54′S 41°46′W	5	80	Peridomestic	26	18	18	2010	PAR
4. Piracuruca, Piauí	03°55′S 41°43′W	4	75	Peridomestic	9	9	8	2010	PIR
5. Jaguaruana, Ceará	04°50′S 37°46′W	5	-	Peridomestic	44	32	44	2004	JAG
6. Serra Talhada, Pernambuco	07°59′S 38°17′W	7	57	Sylvatic	23	15	13	2010	STA
7. Sousa, Paraíba	06°45′S 38°13′W	6	100	Peridomestic	44	28	26	2010	SOU
8. Carnaúba dos Dantas, Rio Grande do Norte	06°33′S 36°35′W	13	54	Sylvatic	49	25	24	2010	CAR
	TOTAL	**53**			**228**	**147**	**155**		

N_P: number of palms infested with *Rhodnius* specimens

[1]: Sylvatic–forested areas >200m-distant from human dwellings, Peridomestic–deforested area <200m-distant from human dwellings; N_R: number of *Rhodnius* specimens collected; N^*: sample sizes used in microsatellite analyses; $N^{\dagger}$: sample sizes used in cyt b analyses.

https://doi.org/10.1371/journal.pntd.0006731.t001

based on a 682 base pair (bp) fragment of the mitochondrial cytochrome b gene (cyt b) [20]. Sequences generated with the amplification of this fragment were also used in the population genetics and phylogeographic analyses described in the next section.

Mitochondrial cyt b analyses

DNA extraction was performed using the Promega Wizard Genomic DNA extraction kit (Promega, Madison, Wisconsin, USA) following the manufacturers recommendations. A 682-bp fragment of the cyt b gene was PCR-amplified as described in Monteiro *et al.* [20]. Amplicons were purified with the Hi Yield Gel/PCR DNA extraction kit (Real Biotech, Banqiao, Taipei, Taiwan), or following the purification protocol with PEG-NaCl 20% - 5mM (modified from [21]). Both DNA strands were subjected to Sanger sequencing reactions with the ABI Prism BigDye Terminator v3.1 Cycle Sequencing kit (Thermo Fisher Scientific, Walthan, Massachusetts, USA) and run on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, California, USA). The removal of primer sequences, editing of both forward and reverse strands, and the generation of a consensus sequence for each sample were performed with Seqman Lasergene v. 7.0 (DNAStar, Inc., Madison, Wisconsin, USA). Consensus sequences were aligned to published ortholog sequences from other members of the 'robustus lineage' (*Rhodnius robustus* I-V, *R. prolixus*, *R. neglectus*, *R. nasutus* and *R. barretti*) and analyzed to confirm species identity based on genetic distances using Mega v. 5 [22]. Taxonomic identification was achieved based on the genetic similarity with other *Rhodnius* sequences available in the GenBank public database through the Basic Local Alignment Search Tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi), considering the following parameters: percentage of identity and coverage, e-value, and also Bayesian phylogenetic analyses.

A Bayesian phylogenetic cyt b tree based on the 147 samples of morphologically-identified *R. nasutus* and the other members of the 'robustus lineage' was inferred in BEAST v. 1.8 [23]. *Rhodnius barretti* [24] was used to root the maximum clade credibility tree. The best fit model of substitution was determined using jModeltest v. 2 [25]. We tested whether a strict or a relaxed molecular clock best fitted our data through a Bayesian random local clock analysis (RLC) [26]. Three independent runs were performed for 10^9 generations, sampling every 20,000 generations. Convergence of parameters and proper mixing were confirmed through the calculation of effective sample sizes (ESS) in Tracer v. 1.6 [27], and ESS estimates above 10^4 were considered appropriate [28].

Gene tree and species tree were reconstructed with *BEAST [29]. Information in the literature about the taxonomic identification of DNA sequences retrieved from GenBank was used to assign prior information on each sequence to a particular species. The 147 cyt b sequences from field-collected specimens without molecular identification were labeled as "*Rhodnius* sp.". Suggested divergence rate for triatomines of 1.1 to 1.8% per Myr was used as a prior with normal distribution [30].

Molecular indices of haplotype diversity (Hd) and nucleotide diversity (π) were computed in DnaSP v. 5 [31] for each population and sequence divergence between populations were calculated in Mega v. 5 [22]. A median-joining network [32] was constructed with Network v. 4.6 (Fluxus Technology Ltd. 2008) for a better visualization of the relationships between *R. nasutus* cyt b haplotypes. Since samples from Jaguaruana were collected in 2004 and all other samples in 2010, we tested whether this could have an impact in our study by carrying out an analysis of variance (AMOVA) grouping samples according to the collection date with Arlequin v. 3.5 [33].

Levels of genetic differentiation among populations were determined with Wright's pairwise F_{ST} comparisons [34] in Arlequin v. 3.5 [33]. Correction for multiple testing was done

with the false discovery rate (FDR) method [35] under 5% of significance level (P-value = 0.013). The coefficient of determination R^2 to evaluate the fit of data to theoretical models of distribution of genetic distances under different demographic scenarios was calculated in the R environment [36]. Deviations from neutrality were assessed with Fu's Fs [37] and Tajima's D [38] tests in DnaSP v. 5 [31]. Significant and negative values of both tests indicate population size expansion or purifying selection. Mismatch distribution tests were also used to infer possible demographic and spatial expansions [39].

We also used Beast v. 2.4 [40] to reconstruct a Bayesian skyline plot (BSP) and test which of the two different demographic models (constant population size or exponential growth) better explained the demographic history of *R. nasutus*. The best model was selected through the comparison of marginal likelihood, calculated with the path sampling algorithm [41] under 10^6 MCMC chain length and 60 steps, based on Bayes Factors (BF) [42]. In all cases, results from two independent runs (2×10^7 generations with the first 2×10^6 discarded as burn-in and parameter values sampled every 2×10^3 generations) were analyzed with Tracer v. 1.6 [27]. Convergence of parameters and proper mixing were confirmed through the calculation of ESS, and estimates above 10^4 were considered appropriate [28]. Because only a single species was analyzed (and thus it is highly probable that specimens from all populations evolve at a single rate), we imposed a strict clock in the analysis. Suggested divergence rates for triatomines of 1.1 to 1.8% per Myr were used [30]. We matched the *R. nasutus* demographic history to paleoclimatic events estimated through oxygen isotope marine paleotemperature records [43], used as etalon for the study of global climate changes.

Microsatellite analyses

Microsatellite alleles of eight loci were amplified in a total reaction volume of 10 uL containing 1 unit of *TaqGold* DNA polymerase, 2 mM of each dNTP, 1.5 mM $MgCl_2$, 10 mM of magnesium free Buffer 10X (Thermo Fisher Scientific Co., Walthan, Massachusetts, USA), 5 pmol of each primer and approximately 10 ng of extracted DNA. PCR conditions were as follows: 1 min at 95˚C, 30 cycles of 30 s at 94˚C, 30 s at Ta ˚C and 45 s at 72˚C, and 72˚C for 30 min; Ta is the annealing temperature for each locus (S1 Table).

All PCR products were run with a size standard GS500 LIZ on an ABI 3730xl Genetic Analyzer, and allele fragment lengths quantified using the Peak Scanner software v. 1.0 (Applied Biosystems, Foster City, California, USA). Because these primers are heterologous (designed to amplify microsatellite loci of other *Rhodnius* species) [44–46], the orthology of microsatellite regions and repetition motifs were molecularly validated. Amplified loci were submitted to Sanger sequencing reactions with the ABI Prism BigDye Terminator v3.1 Cycle Sequencing kit (Thermo Fisher Scientific, Walthan, Massachusetts, USA) and run on an ABI 3730 automated sequencer (Applied Biosystems, Foster City, California, USA).

Microsatellite genotypes were screened for likely scoring errors, allele dropout, and presence of null alleles with Micro-Checker v. 2.2 [47]. Number of shared genotypes, number of private alleles for each sampling site and Shannon's allele information index ($^{S}H_A$) as a measure of gene diversity were performed with GenAlEx v. 6.5 [48]. Deviations from Hardy–Weinberg equilibrium and tests of linkage disequilibrium for each locus were performed with Arlequin v. 3.5 software [33].

Genetic differentiation between populations (pairwise F_{ST}) between sampling localities and inbreeding coefficients (F_{IS}) were carried out with Arlequin v. 3.5 software [33]. Sampling localities for which F_{ST} was non-significant were considered as belonging to the same population. Mutual index ($^{S}H_{UA}$) [49] were also estimated with GenAlEx v. 6.5 [48].

The Bayesian clustering program Structure v. 2.3 [43] was used to estimate population assignment without prior assumptions of population subdivision. We used the *admixture model* due to the lack of information about the ancestry of the field-collected specimens, with correlated allele frequencies, which means that these frequencies in different populations are likely to be similar as a consequence of migration or shared ancestry [50]. Burn-in and simulation were set at 2.5 x 10^5 iterations and 10^6 Markov Chain Monte Carlo (MCMC) generations, respectively. Ten independent runs were performed for each value of K (for 2–8), as suggested by Pritchard et al. [50]. The most likely value of K was estimated with the ΔK method [51].

The Ewens–Watterson neutrality test [52], which relies on the comparison of observed and expected homozygosity, was performed with the Arlequin v. 3.5 software [33].The program Bottleneck v. 1.2 [53] was used to test bottleneck events by evaluating deviations of mutation-drift equilibrium. Three different mutation models of microsatellite evolution were employed: Infinite Allele model (IAM), Stepwise Mutation Model (SMM), and Two-Phased model (TPM), which is an intermediate to the SMM and IAM as it incorporates the mutational process of the former, but allows for mutations of a larger magnitude to occur. The significance was assessed with the Wilcoxon sign-rank test [54], a more powerful and robust test when few polymorphic loci are available (< 20).

We also constructed a population network with EDENetwork v. 2.1 [55] to determine possible source and sink populations of *R. nasutus*. Basically, this analysis reconstructs a minimum-spanning tree based on the pairwise F_{ST} matrix and calculates three different parameters: (a) degree, which is defined as the number of connections a node has in the network, summarizing which populations are exchanging migrants; (b) clustering, which measures the number of subpopulations within a population; and (c) betweenness, which is the number of shortest paths between populations, reflecting areas of intense gene flow. These measures allow identifying "source" and "sink" populations (i.e. those with the highest degree and betweenness) [56].

