## **ORIGINAL ARTICLE**

# Molecular phylogeography of the Chagas' disease vector *Triatoma infestans* in Argentina

AR Pérez de Rosas<sup>1</sup>, EL Segura<sup>2</sup> and BA García<sup>1</sup>

<sup>1</sup>Cátedra de Bioquímica y Biología Molecular, Facultad de Ciencias Médicas, Universidad Nacional de Córdoba, Pabellón Argentina 2° Piso, Ciudad Universitaria, Córdoba, Argentina and <sup>2</sup>Instituto Nacional de Parasitología Dr Mario Fatala Chabén, Avenida Paseo Colón 568, Buenos Aires, Argentina

*Triatoma infestans* is the main vector of Chagas' disease in South America between latitudes  $10^{\circ}$ S and  $46^{\circ}$ S. A multilocus microsatellite data set of 836 individuals from 27 populations of *T. infestans*, from all its range of distribution in Argentina, was analyzed. Our results favor the hypothesis of two independent migration events of colonization in Argentina and secondary contacts. The majority of the populations of the western provinces of Catamarca, La Rioja, San Juan and the west of Cordoba province, had almost no shared ancestry with the rest of the populations analyzed. Probably those populations, belonging to localities close to the Andean region, could have been established by the dispersal line of *T. infestans* that would have arrived to Argentina through the Andes, whereas most of the rest of the populations analyzed may have derived from the dispersal line of *T. infestans* in non-Andean lowlands. Among them, those from the provinces of Formosa, Chaco, Santiago del Estero and Santa Fe shared different percentages of ancestry and presented lower degree of genetic differentiation. The migratory movement linked to regional economies and possibly associated with passive dispersal, would allow a higher genetic exchange among these populations of *T. infestans*. This study, using microsatellite markers, provides a new approach for evaluating the validity of the different hypotheses concerning the evolutionary history of this species. Two major lineages of *T. infestans*, an Andean and non-Andean, are suggested. *Heredity* (2011) **107**, 71–79; doi:10.1038/hdy.2010.159; published online 12 January 2011

Keywords: Triatoma infestans; Chagas' disease vector; phylogeography; microsatellites; Argentina

## Introduction

Chagas' disease (American trypanosomiasis) is produced by infection with Trypanosoma cruzi (Chagas), which is transmitted by hematophagous insects of the subfamily Triatominae (Hemiptera: Reduviidae). The disease is a serious public health problem in Latin America, with about 12 million persons infected and around 90 million living in endemic areas (Schmunis, 1999). The subfamily Triatominae includes more than 130 species of insects, most of which can be either naturally or experimentally infected with T. cruzi. However, only a few of these species are considered important from an epidemiologically standpoint. Among them, Triatoma infestans (Klug) is widely distributed in South America between latitudes 10°S and 46°S, where it has been recognized as the main vector of Chagas' disease because it is one of the species with a high degree of adaptation to the domestic environment and therefore, one of the most closely associated with humans.

In Argentina, the National Chagas Control Program established in 1962 pursued the elimination of domestic and peridomestic populations of *T. infestans* by insecti-

Correspondence: Dr BA García, Cátedra de Bioquímica y Biología Molecular, Facultad de Ciencias Médicas, Universidad Nacional de Córdoba, Pabellón Argentina 2° Piso, Ciudad Universitaria, Córdoba 5000, Argentina.

E-mail: bgarcia@biomed.uncor.edu

Received 5 April 2010; revised 20 October 2010; accepted 15 November 2010; published online 12 January 2011

cide spraying. From 1962 to 2000 almost all the localities within the endemic area received insecticide treatment, but in most of them the entomological surveillance was not maintained after 20 years of the implemented programme (Segura *et al.*, 1999). As a consequence, resurgence of vector transmission of the disease was observed in the endemic region, where new acute cases since the year 2000 have been detected (Segura, 2007). The long-term effectiveness of the control campaigns is greatly dependent upon the vector population structure. The genetic analysis of these populations is important in this respect and may provide information on the evolutionary history and migration patterns of this species, which is of central importance for the development of control strategies.