To investigate the patterns of gene flow between the populations defined by Structure we performed a Bayesian approach implemented in Migrate-n v. 3.2 [57]. Parameters were first estimated under a full migration model that allowed gene flow to occur among all regions. Then we tested 10 reduced models (based on F_{ST} results) representing different patterns of migration. Migrate-n analysis was performed with a single long cold chain and four hot swapping chains using a static heating scheme (temperature: 1.0, 1.5, 3.0, 10^6), with two independent runs, sampling at every 100^{th} step for a total of 2 x 10^5 recorded steps (burn-in = 3 x 10^4 steps). The 10 migration models were compared based on their marginal likelihood and probability using thermodynamic integration with Bezier approximation (implemented in Migrate-n) according to Beerli and Palczewski [58]. The plot for visualizing of migration pattern inferred by Migrate-n analysis was drawn in the R environment [36] using Migest Package [59]. The migration scenario with highest probability was tested through an AMOVA with Arlequin v. 3.5 [33].

Isolation by distance was estimated for cyt b and microsatellite data through linear regression analyses of dependence, comparing the logarithms of genetic distances (pairwise F_{ST} values) [60], and geographic distances between the eight localities, in the R environment [36]. The significance level of hypothesis testing was set at $\alpha = 0.05$.

Results

Morphological taxonomy and ecology

Thirty-three of the 53 *C. prunifera* palm trees investigated (62%) were infested with *Rhodnius* specimens. Higher infestation percentages were detected in deforested areas (Table 1). Thirty-

three palm trees were sampled in four forested locations and showed lower infestation percentages (Mean: 58%, range = 50–80%) than the 15 palm trees collected in four deforested areas (Mean: 87%, 75–100%). From the 228 *Rhodnius* specimens collected, 163 insects were morphologically identified as *Rhodnius nasutus*.

Mitochondrial cyt b analyses

Only two of the DNAs extracted from all 157 specimens resulted in unsuccessful amplification of the desired cyt b fragment. Eight specimens exhibited low-quality sequences (i.e. < 500-bp) and thus were discarded from the analysis. Therefore, our dataset included 147 cyt b sequences of 595-bp long from *R. nasutus* individuals sampled from eight localities in the Brazilian Caatinga (Table 1).

Pairwise comparison of the sequences obtained with those available in GenBank revealed 99–100% identity with another *R. nasutus* "reference" sequence (JX273155) generated by our group [24]. This sequence was obtained from a specimen collected in Jaguaruana, Ceará, where the morphologically similar species *R. neglectus* is not known to occur [61].

Hasegawa-Kishino-Yano (HKY [62]) was selected as the best evolutionary model for the data, following the Akaike and Bayesian Information criteria. RLC analysis showed that a "strict" clock is more suitable to our cyt b dataset (rate change: median = 0; variance = 0.67; ESS = 4020). Bayesian species tree (Fig 1) revealed that all sequences from *Rhodnius* specimens collected formed a monophyletic and well-supported clade with the *R. nasutus* reference sequence (PP = 1.0), corroborating further their taxonomic identity. A bayesian coalescent gene tree disclosed the short branches (i.e. low sequence divergence) among *R. nasutus* samples. Although posterior probabilities for *R. nasutus* clades were low (PP < 0.9), it is noticeable

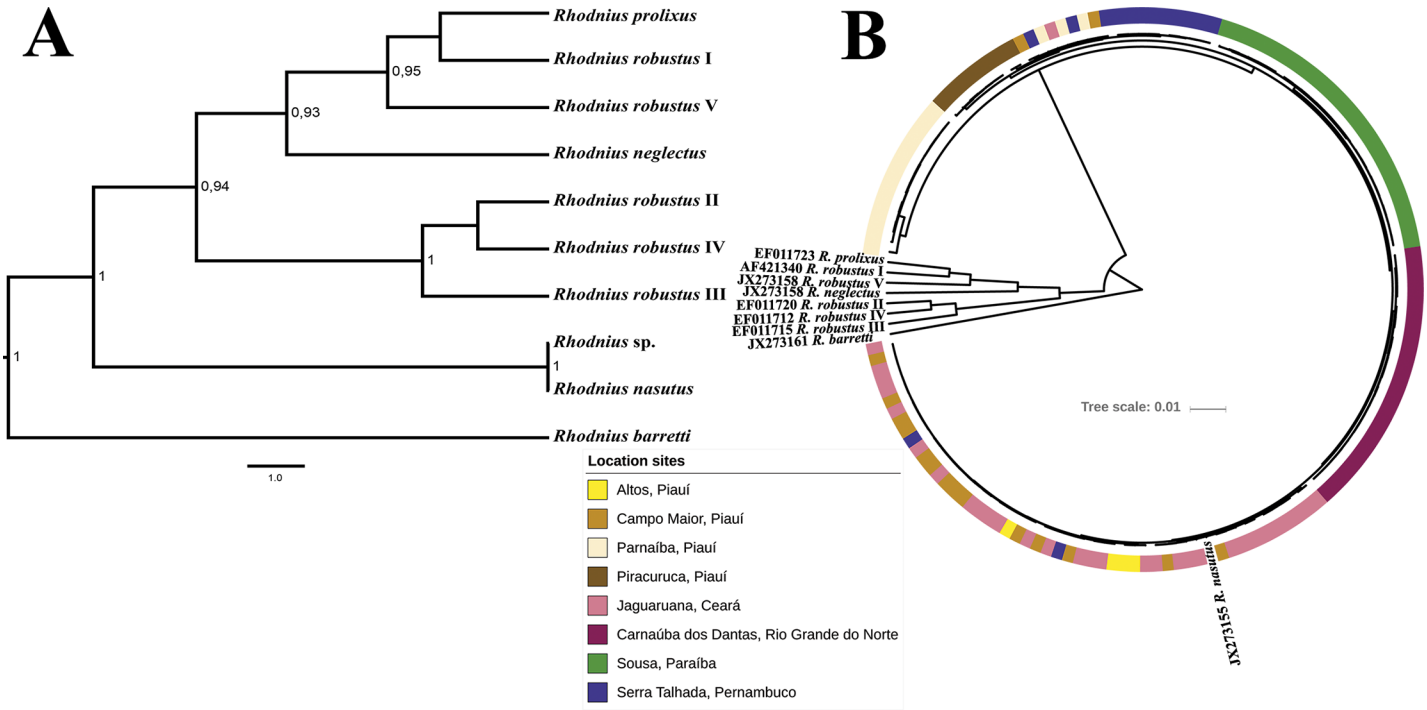

Fig 1. A–Maximum clade credibility species tree reconstructed with 595-bp cyt b sequences of 156 *Rhodnius* specimens. Posterior probabilities above 0.95 are show for key nodes. B–Coalescent gene tree with 147 sequences of 595-bp cyt b. GenBank accession numbers of *R. nasutus* sequenced in this study are MG734978-MG735124. The accession numbers of the sequences retrieved from GenBank are shown on the tips of their respective tree stems.

https://doi.org/10.1371/journal.pntd.0006731.g001

that only the *R. nasutus* sequences from Carnaúba dos Dantas and Sousa clustered in separate clades without the presence of at least one sequence from another locality. Sequences from the other six localities clustered together in other three different clades.

AMOVA analysis between Jaguaruana and the other populations did not indicate any sampling effect of collecting at different times as genetic differences between these groups were not significant (Φ_{ST} = 0.18, P = 0.98). Therefore, these samples were included in further analyses. Molecular divergence of *R. nasutus* cyt b sequences varied between 0%-0.8% (S2 Table), as expected for intraspecific comparisons in *Rhodnius* species [20]. Pairwise comparisons within and between localities ranged from 0% to 0.2% and 0.2% to 0.8%, respectively. Overall, the most divergent sequences were present in Parnaíba (mean divergence = 0.62%) and the less divergent sequences were in Jaguaruana (mean divergence = 0.27%). Inspection of the sequences revealed 15 polymorphic sites and 16 haplotypes, with two haplotypes (3 and 9) shared between different localities (Fig 2A). Molecular diversity indices (Table 2) showed high

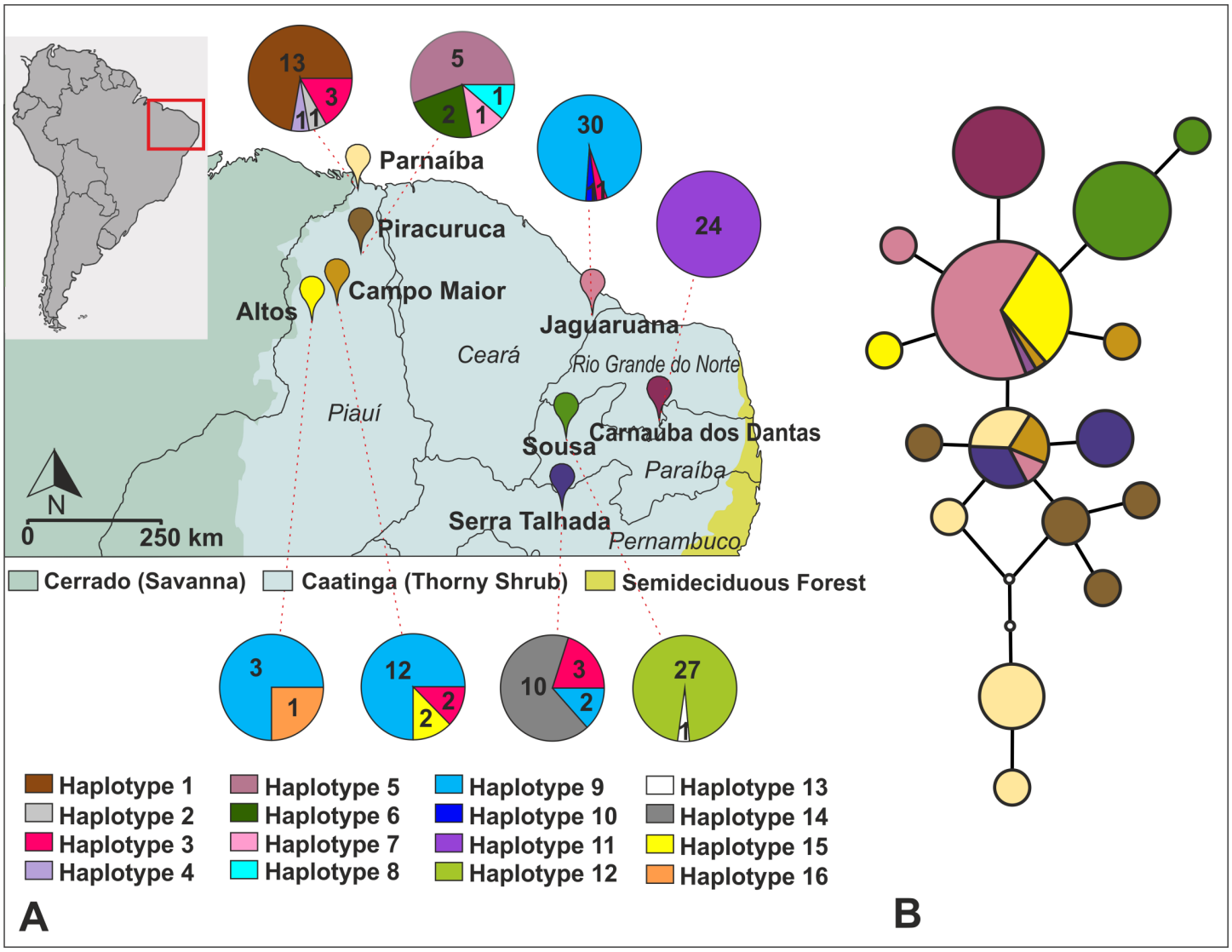

Fig 2. A–Geographical distribution of *Rhodnius nasutus* cytb haplotypes in each sampled locality in northeastern Brazil. Slice sizes are proportional to haplotype frequency. B—Haplotype network based on cyt b sequences. Circle sizes are proportional to haplotype frequency.

https://doi.org/10.1371/journal.pntd.0006731.g002

haplotype diversity (Hd = 0.45–0.83 ± 0.10), but low nucleotide diversity between haplotypes (π = 0.0032 ± 0.0002). The highest haplotype diversities were found in Piracuruca (N = 9, Hd = 0.694±0.147) and Altos (N = 4, Hd = 0.667±0.204), and the lowest haplotype diversity in Sousa (N = 28, Hd = 0.071±0.065). All sequences from Carnaúba dos Dantas (N = 25) were identical. The highest nucleotide diversities were found in sequences from Parnaíba (N = 18, π = 0.00253±0.00073) and Piracuruca (N = 9, π = 0.00205±0.00058), and the lowest nucleotide diversities in sequences from Sousa (N = 28, π = 0.00012±0.00011) and Jaguaruana (N = 32, π = 0.00021±0.00013). Pairwise population F_{ST} estimates are given in Table 3. Most of the comparisons revealed high and significant values (> 0.6). Non-significant F_{ST} values were observed only when the populations of Altos, Campo Maior and Jaguaruana were compared. The excessive number of high F_{ST} values is due to the presence of 14 (out of 16) unique haplotypes.

Haplotypes derived from cyt b sequences revealed a network (Fig 2B) with two central haplotypes (haplotypes 3 and 9) that are very abundant and widespread, and to which several less common haplotypes are closely related (1–2 mutational steps). The most frequent haplotype (Haplotype 9, N = 46) is shared with specimens from Altos, Campo Maior and Jaguaruana, and the other haplotype (Haplotype 3, N = 9) is shared with specimens from Campo Maior, Jaguaruana, Parnaíba and Serra Talhada. This type of haplotype connection suggests population expansion or retention of ancestral polymorphism with little migration, evidenced by the geographically restricted haplotypes.

Considering that all *R. nasutus* sequences are separated by small genetic distances (most haplotypes are connected by a single mutational step), neutrality tests, mismatch distribution and BSP analyses were performed with the entire dataset. Mismatch distribution analysis (Fig 3A) exhibits a relatively good fit to the expected mismatch distributions under the model of population expansion (coefficient of determination R^2 for population growth model is 0.95 (p = 0.008) while for constant population size model is 0.72 (p = 0.160)). Neutrality tests did not indicate significant departures from neutrality (Fu's *Fs* = -4.27, *p* = 0.085; Tajima's *D* = -0.75, *p* = 0.246). Bayesian analyses of population growth revealed that marginal likelihood for constant population was -1240.82 and for exponential growth was -1228.27. Thus, the demographic model with the best fit was the exponential growth model (BF = 25.11).

Demographic history of the species (Fig 3B) depicts a population expansion that started at about 10 thousand years ago (ka), which coincided with a global climate warming corresponding to the first Marine Isotope Stage (MIS). Migrate-n analysis with the cyt b dataset resulted in no clear pattern of migration probably due to the small amount of variation among sequences, and thus it was not included in the paper.

Table 2. Molecular diversity indices for the *R. nasutus* sequences collected in eight localities.

Localities	N	S	π (±SD)	h	Hd (± SD)
1. ALT	4	1	0.00112 (± 0.00034)	2	0.667 (± 0.204)
2. CAM	16	2	0.00078 (± 0.00028)	3	0.433 (± 0.138)
3. CAR	25	0	0.00000	1	0.000
4. JAG	32	2	0.00021 (± 0.00013)	3	0.123 (± 0.078)
5. PAR	18	5	0.00253 (± 0.00074)	4	0.471 (± 0.130)
6. PIR	9	4	0.00205 (± 0.00058)	4	0.694 (± 0.147)
7. SOU	28	1	0.00012 (± 0.00011)	2	0.071 (± 0.065)
8. STA	15	2	0.00124 (± 0.00033)	3	0.553 (± 0.126)
R. nasutus	147	15	0.00322 (± 0.00023)	16	0.826 (± 0.018)

N = sample size, S = number of polymorphic sites, π = nucleotide diversity, SD = standard deviation, h = number of haplotypes; Hd = haplotype diversity.

https://doi.org/10.1371/journal.pntd.0006731.t002

Table 3. Pairwise F_{ST} values for microsatellite data (above the diagonal) and cyt b data (below the diagonal).

Localities	1	2	3	4	5	6	7	8
1-ALT	-	0.07	0.24*	0.11	0.30*	0.49*	0.25*	0.64*
2-CAM	0.20	-	0.08	0.04	0.27*	0.32*	0.26*	0.53*
3-PAR	0.74*	0.77*	-	0.16	0.42*	0.30*	0.36*	0.64*
4-PIR	0.62*	0.66*	0.58*	-	0.28*	0.40*	0.22*	0.55*
5-CAR	0.93*	0.85*	0.88*	0.89*	-	0.51*	0.31*	0.19
6-SOU	0.89*	0.83*	0.88*	0.89*	0.98*	-	0.47*	0.65*
7-STA	0.71*	0.69*	0.74*	0.58*	0.92*	0.91*	-	0.53*
8-JAG	0.44	0.06	0.85*	0.82*	0.93*	0.91*	0.83*	-

* Significant results ($P < 0.013$)

https://doi.org/10.1371/journal.pntd.0006731.t003

Microsatellite loci

Eight microsatellite loci (S1 Table) were analyzed for 155 of the 157 *R. nasutus* specimens collected in the Caatinga biome.

The number of alleles per locus ranged from 1 (R8) to 14 (List14-064). Since the R8 locus was monomorphic for all individuals, it was excluded from the analysis.

The basic statistics for the microsatellite data is summarized in Table 4. Three loci showed significant deviations from HWE in at least one population, possibly due to the admixture of

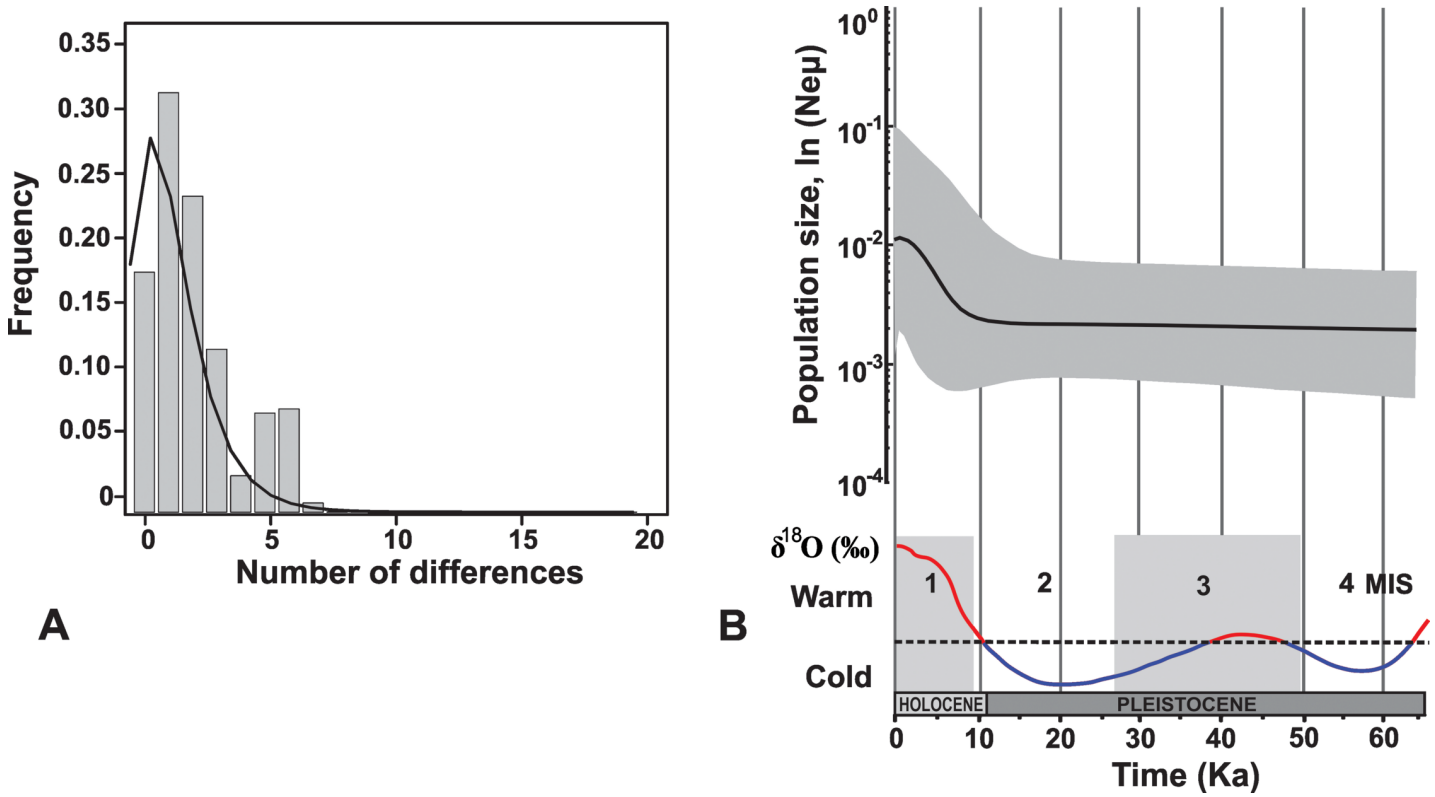

Fig 3. A—Mismatch distribution for *R. nasutus* based on cyt b sequences. Bars represent observed values, lines represent expected values under the model of sudden population growth; B—Calibrated demographic history of *R. nasutus* and reconstructed paleoclimatic history. Shaded areas correspond to 1st and 3rd Marine Isotope Stages.

https://doi.org/10.1371/journal.pntd.0006731.g003

Table 4. Microsatellite genetic variation (seven loci) for the four groups of *R. nasutus*.