Mitochondrial DNĂ (mtDNA) genes have been used in a number of population genetic analyses and have been recognized particularly useful for phylogeographical studies in many species of insects (Roderick, 1996); however, the maternally inherited markers analyzed until the present in *T. infestans*, either exhibited low levels of variation (Monteiro *et al.*, 1999; García *et al.*, 2003; Segura *et al.*, 2009) or have not been very useful for phylogeographical inferences of the Chagas' disease vector in Argentina (Piccinali *et al.*, 2009). Among the new DNA-based methods, microsatellites or simple sequence repeat markers have permitted greater resolution of genetic variation in populations (Bruford and Wayne, 1993; Schlötterer and Pemberton, 1994). In a previous work, 93 microsatellite loci were isolated from partial genomic libraries of T. infestans, 30 were amplified and 10 of these polymorphic microsatellite loci for which different allele types could be clearly resolved were selected for genotyping (García et al., 2004). With these nuclear markers, we evaluated the possible effect of eradication campaigns on the genetic structure of the vector populations, comparing levels of genetic variability in natural populations of T. infestans from areas with different elapsed periods since the last insecticide treatment and from areas that never received treatment. Surprisingly, the most of T. infestans populations from insecticide-treated localities seemed to have retained a substantial proportion of genetic diversity. As microgeographic genetic structure analyses supported the existence of subdivision in the populations of T. infestans, a population bottleneck would result in independent genetic-drift effects that could randomly preserve different combinations of alleles in each sub-population. These events followed by a rapid population growth could have preserved high levels of genetic diversity. Our studies supported the hypothesis of vector population recovery from survivors of the insecticide-treated areas, highlighting the value of population genetic analyses in assessing the effectiveness of Chagas' disease vector control programmes (Pérez de Rosas et al., 2007, 2008).

In this study, we propose a multilocus approach with 10 microsatellite neutral loci (García *et al.*, 2004), which have demonstrated to be a valuable molecular tool (Richer *et al.*, 2007; Pizarro *et al.*, 2008; Pérez de Rosas *et al.*, 2007, 2008) to infer the phylogeography and migration patterns from different *T. infestans* populations covering almost the entire species range in Argentina.

## Materials and methods

#### Specimens

A total of 27 populations of *Triatoma infestans* were sampled in Argentina. Sampling sites are shown in Table 1, 24 of them were reported previously (Pérez de Rosas *et al.*, 2007, 2008) and three (Salvador Mazza (19, Salta Province), and El Nochero and Morajú (17 and 18, Santa Fe Province)) were examined in this study. The specimens obtained in each locality were considered belonging to one population. The insects were captured in each locality from several houses or peridomiciliary sites, except in Santa Rosa (16, La Pampa province) where they were obtained from an attic invaded by pigeons. All sites were treated with insecticide after insect collection. The specimens were processed between 10 and 72 h after collection and their legs were preserved at -20 °C. In all, 10 microsatellite loci were analyzed.

#### DNA extraction and amplification

The DNA was extracted from frozen preserved legs of each bug using a simplified protocol of a phenolchloroform extraction procedure reported by Werman *et al.* (1990). The 10 primer pairs used for PCR were previously designed and tested for proper amplification of microsatellite DNA from whole-genomic DNA of *T. infestans* (García *et al.*, 2004). The forward primer from each primer set was 5'-fluorescent labeled with one of three dyes, 6-FAM, HEX or NED (Applied Biosystems, Foster City, CA, USA). PCR amplifications were carried out in a Hybaid thermal cycler (Omnigene, Hampton Hill, Middlesex, UK) in 25 µl of a solution containing