	Group 1 (ALT+CAM+PAR+PIR)								
Locus	**R31**	**List14-010**	**List14-025**	**L43**	**List14-013**	**List14-021**	**List14-064**	**Average gene diversity**	**F_{IS} (*P* value)**
N	52	48	50	44	42	43	44	0.157± 0.153	-0.063 (0.851)
N_A	1	2	1	5	6	4	10		
P_A	0	0	0	0	5	2	3		
H_O	ND	**0.234**	ND	**0.955**	0.381	**0.047**	0.795		
H_E	ND	**0.491**	ND	**0.749**	0.372	**0.069**	0.728		
	Group 2 (CAR+JAG)								
N	68	68	68	68	64	68	67	0.255 ± 0.169	**0.122 (0.018)**
N_A	2	3	2	4	2	2	6		
P_A	1	1	0	0	1	0	1		
H_O	0.015	0.044	0.162	0.485	0.266	**0.103**	0.328		
H_E	0.015	0.044	0.152	0.543	0.255	**0.220**	0.326		
	Group 3 (STA)								
N	13	13	13	13	13	13	13	0.255 ± 0,169	-0.036(0.674)
N_A	2	2	1	3	2	6	1		
P_A	1	0	0	0	0	1	0		
H_O	0.692	0.077	ND	0.462	**0.000**	0.615	ND		
H_E	0.508	0.077	ND	0.394	**0.147**	0.658	ND		
	Group 4 (SOU)								
N	26	26	26	25	24	26	26	0.184 ± 0.136	-0.009(0,610)
N_A	1	3	2	3	1	2	4		
P_A	0	1	0	1	0	1	0		
H_O	ND	0.077	0.115	0.640	ND	0.077	0.231		
H_E	ND	0.147	0.111	0.525	ND	0.145	0.217		

N—sample size for each group; N_A—number of alleles at each locus; H_O and *H*E—observed and expected heterozygosities, respectively; P_A—private alleles; F_{IS}—inbreeding coefficient; **ALT**–Altos, **CAM**–Campo Maior, **PAR**–Parnaíba, **PIR**–Piracuruca; **CAR**—Carnaúba dos Dantas; **JAG**–Jaguaruana; **SOU**–Sousa; **STA**—Serra Talhada. Significant deviations from Hardy–Weinberg equilibrium are shown in bold.

https://doi.org/10.1371/journal.pntd.0006731.t004

null alleles or LD. Microsatellite loci analyses revealed moderate levels of genetic diversity and low allelic diversity. Null alleles were detected for loci List14-010 in Group 1 and List14-021 in Group 2. Significant LD between List-025 and List-064 loci, and List-025 and L43 were found in Group 1.

Among 155 individuals, 111 had unique genotypes: ALT+CAM– 22; PAR– 18; PIR– 8; CAR– 17; JAG– 25; STA– 9; SOU—15. There are two shared genotypes between CAR and JAG, one genotype shared between STA and JAG, one shared between PAR and PIR, and another shared between PAR and SOU. The highest number of shared genotypes was observed among *R. nasutus* specimens from Jaguaruana. In addition, positive F_{IS} values were significant in Group 2, which comprised Jaguaruana specimens.

The Bayesian clustering using seven microsatellite loci for 155 individuals revealed the existence of four groups (Fig 4). The ΔK method indicated 4 groups as the most likely population structure. Group 1 includes all sampling locations of *R. nasutus* in Altos (ALT), Campo Maior (CAM), Parnaíba (PAR), and Piracuruca (PIR). Group 2 is the most numerous and genetically variable and contains samples from Jaguaruana (JAG) and Carnaúba dos Dantas (CAR). Group 3 includes specimens from Serra Talhada (STA), and Group 4 is composed of specimens from Sousa (SOU). These groups were in accordance with the pairwise F_{ST} comparisons among sequences from different localities. Pairwise F_{ST} values between sampling localities

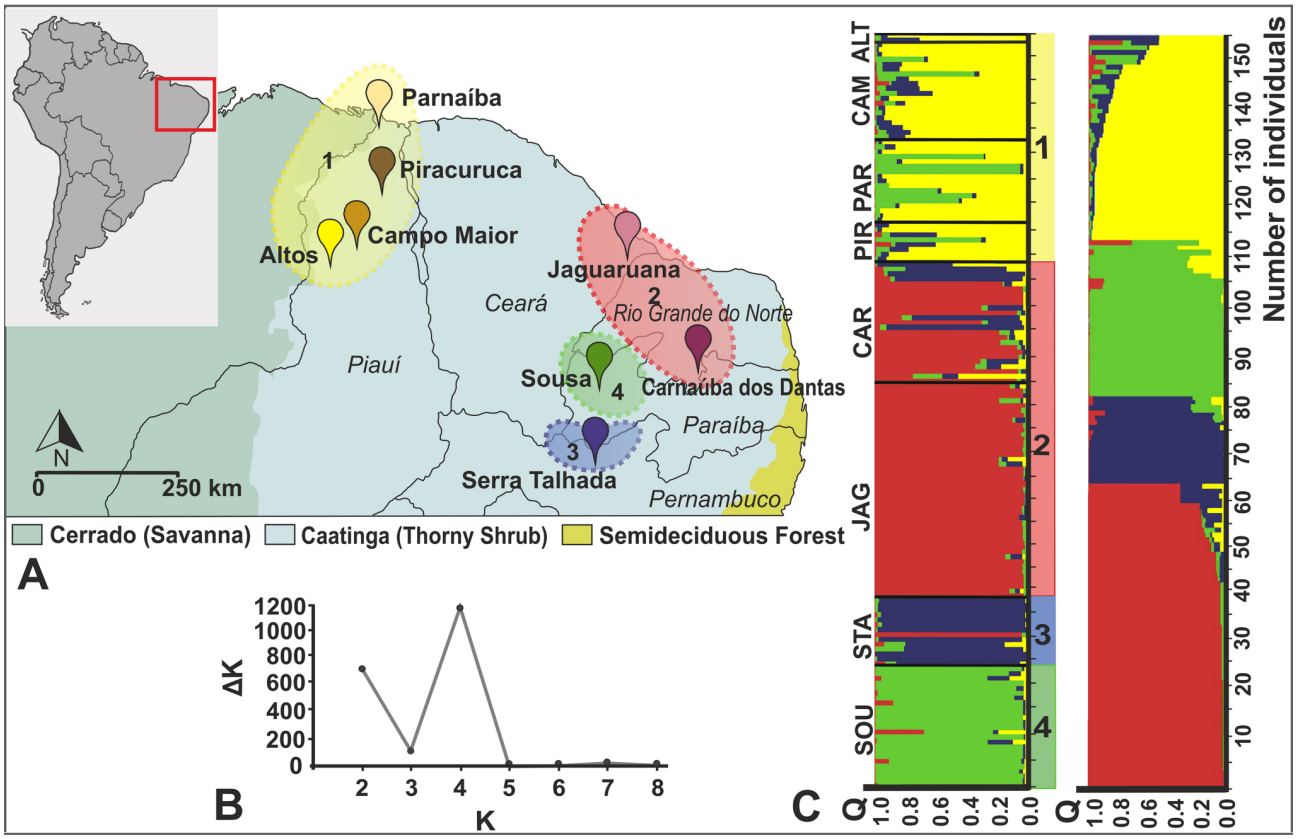

Fig 4. A—Sampling sites map of *R. nasutus* specimens; B– ΔK values are shown for K ranging between 2 and 8; C—population structure of *R. nasutus* estimated with Structure for $K = 4$ organized by geographic region (on the left) and by Q value (on the right). Each vertical bar represents the individual probability of a single bug to belong to a specific genetic cluster; the color(s) in each bar represent the proportion of the individual's genome coming from each of the four assumed clusters. Black lines separate the sampling localities. Structure diagrams for $K = 2$–8 are presented as supporting information (S1 Fig).

https://doi.org/10.1371/journal.pntd.0006731.g004

(Table 3) revealed non-significant differences between Altos, Campo Maior, Piracuruca and Parnaíba, and also between Jaguaruana and Carnaúba dos Dantas. Pairwise F_{ST} values between Sousa, Serra Talhada and all other localities were significant.

The highest value of Shannon's allele information index ($^{S}H_{A}$) describing genetic variation in population was shown in Group 1 (0.955). Group 1 also displayed the highest number of private alleles (10) over all loci. The other groups had a smaller number of private alleles: four

Table 5. $^{S}H_{UA}$ indexes (above the diagonal), F_{ST} values (below the diagonal) and $^{S}H_{A}$ indexes (in diagonal) among the four groups of *R. nasutus* determined based on seven microsatellite loci.

Pairwise F_{ST}, $^{S}H_{UA}$	Group 1	Group 2	Group 3	Group 4
Group 1	0.955	0.270	0.159	0.144
Group 2	0.438*	0.584	0.122	0.203
Group 3	0.275	0.418*	0.624	0.197
Group 4	0.277*	0.562*	0.470*	0.431

Group 1: ALT—Altos (Piauí), CAM—Campo Maior (Piauí), PAR—Parnaíba (Piauí), PIR—Piracuruca (Piauí); Group 2: CAR—Carnaúba dos Dantas (Rio Grande do Norte), JAG—Jaguaruana (Ceará); Group 3: STA—Serra Talhada (Pernambuco), Group 4—SOU—Sousa (Paraíba).

* $P < 0.05$

https://doi.org/10.1371/journal.pntd.0006731.t005

Table 6. Ewens–Watterson neutrality test.

	Group 1		Group 2		Group 3		Group 4	
Locus	F_O	F_E	F_O	F_E	F_O	F_E	F_O	F_E
R31	ND	ND	0.985	0.814	0.512	0.752	ND	ND
List14-010	0.515	0.810	0.510	0.690	0.926	0.748	0.388	0.456
List14-025	ND	ND	0.851	0.815	ND	ND	0.891	0.785
L43	**0.259**	**0.475**	0.461	0.582	0.621	0.584	0.486	0.631
List14-021	0.931	0.563	0.781	0.815	0.858	0.754	0.858	0.790
List14-064	0.255	0.266	0.688	0.448	0.367	0.318	0.786	0.444
List14-013	0.633	0.420	0.747	0.811	ND	ND	ND	ND

F_O–observed homozygosity, F_E–expected homozygosity. F_O and F_E values indicating significant difference ($P < 0.05$) are shown in bold.

https://doi.org/10.1371/journal.pntd.0006731.t006

private alleles were found in Groups 2 and 4, and only two private alleles were found in Group 3 (Table 4).

Genetic differentiation based on the F_{ST} and Shannon's mutual information index $^{S}H_{UA}$ between groups estimated in the Structure program [50] exhibited similar but not identical patterns (Table 5). The highest differentiation values were found between pairs of groups 2–4 (F_{ST} = 0.562, $^{S}H_{UA}$ = 0.203), 3–4 (F_{ST} = 0.470, $^{S}H_{UA}$ = 0.197) and 1–2 (F_{ST} = 0.438, $^{S}H_{UA}$ = 0.270). The lowest values of F_{ST} were obtained for groups 1 and 3, although the lowest values of $^{S}H_{UA}$ were indicated between groups 2 and 3. Also, relatively low values of F_{ST} and $^{S}H_{UA}$ were found between groups 1 and 4 (0.277 and 0.144, respectively).

Results of the Ewens-Watterson test indicated neutral evolution of all loci in all groups, but in Group 1, in which neutrality was rejected in locus L43 as observed homozygosity values were lower than expected (Table 6). Because past demographic events can cause significant departures from neutrality, we tested Group 1 for the occurrence of possible bottleneck events. Results of the analysis performed with the program Bottleneck did not indicate any signs of recent population decline among *R. nasutus* specimens from Group 1. One-tailed Wilcoxon sign-rank test for heterozygote excess indicated that each group was in mutation-drift equilibrium for all mutation models evaluated: IAM, TPM, and SMM ($P > 0.05$).

EDENetwork analysis revealed strong connections between Groups 1 and 3 and 1 and 4. Groups 2 and 3 were also connected. We did not observe any indication of substructuring within groups (Clustering = 0). Groups 1 and 3 were identified as possible source-sink populations based on the network analysis. These groups exhibited the highest connection values (calculated through the degree parameter) and betweenness, which indicate that they are exchanging migrants with other groups (Fig 5A).

Two of seven microsatellite loci (L43, List14-025) had complex repetition motifs that did not fit a SMM. Although List14-064 and List14-021 were reported with the same dinucleotide repeat motifs [43], we observed nucleotide substitutions (S2 Fig). Thus, we imposed the IAM mutation model as more appropriate for our data set for migration analysis. The estimates of the marginal likelihood (lmL) and the posterior model probability (*P*) for the 10 different models representing various scenarios of migration in Migrate-n revealed support for the 7th model (Table 7). Migration pattern of *R. nasutus* in the Caatinga biome inferred by Migrate-n analysis with "full migration" model (Fig 5B) revealed the existence of bidirectional gene flow between Groups 2 and 3. Unidirectional gene flow was detected from Group 2 to 4, from Group 1 to 3, and from Group 4 to 3. Group 1 seems to be a source population, because it only provides emigrants to other Groups, whereas Group 3 is a sink population as it is composed by immigrants from the other three populations. AMOVA between east (Carnaúba dos Dantas, Jaguaruana, Sousa and Serra Talhada) and west (Altos, Campo Maior, Parnaíba and

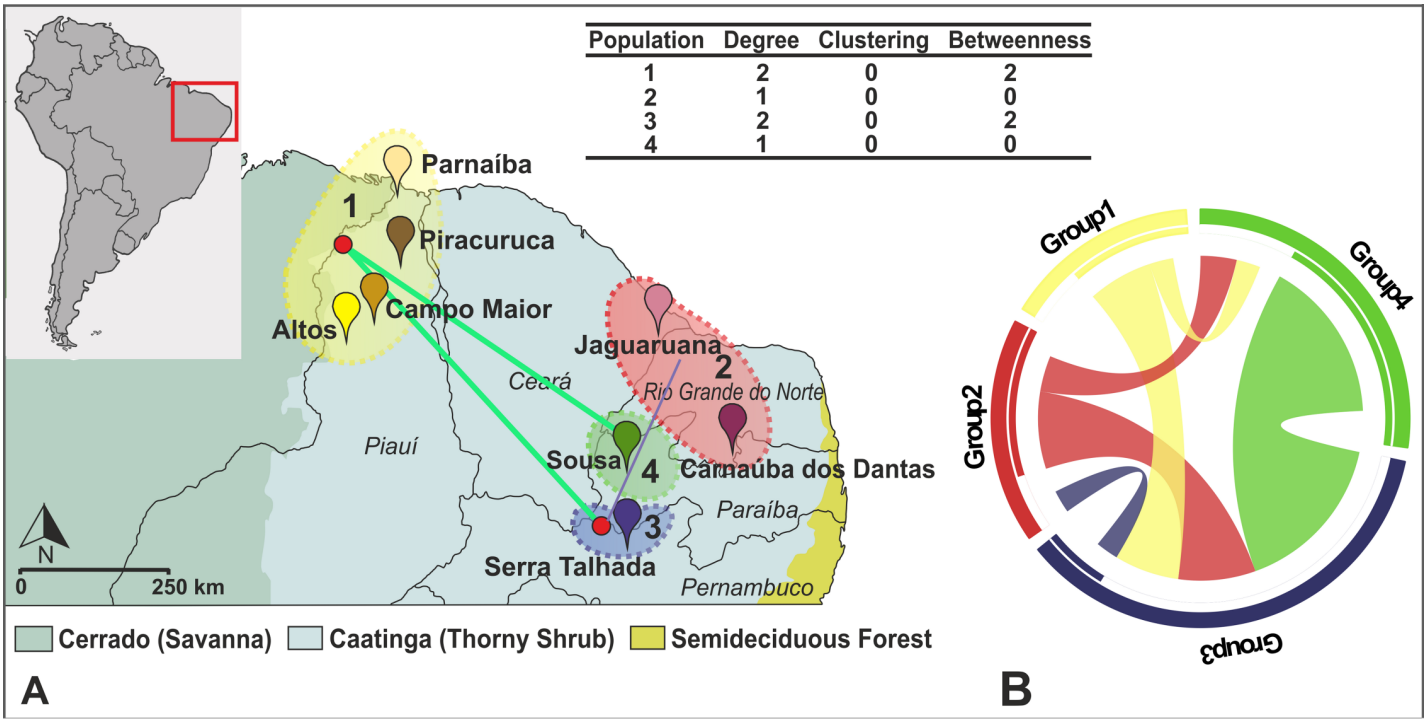

Fig 5. A—Populational network based on microsatellite data showing two populations (1 and 3) as source of migrants for all others, based on higher values of degree and betweenness components. Red circles represent the fundamental units of the system (i.e. source-sink populations). Each line represents the connection of a source population to a sink population. Line thickness is proportional to linkage strength of such relationships. B—Migration pattern of *R. nasutus* in the Caatinga region based on Migrate-n analysis. The color of each stripe in the middle of the circle represents the color of the source group of migrants. The width of each stripe is proportional to the number of migrants.

https://doi.org/10.1371/journal.pntd.0006731.g005

Piracuruca) populations shows that the percentage of within-population variation is higher (53.18%, $\Phi_{ST} = 0.46$, $P < 0.0001$) than among populations within groups (33.51%, $\Phi_{SC} = 0.39$, $P < 0.0001$) or between groups (13.31%, $\Phi_{CT} = 0.13$, $P < 0.21$). The latter non-significant value of AMOVA statistic thus indicates gene flow between west and east groups.

Linear regression analysis (S3 Fig) did not reveal a statistically significant correlation between cyt b-derived genetic distances and geographic distances ($R^2 = 0.05$, $P = 0.258$). However, a weak correlation was observed between microsatellite-based F_{ST} distances and geographic distances ($R^2 = 0.29$, $P = 0.003$), suggesting that distance contributes ~29% to the shaping of genetic diversity [63].

Discussion

This paper describes the population structure and demographic history of *R. nasutus* specimens collected from *C. prunifera* palm trees throughout most of the Caatinga biome. The pattern of genetic diversity found with the analyses of seven microsatellite loci and cytb sequences suggests that the colonization of the Caatinga by *R. nasutus* took place from west to east through the region (Fig 5), partially coinciding with a population expansion event that occurred during the last 12–10 thousand years (Fig 3).

R. nasutus genetic diversity and population structure

The genetic structure revealed by microsatellite analyses (Figs 4 and 5) indicate the existence of four highly differentiated groups with pairwise F_{ST} values (>0.25; Table 5) well above those

Table 7. Model selection with Migrate-n.