Table 1 Sampling site and sample size of Triatoma infestans from Argentina

Population number	Sampling site: locality, county and province	Latitude	Longitude	Sample size
1 <sup>b</sup>	El Jardín, General San Martín, La Rioja	31° 26′ S	65° 58′ W	25
2 <sup>b</sup>	Chancaní, Pocho, Córdoba	31° 26′ S	65° 28' W	34
3 <sup>b</sup>	Sauce Arriba, San Alberto, Córdoba	31° 54′ S	65° 10′ W	36
4 <sup>b</sup>	Caucete, Caucete, San Juan	31° 40′ S	68° 16' W	18
5 <sup>b</sup>	Sabagasta, Salavina, Santiago del Estero	28° 37′ S	63° 29' W	31
6 <sup>b</sup>	Vaca Human, Salavina, Santiago del Estero	28° 47′ S	63° 37' W	37
7 <sup>b</sup>	Taco Totoravo, Salavina, Santiago del Estero	28° 49′ S	63° 26' W	24
8 <sup>b</sup>	Vaca Huañuna, Figueroa, Santiago del Estero	27° 27′ S	63° 27' W	36
9 <sup>b</sup>	Tres Isletas, Maipú, Chaco	26° 21′ S	60° 26' W	36
10 <sup>b</sup>	Siete Árboles, General San Martín, Chaco	26° 24′ S	59° 25' W	74
11 <sup>b</sup>	Las Lomitas, Patiño, Formosa	24° 42′ S	60° 35' W	28
12 <sup>b</sup>	Palo Santo, Bermejo, Formosa	24° 19′ S	60° 55' W	36
13 <sup>b</sup>	Pozo del Zuri, Bermejo, Formosa	23° 41′ S	61° 09' W	36
14 <sup>b</sup>	La Esperanza, Bermejo, Formosa	23° 40′ S	61° 09' W	20
15 <sup>b</sup>	El Zapallo, General Paz, Corrientes	27° 41′ S	57° 37' W	29
16 <sup>b</sup>	Santa Rosa, Capital, La Pampa	36° 35′ S	64° 20' W	28
17 <sup>a</sup>	El Nochero, 9 de Julio, Santa Fe	29° 01′ S	61° 15′ W	25
18 <sup>a</sup>	Moraju, General Obligado, Santa Fe	28° 30′ S	59° 30' W	24
19 <sup>a</sup>	Salvador Mazza, San Martín, Salta	22° 03′ S	63° 42' W	11
20 <sup>b</sup>	San José Banda, Santa María, Catamarca	26° 41′ S	66° 02' W	19
21 <sup>b</sup>	Santa Rosa, Valle Viejo, Catamarca	28° 28′ S	65° 47′W	27
22 <sup>c</sup>	Saujil, Pomán, Catamarca	28° 12′ S	66° 14' W	18
23 <sup>c</sup>	Saujil, Tinogasta, Catamarca	27° 34′ S	67° 37' W	28
24 <sup>c</sup>	Fiambalá, Tinogasta, Catamarca	27° 41′ S	67° 37' W	31
25 <sup>c</sup>	Copacabana, Tinogasta, Catamarca	28° 12′ S	67° 29' W	19
26 <sup>c</sup>	Medanitos, Tinogasta, Catamarca	27° 31′ S	67° 36' W	70
27 <sup>c</sup>	Huillapima, Capayán, Catamarca	28° 44′ S	65° 59' W	36

<sup>a</sup>New populations added in this study.

<sup>b</sup>Populations previously analyzed (Pérez de Rosas *et al.*, 2007). <sup>c</sup>Populations previously analyzed (Pérez de Rosas *et al.*, 2008).

72

10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, each dNTP at 200  $\mu$ M, each primer at 1  $\mu$ M, genomic DNA (10– 50 ng), and 1 U of Amplitaq Gold (Applied Biosystems, Branchburg, NJ, USA). Thermal profiles consisted of an initial denaturation step at 94 °C for 5 min, followed by 30 cycles of 20 s at 94 °C (denaturation), 30 s at 55 °C (annealing), and 40 s at 72 °C (extension), with a final extension step of 15 min at 72 °C. Reaction products were visualized after electrophoresis on a 3% agarose gel (MetaPhor, Rockland, ME, USA). Size of the PCR products was estimated relative to an internal standard (GeneScan-500 ROX) in polyacrylamide gels using an ABI PRISM 377 automated DNA sequencer, GeneScan 3.1 and Genotyper 2.5 softwares (all from Perkin-Elmer Applied Biosystems, Foster City, CA, USA).