No	Migration pattern	Harmonic lmL	*P*
1	1—>2—>3<—4<—1; 4<—>2	-487.39	0.00000
2	3—>1—>2<—3—>4	-709.12	0.00000
3	4<—>1—>3—>2; 1—>2	-571.86	0.00000
4	4<—>1<—>2; 1—>3	-632.87	0.00000
5	2<—>3<—4<—>2	-625.25	0.00000
6	1<—3<—>2—>4	-758.89	0.00000
7	**3<—1—>4—>3<—>2**	**-451.98**	**0.99999**
8	4<—1—>3—>2—>4—>3	-579.88	0.00000
9	1<—2—>3—>1; 4<—>2	-535.22	0.00000
10	4—>1<—2<—3—>1	-690.89	0.00000

Arrow heads indicate the direction of migration. The highest migration pattern probability is shown in bold. *No*—model number, *Harmonic lmL*—harmonic estimates of the marginal likelihood, and *P*—posterior probability for each model. Population numbers correspond to: 1 –Group 1; 2 –Group 2; 3 –Group 3; 4 –Group 4.

https://doi.org/10.1371/journal.pntd.0006731.t007

estimated for other triatomine populations at similar geographical scales, such as for *R. prolixus* in Venezuela [5] and *T. infestans* in Peru and Argentina [64, 65]. Although this could indicate low levels of gene flow between populations or even the possibility of cryptic speciation (as seen for other *Rhodnius* species [5]), results from the cyt b marker do not support the latter hypothesis. There is no evidence of present potential isolation by distance that could explain the population substructure observed herein. A weak correlation was found between genetic and geographical distances (S2 Fig), assuming that the correlation between geographic and genetic distances indicate isolation by distance (but see reference [66] for a discussion about the limitations of this analysis). Indeed, populations of Group 1 are geographically close and have low and non-significant pairwise F_{ST} values, but other populations such as Sousa and Carnaúba dos Dantas, which are also close, are highly structured ($F_{ST} = 0.51$, $P < 0.01$). Therefore, we can hypothesize that restricted gene flow in the Caatinga region might result from human modifications, since the main sylvatic ecotope of this species, *C. prunifera* palm trees, has been reduced significantly due to continuous deforestation [67]. Biodiversity losses in tropical areas have been strongly determined by deforestation, which can interrupt environmental patches that uphold connectivity with other habitats [68]. Future research considering the possible routes of human migration among the studied communities, as well as a more complete analysis of the distribution of *C. prunifera* palm trees are necessary to adequately address this hypothesis.

Reduced genetic diversity was detected in the three *R. nasutus* groups from eastern Caatinga (groups 2, 3 and 4) evidenced by the low number of alleles present (< 5) in the majority of the loci analyzed, and low heterozygosity levels. Reduced diversity commonly occurs in endemic species which have low effective population sizes caused by habitat fragmentation [69]. In this case, stochastic processes such as genetic drift and inbreeding within isolated patches are expected to shape the genetic background of populations [70]. Indeed, Group 2 shows moderate levels of inbreeding ($F_{IS} = 0.122$, $P = 0.018$).

Although two loci are monomorphic (List14-025 and R31), Group 1 from western Caatinga presents the highest heterozygosity levels. Moreover, this group has the highest Shannon's allele information index ${}^{S}H_{A}$ (0.955), further indicating its high variability. Two loci display heterozygote excess (L43 and List14-021), which might result from isolate-breaking (when two previously isolated populations are mixed [71]). It is important to note that Group 1 also has the highest number of private alleles. Perhaps the high number of private alleles observed in this group results from long-term isolation during glaciation periods.

Evidence for departures from neutrality and bottleneck events were not detected in the *R. nasutus* populations analyzed. However, this result should be interpreted with caution, since the small sample size of some locations could prevent proper testing of this hypothesis [72]. Moreover, the reduction of effective population size, heterozygosity levels and allelic diversity may go undetected when (i) they occurred for a limited time interval or very recently [73], (ii) are followed by a sudden population growth [74], or (iii) when populations experienced low levels of subsequent immigration [75].

Gene flow among groups and colonization scenarios based on microsatellite data

A scattered mosaic of habitat fragments, such as observed in the Caatinga region, might cause population reduction due to an impact in the connectivity of habitat networks, favoring dispersal through a series of successive stepping-stones events [76, 77]. Network and Bayesian coalescent analyses indicated a western-eastern colonization of the biome. Group 1 is a source population, from which individuals emigrated to Groups 3 and 4, while Group 3 is a sink population, also receiving emigrants from Groups 2 and 4 (Fig 5A). Geographically close groups of eastern Caatinga have a complex pattern of gene exchange, with unidirectional migration of Group 2 individuals to Group 4, and Group 4 individuals to Group 3, as well as bidirectional gene flow between Groups 2 and 3 (Fig 5B).

Unidirectional gene flow may be explained by a decrease in capacity at the source habitat and availability of expansion potential at the target area [78]. The niche capacity of *R. nasutus* may be limited by several ecological factors, such as the sparse distribution of preferred food sources (fauna associated with the palms) and human activity.

Phylogeography, demographic history based on mitochondrial DNA sequence analysis

The network shows that most haplotypes differ by a single substitution, indicating reduced diversity. These haplotypes seem to be under neutral evolution, since 13 out of 16 variable sites resulted in synonymous substitutions. Two haplotypes are shared by five geographically distant locations (Serra Talhada, Parnaíba, Jaguaruana, Campo Maior and Altos), which could indicate high gene flow among populations or reflect retention of ancestral polymorphism [79]. The remaining 14 haplotypes are exclusive (Fig 2), and thus support restricted migration. Pairwise F_{ST} comparisons corroborate the latter hypothesis, as 24 out of 28 values are high (> 0.6) and significant (Table 3).

Populations can become structured quickly during a range expansion event as a consequence of the colonization of newly available patchy areas by pioneering groups with reduced population sizes [76]. In the case of a patchy environment, there are limitations to dispersal and therefore the colonization of new habitats will increase genetic structure among populations as only a subset of the original gene pool will be transferred [72].

Evidence suggests that populations of *R. nasutus* expanded from west to east throughout the Caatinga in the last 12–10 ka. Molecular diversity indices showed high haplotype diversity (Hd) and low nucleotide diversity (π), which indicates that despite the high number of different haplotypes per population, they only differ by 1–5 nucleotides. This observation is consistent with expansion of source populations with small effective size [79]. Mismatch distribution analysis showed a good fit to the expected under the model of population growth (Fig 3A). Moreover, the demographic model that better represents the population dynamics of *R. nasutus* is the exponential growth (BF = 25.11), evidenced by a 5-fold increase in the species population size in the Holocene (Fig 3B).

The demographic history of *R. nasutus* is estimated to have begun 66 thousand years ago thus covering Pleistocene–Holocene geological epochs (Fig 3B). Based on the Bayesian Skyline Plot results, the population size of *R. nasutus* was stable for approximately 50,000 years. This demographic stability may reflect the presence of a putative refuge area for Seasonally Dry Tropical Forests in the Caatinga, spanning most of the species current distribution area during the Last Glacial Maximum (24–12 ka) [16].

R. nasutus went through a population expansion 12–10 ka, during the Holocene. This expansion is consistent with global climate changes (Fig 3B), as it coincides with the end of the largest dry season in South America, during the glaciation Wisconsin-Würm (18–12 ka). This period is characterized by the expansion of semi-humid vegetation in South America and diversification of Neotropical biota in seasonally dry areas [80, 81]. Moreover, the Caatinga biome experienced an increase in moisture levels during the late Pleistocene (12–6 ka), which possibly contributed to the expansion of *Mauritia* palm tree species in the region [82] and likely also *Copernicia* palms.

Incongruence between nuclear and mitochondrial data

The genetic structure of *R. nasutus* populations inferred based on the two markers differed considerably (Figs 2 and 4). Cyt b analysis indicated more complex substructure than the microsatellite data (Table 3). Although Group 3 (Serra Talhada) and 4 (Sousa) remained separate by both microsatellites and cyt b data, microsatellite-defined Group 1 comprises four populations (Altos, Campo Maior, Piracuruca and Parnaíba), and Group 2 includes two populations (Jaguaruana and Carnaúba dos Dantas), whereas cyt b data grouped together Altos and Campo Maior with the geographically distant population of Jaguaruana. Assuming that sampled populations represent panmitic units, we might speculate that in the past, Altos, Campo Maior and Jaguaruana were connected, while Piracuruca and Parnaíba were separated. In a contemporary event, the Jaguaruana population became isolated from the western group, which clustered with the other two western populations, Piracuruca and Parnaíba. It is worth mentioning, however, that F_{ST} results based on cyt b data should be interpreted with caution, due to the low divergence found between *R. nasutus* sequences. In a contemporary event, the Jaguaruana population became isolated from the western group, which clustered with the other two western populations, Piracuruca and Parnaíba. Group 1 heterozygote excess and pairwise cyt b and microsatellite F_{ST} values among populations (0.20–0.77 and 0.07–0.24, respectively), thus corroborate the hypothesis of an isolate-breaking effect.

Incongruent results may also be explained by different evolutionary mechanisms (inheritance modes, population size, rates of evolution, response to selective pressure) and intrinsic biological peculiarities of triatomines (e.g. males fly more than females). Poor accordance between population genetic structures inferred from mitochondrial and microsatellite markers have also been reported for various animal groups, including mammals [83], amphibians [84] and insects [85].

Dispersal capacity and epidemiological implications

R. nasutus was able to recently disperse within the Caatinga in response to favorable climatic events and environmental conditions. The species is presently facing an important ecological impact caused by the continuous deforestation of *C. prunifera* palm trees. Perhaps this ecological disturbance is contributing to the observation that *R. nasutus* now colonizes other species of palms (*A. speciosa*, *M. flexuosa*, *S. oleracea*, *A. intumescens* [12, 13]) and occasionally, even trees (*L. rigida* [14]). These observations show that *R. nasutus* populations have the ability to adapt to new microhabitats. In addition, it also invades human-modified areas and

sporadically colonizes dwellings [10, 11, 86]. We observed that *C. prunifera* palm trees in deforested areas seem to be more prone to become infested with *R. nasutus* than those in forested areas (87% and 58%, respectively), which raises the risk of *T. cruzi* transmission to humans.

Demographic history data indicates the capacity of *R. nasutus* to increase its population size when environmental conditions are favorable. Therefore, closer attention should be paid to the distribution of *R. nasutus*, as it can maintain *T. cruzi* transmission within the Caatinga. Given the limitations inherent to insecticide spraying such as continuous re-infestation of insecticide-treated households by abundant native vectors, molecular studies focusing on the distribution patterns and demographic trends of sylvatic triatomines should be conducted in order to further improve the effectiveness of Chagas disease control.

Supporting information

S1 Table. Microsatellite loci used in this study.
(DOCX)

S2 Table. Molecular divergence of *R. nasutus* cyt b sequences from the same locality (in bold) and from different localities. N–number of sequences analyzed.
(DOCX)

S1 Fig. Population structure of *R. nasutus* estimated with Structure for $K = 2–8$.
(TIF)

S2 Fig. Electropherograms of four different alleles of *locus* List14-064.
(JPG)

S3 Fig. Linear regression analysis between logarithms of pairwise $F_{ST}/(1 - F_{ST})$ ratios and geographic distances.
(TIF)

Acknowledgments

The comments and suggestions of three anonymous reviewers greatly improved an earlier version of the manuscript. Dr Karina Morelli provided technical assistance on microsatellite amplification. DNA sequencing and microsatellite amplification were conducted at the Oswaldo Cruz Institute sequencing facility.

Author Contributions

Conceptualization: Tatiana Peretolchina, Márcio G. Pavan, Rodrigo Gurgel-Gonçalves, Fernando A. Monteiro.

Formal analysis: Tatiana Peretolchina, Márcio G. Pavan, Jessica Corrêa-Antônio.

Funding acquisition: Rodrigo Gurgel-Gonçalves, Marli M. Lima, Fernando A. Monteiro.

Methodology: Jessica Corrêa-Antônio.

Resources: Rodrigo Gurgel-Gonçalves, Marli M. Lima, Fernando A. Monteiro.

Software: Tatiana Peretolchina, Márcio G. Pavan, Jessica Corrêa-Antônio.

Supervision: Fernando A. Monteiro.

Visualization: Tatiana Peretolchina.

Writing – original draft: Tatiana Peretolchina, Márcio G. Pavan.

Writing – review & editing: Márcio G. Pavan, Rodrigo Gurgel-Gonçalves, Fernando A. Monteiro.

References

1. Coura J.R., Chagas disease: what is known and what is needed—a background article. Mem Inst Oswaldo Cruz, 2007. 102 Suppl 1: p. 113–22.
2. WHO, Chagas Disease (American trypanosomiasis). Fact sheet 340. http://www.who.int/mediacentre/factsheets/fs340/en/ 2017 [cited 2017 December 23].
3. Morillo C.A., et al., Randomized Trial of Benznidazole for Chronic Chagas' Cardiomyopathy. N Engl J Med, 2015. 373(14): p. 1295–306. https://doi.org/10.1056/NEJMoa1507574 PMID: 26323937
4. Abad-Franch F., et al., Modeling disease vector occurrence when detection is imperfect: infestation of Amazonian palm trees by triatomine bugs at three spatial scales. PLoS Negl Trop Dis, 2010. 4(3): p. e620. https://doi.org/10.1371/journal.pntd.0000620 PMID: 20209149
5. Fitzpatrick S., et al., Molecular genetics reveal that silvatic Rhodnius prolixus do colonise rural houses. PLoS Negl Trop Dis, 2008. 2(4): p. e210. https://doi.org/10.1371/journal.pntd.0000210 PMID: 18382605
6. Abad-Franch F., et al., Certifying the interruption of Chagas disease transmission by native vectors: cui bono? Mem Inst Oswaldo Cruz, 2013. 108(2): p. 251–4. https://doi.org/10.1590/0074-0276108022013022 PMID: 23579810
7. Ribeiro G., et al., Frequent house invasion of Trypanosoma cruzi-infected triatomines in a suburban area of Brazil. PLoS Negl Trop Dis, 2015. 9(4): p. e0003678. https://doi.org/10.1371/journal.pntd.0003678 PMID: 25909509
8. Brito R.N., et al., Drivers of house invasion by sylvatic Chagas disease vectors in the Amazon-Cerrado transition: A multi-year, state-wide assessment of municipality-aggregated surveillance data. PLoS Negl Trop Dis, 2017. 11(11): p. e0006035. https://doi.org/10.1371/journal.pntd.0006035 PMID: 29145405
9. Gurgel-Gonçalves R., et al., Geographic distribution of chagas disease vectors in Brazil based on ecological niche modeling. J Trop Med, 2012. 2012: p. 705326. https://doi.org/10.1155/2012/705326 PMID: 22523500
10. Abad-Franch F., et al., Ecology, evolution, and the long-term surveillance of vector-borne Chagas disease: a multi-scale appraisal of the tribe Rhodniini (Triatominae). Acta Trop, 2009. 110(2–3): p. 159–77. https://doi.org/10.1016/j.actatropica.2008.06.005 PMID: 18619938
11. Parente C.C., et al., Community-Based Entomological Surveillance Reveals Urban Foci of Chagas Disease Vectors in Sobral, State of Ceará, Northeastern Brazil. PLoS One, 2017. 12(1): p. e0170278. https://doi.org/10.1371/journal.pone.0170278 PMID: 28103294
12. Dias F.B., et al., Ecological aspects of Rhodnius nasutus Stål, 1859 (Hemiptera: Reduviidae: Triatominae) in palms of the Chapada do Araripe in Ceará, Brazil. Mem Inst Oswaldo Cruz, 2008. 103(8): p. 824–30. PMID: 19148424
13. Abad-Franch F., et al., On palms, bugs, and Chagas disease in the Americas. Acta Trop, 2015. 151: p. 126–41. https://doi.org/10.1016/j.actatropica.2015.07.005 PMID: 26196330
14. Lima M.M. and Sarquis O., Is Rhodnius nasutus (Hemiptera; Reduviidae) changing its habitat as a consequence of human activity? Parasitol Res, 2008. 102(4): p. 797–800. https://doi.org/10.1007/s00436-007-0823-1 PMID: 18095000
15. Pennington R.T., Prado D.E., and Pendry C.A., Neotropical seasonally dry forests and Quaternary vegetation changes. J Biogeogr, 2000. 27: p. 261–273.
16. Werneck F.P., The diversification of eastern South American open vegetation biomes: historical biogeography and perspectives. Quat. Sci. Rev., 2011. 30: p. 1630–1648.
17. Miranda E.A., et al., Pleistocene climate changes shaped the population structure of Partamona seridoensis (Apidae, Meliponini), an endemic stingless bee from the Neotropical dry forest. PLoS One, 2017. 12(4): p. e0175725. https://doi.org/10.1371/journal.pone.0175725 PMID: 28410408
18. Gurgel-Gonçalves R., et al., Infestation of Mauritia flexuosa palms by triatomines (Hemiptera: Reduviidae), vectors of Trypanosoma cruzi and Trypanosoma rangeli in the Brazilian savanna. Acta Trop, 2012. 121(2): p. 105–11. https://doi.org/10.1016/j.actatropica.2011.10.010 PMID: 22037200
19. Lent H. and Wygodzinsky P., Revision of the Triatominae (Hemiptera, Reduviidae) and their significance as vectors of Chagas disease. 1979: Bulletin of American Museum of Natural History. p. 125–520.

20. Monteiro F.A., et al., Molecular phylogeography of the Amazonian Chagas disease vectors Rhodnius prolixus and R. robustus. Mol Ecol, 2003. 12(4): p. 997–1006. PMID: 12753218

21. Lis J.T., Fractionation of DNA fragments by polyethylene glycol induced precipitation. Methods Enzymol., 1980. 65(1): p. 347–353. PMID: 6246357

22. Tamura K., et al., MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol, 2007. 24(8): p. 1596–9. https://doi.org/10.1093/molbev/msm092 PMID: 17488738

23. Drummond A.J., et al., Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol, 2012. 29(8): p. 1969–73. https://doi.org/10.1093/molbev/mss075 PMID: 22367748

24. Abad-Franch F., et al., Rhodnius barretti, a new species of Triatominae (Hemiptera: Reduviidae) from western Amazonia. Mem Inst Oswaldo Cruz, 2013. 108 Suppl 1: p. 92–9.

25. Darriba D., et al., jModelTest 2: more models, new heuristics and parallel computing. Nat Methods, 2012. 9(8): p. 772.

26. Drummond A.J. and Suchard M.A., Bayesian random local clocks, or one rate to rule them all. BMC Biol, 2010. 8: p. 114. https://doi.org/10.1186/1741-7007-8-114 PMID: 20807414

27. Rambaut, A., et al., Tracer v.1.6, Available from http://beast.bio.ed.ac.uk/Tracer. 2014.

28. Drummond A.J. and Bouckaert R.R., *Bayesian Evolutionary Analysis with BEAST*. 1 ed. 2015: Cambridge University Press. 260.

29. Heled J. and Drummond A.J., Bayesian inference of species trees from multilocus data. Mol Biol Evol, 2010. 27(3): p. 570–80. https://doi.org/10.1093/molbev/msp274 PMID: 19906793

30. Pfeiler E., et al., Genetic variation, population structure, and phylogenetic relationships of Triatoma rubida and T. recurva (Hemiptera: Reduviidae: Triatominae) from the Sonoran Desert, insect vectors of the Chagas' disease parasite Trypanosoma cruzi. Mol Phylogenet Evol, 2006. 41(1): p. 209–21. https://doi.org/10.1016/j.ympev.2006.07.001 PMID: 16934496

31. Librado P. and Rozas J., DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics, 2009. 25(11): p. 1451–2. https://doi.org/10.1093/bioinformatics/btp187 PMID: 19346325

32. Bandelt H.J., Forster P., and Röhl A., Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol, 1999. 16(1): p. 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036 PMID: 10331250

33. Excoffier L. and Lischer H.E., Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour, 2010. 10(3): p. 564–7. https://doi.org/10.1111/j.1755-0998.2010.02847.x PMID: 21565059

34. Weir B.S. and Cockerham C.C., Estimating F-statistics for the analysis of population structure. Evolution, 1984. 38(6): p. 1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x PMID: 28563791

35. Benjamini Y, and Hochberg Y., Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol, 1995: p. 289–300.

36. R Development Core, R: *A language and environment for statistical computing*. 2013, R Foundation for Statistical Computing: Vienna, Austria.

37. Fu Y.X., Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics, 1997. 147(2): p. 915–25. PMID: 9335623

38. Tajima F., Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 1989. 123(3): p. 585–95. PMID: 2513255

39. Rogers A.R. and Harpending H., Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol, 1992. 9(3): p. 552–69. https://doi.org/10.1093/oxfordjournals.molbev.a040727 PMID: 1316531

40. Bouckaert R., et al., BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol, 2014. 10(4): p. e1003537. https://doi.org/10.1371/journal.pcbi.1003537 PMID: 24722319

41. Baele G., Lemey P., and Suchard M.A., Genealogical Working Distributions for Bayesian Model Testing with Phylogenetic Uncertainty. Syst Biol, 2016. 65(2): p. 250–64. https://doi.org/10.1093/sysbio/syv083 PMID: 26526428

42. Kass R.E. and Raftery A.E., Bayes Factors. J. Am. Stat. Assoc., 1995. 90: p. 773–795.

43. Shackleton N.J., The last interglacial in the marine and terrestrial records. Proc R Soc Lond B Biol Sci, 1969. 174(1034): p. 135–154.

44. Harry M., et al., Isolation and characterization of microsatellite markers in the bloodsucking bug Rhodnius pallescens (Heteroptera, Reduviidae). Mol Ecol, 1998. 7(12): p. 1784–6. PMID: 9859210

45. Harry M., et al., Microsatellite markers from the Chagas disease vector, Rhodnius prolixus (Hemiptera, Reduviidae), and their applicability to Rhodnius species. Infect Genet Evol, 2008. 8(3): p. 381–5. https://doi.org/10.1016/j.meegid.2008.01.006 PMID: 18304894

46. Fitzpatrick S., et al., A panel of ten microsatellite loci for the Chagas disease vector Rhodnius prolixus (Hemiptera: Reduviidae). Infect Genet Evol, 2009. 9(2): p. 206–9. https://doi.org/10.1016/j.meegid.2008.10.017 PMID: 19061974

47. Van Oosterhout C., et al., MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes, 2004. 4: p. 535–538.

48. Peakall R. and Smouse P.E., GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics, 2012. 28(19): p. 2537–9. https://doi.org/10.1093/bioinformatics/bts460 PMID: 22820204

49. Sherwin W.B., et al., Measurement of biological information with applications from genes to landscapes. Mol Ecol, 2006. 15(10): p. 2857–69. https://doi.org/10.1111/j.1365-294X.2006.02992.x PMID: 16911206

50. Pritchard J.K., Stephens M., and Donnelly P., Inference of population structure using multilocus genotype data. Genetics, 2000. 155: p. 945–959. PMID: 10835412

51. Evanno G., Regnaut S., and Goudet J., Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol, 2005. 14(8): p. 2611–20. https://doi.org/10.1111/j.1365-294X.2005.02553.x PMID: 15969739

52. Watterson G.A., The homozygosity test of neutrality. Genetics, 1978. 88(2): p. 405–17. PMID: 17248803

53. Piry S., Luikart G., and Cornuet J.M., Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. J Hered, 1999. 90(4): p. 502–503.

54. Wilcoxon F., Individual comparisons by ranking methods. Biometrics Bulletin, 1945. 1(6): p. 80–83.

55. Kivelä M., Arnaud-Haond S., and Saramäki J., EDENetworks: a user-friendly software to build and analyse networks in biogeography, ecology and population genetics. Mol Ecol Resour, 2015. 15(1): p. 117–22. https://doi.org/10.1111/1755-0998.12290 PMID: 24902875

56. Rozenfeld A.F., et al., Network analysis identifies weak and strong links in a metapopulation system. Proc Natl Acad Sci U S A, 2008. 105(48): p. 18824–9. https://doi.org/10.1073/pnas.0805571105 PMID: 19022909

57. Beerli P. and Felsenstein J., Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proc Natl Acad Sci U S A, 2001. 98(8): p. 4563–8. https://doi.org/10.1073/pnas.081068098 PMID: 11287657

58. Beerli P. and Palczewski M., Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics, 2010. 185(1): p. 313–26. https://doi.org/10.1534/genetics.109.112532 PMID: 20176979

59. Abel G.J. and Sander N., Quantifying global international migration flows. Science, 2014. 343(6178): p. 1520–2. https://doi.org/10.1126/science.1248676 PMID: 24675962

60. Slatkin M., Isolation by distance in equilibrium and non-equilibrium populations. Evolution, 1993. 47(1): p. 264–279. https://doi.org/10.1111/j.1558-5646.1993.tb01215.x PMID: 28568097

61. Batista T.A. and Gurgel-Gonçalves R., Ecological niche modelling and differentiation between Rhodnius neglectus Lent, 1954 and Rhodnius nasutus Stål, 1859 (Hemiptera: Reduviidae: Triatominae) in Brazil. Mem Inst Oswaldo Cruz, 2009. 104(8): p. 1165–70. PMID: 20140378

62. Hasegawa M., Kishino H., and Yano T., Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol, 1985. 22(2): p. 160–74. PMID: 3934395

63. Slatkin M., A measure of population subdivision based on microsatellite allele frequencies. Genetics, 1995. 139: p. 457–462. PMID: 7705646

64. Khatchikian C.E., et al., Population structure of the Chagas disease vector Triatoma infestans in an urban environment. PLoS Negl Trop Dis, 2015. 9(2): p. e0003425. https://doi.org/10.1371/journal.pntd.0003425 PMID: 25646757

65. Piccinali R.V. and Gürtler R.E., Fine-scale genetic structure of Triatoma infestans in the Argentine Chaco. Infect Genet Evol, 2015. 34: p. 143–52. https://doi.org/10.1016/j.meegid.2015.05.030 PMID: 26027923

66. Diniz-Filho J.A., et al., Mantel test in population genetics. Genet Mol Biol, 2013. 36(4): p. 475–85. https://doi.org/10.1590/S1415-47572013000400002 PMID: 24385847

67. Leal I.R., et al., Changing the course of biodiversity conservation in the caatinga of northeastern *Brazil* Conservation Biology, 2005. 19: p. 701–706.

68. Laurance W.F., et al., Averting biodiversity collapse in tropical forest protected areas. Nature, 2012. 489: p. 290–294. https://doi.org/10.1038/nature11318 PMID: 22832582

69. Spielman D., Brook B.W., and Frankham R., Most species are not driven to extinction before genetic factors impact them. Proc Natl Acad Sci U S A, 2004. 101(42): p. 15261–4. https://doi.org/10.1073/pnas.0403809101 PMID: 15477597

70. Templeton A.R., et al., Disrupting evolutionary processes: the effect of habitat fragmentation on collared lizards in the Missouri Ozarks. Proc Natl Acad Sci U S A, 2001. 98(10): p. 5426–32. https://doi.org/10.1073/pnas.091093098 PMID: 11344289

71. Hartl D.L. and Clark A.G. *Principles of Population Genetics*. 4th ed. 1997, Sunderland: Sinauer Associates. 542.

72. Short K.H. and Petren K., Multimodal dispersal during the range expansion of the tropical house gecko Hemidactylus mabouia. Ecol Evol, 2011. 1(2): p. 181–90. https://doi.org/10.1002/ece3.18 PMID: 22393494

73. Luikart G. and Cornuet J.M., Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conserv Biol, 1998. 12(1): p. 228–237.

74. Zenger K.R., Richardson B.J., and Vachot-Griffin A.M., A rapid population expansion retains genetic diversity within European rabbits in Australia. Mol Ecol, 2003. 12(3): p. 789–94. PMID: 12675833

75. Keller L.F., et al., Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. Proc Biol Sci, 2001. 268(1474): p. 1387–94. https://doi.org/10.1098/rspb.2001.1607 PMID: 11429139

76. Saura M., et al., Genome-wide estimates of coancestry and inbreeding in a closed herd of ancient Iberian pigs. PLoS One, 2013. 8(10): p. e78314. https://doi.org/10.1371/journal.pone.0078314 PMID: 24205195

77. Kramer-Schadt S., et al., Analyzing the effect of stepping stones on target patch colonisation in structured landscapes for Eurasian lynx. Landsc Ecol, 2011. 26: p. 501–513.

78. Fahrig L., Effects of habitat fragmentation on biodiversity. Annu Rev Ecol Evol Syst, 2003. 34: p. 487–515.

79. Avise J.C., *Phylogeography: the history and formation of species*. 2000, Cambridge: Harvard University press. 447.

80. Carnaval A.C., et al., Prediction of phylogeographic endemism in an environmentally complex biome. Proc Biol Sci, 2014. 281(1792).

81. Vitorino L.C., et al., Demographical history and palaeodistribution modelling show range shift towards Amazon Basin for a Neotropical tree species in the LGM. BMC Evol Biol, 2016. 16(1): p. 213. https://doi.org/10.1186/s12862-016-0779-9 PMID: 27737632

82. De Oliveira P.E., Barreto A.M.F., and Suguio K., Late Pleistocene/Holocene climatic and vegetational history of the Brazilian caatinga: the fossil dunes of the middle São Francisco River. Palaeogeography, Palaeoclimatology, Palaeoecology, 1999. 152(3): p. 319–337.

83. Nyakaana S. and Arctander P., Population genetic structure of the African elephant in Uganda based on variation at mitochondrial and nuclear loci: evidence for male-biased gene flow. Mol Ecol, 1999. 8(7): p. 1105–15. PMID: 10447852

84. Scribner K.T., et al., Environmental correlates of toad abundance and population genetic diversity. Biological Conservation, 2001. 98(2): p. 201–210.

85. Doums C., Cabrera H., and Peeters C., Population genetic structure and male-biased dispersal in the queenless ant Diacamma cyaneiventre. Molecular Ecology, 2002. 11(11): p. 2251–2264. PMID: 12406237

86. Lima M.M., et al., Risk presented by Copernicia prunifera palm trees in the Rhodnius nasutus distribution in a Chagas disease-endemic area of the Brazilian northeast. Am J Trop Med Hyg, 2008. 79(5): p. 750–4. PMID: 18981517