#### Analyses

The genotypes were assigned with the same criteria for all individuals comparing previous and new data sets. Linkage disequilibrium was tested between all pairs of loci and for all locations using the program FSTAT version 2.9.3 (Goudet, 2001). Genotypic disequilibrium was not apparent for any pair of loci according to a global test for each of the 45 different pairs of loci across all populations based on 900 permutations; therefore, we proceeded under the assumption of statistical independence between loci.

The degree of differentiation between and across all sites was determined using Weir and Cockerham (1984) estimator ( $\theta$ ) of Wright's F<sub>ST</sub>. Global and pairwise comparisons of  $\theta$  were calculated by the program FSTAT version 2.9.3 (Goudet, 2001). The significance of all  $\theta$  was tested using permutation. Bonferroni corrections (Rice, 1989) for multiple comparisons were applied when necessary. Standard errors of  $\theta$  were calculated by jackknifing over populations and loci, and a 95% confidence interval was generated by bootstrapping over loci. Significance of global  $\theta$  estimate was further evaluated with an exact G-test after 1000 randomizations of allele among sites (Goudet et al., 1996). Isolation by distance was examined by testing the correlation between  $\theta/(1-\theta)$  and the natural logarithm of pairwise geographical distances between populations (Rousset, 1997) in different geographical scales. The significance of the correlation was examined by a Mantel test with permutations as implemented in FSTAT version 2.9.3 (Goudet, 2001).

As population genetics data are known to be multivariate, because of the multilocus data set, univariate statistics such as F<sub>ST</sub> may not be sufficient to fully describe them. On the basis of this presumption, evidence for genetic structure, revealed by discontinuities among populations, was assessed by a factorial correspondence analysis performed using GENETIX version 4.03 (Belkhir et al., 2000). This technique was used to project populations into multivariate space and weigh and partition the influences of specific alleles on the global pattern of genetic variance. The method uses a similar strategy of a multivariate model adapted to the design of population genetics data to infer patterns of genetic differentiation, which is not completely independent from differentiation parameters such as  $F_{ST}$  (Guinand, 1996). This model is derived from Multiple Correspondence Analysis to study distribution profiles for a number of discrete variables (loci), which could have different states or modalities (alleles). The aim of the method is to find composite axes, which are a combination of the variables and optimize the differences between the analyzed populations.

A Bayesian approach implemented in the program STRUCTURE version 2 (Pritchard et al., 2000) was used to infer spatial population structure. The program STRUCTURE uses a Markov chain Monte Carlo approach to infer the number of populations (K) in a data set without prior information of the sampling locations. Considering the significant levels of genetic differentiation among T. infestans populations inferred previously (Pérez de Rosas et al., 2007, 2008), we assumed no admixture model, which is very appropriate to analyze discrete populations and it may be more successful in detecting subtle structure (Pritchard et al., 2000). On the other hand, uncorrelated allele frequencies within populations were assumed, given that Pritchard et al. (2000) suggested that this model is a convenient approximation for populations that are not too closely related and can produce accurate clustering. To select an appropriate K value for data analysis, a series of eight independent runs for each value of K between 1 and 10 were conducted. In each run, a burn-in period of 100000 and a 1000000 run length were performed. Ideally, a comparison of the replicates and the probability of data (called LnP(D) in STRUCTURE output) for each value of *K* allows the estimation of the point of inflection and the more likely numbers of clusters (Evanno et al., 2005).

An assignment analysis was performed to identify possible first-generation migrants as implemented by the program GENECLASS2 (Piry *et al.*, 2004). The program uses multilocus genotypes to identify putative immigrants within each population and the most likely source of these immigrants. The marginal probability of given individual multilocus genotype was compared with the distribution of marginal probabilities of randomly generated multilocus genotypes (10 000 replicates), when the *P*-value was <0.01 the individual was considered not belonging to the population. First-generation migrant test was carried out using Rannala and Mountain (1997) Bayesian individual assignment method and the Monte Carlo resampling method of Paetkau *et al.* (2004).

## Results

A multilocus microsatellite data set of 836 individuals from 27 populations of Triatoma infestans, from all its range of distribution in Argentina, was analyzed (see microsatellite genotype data set in the supplementary Table 1). Evidence of genetic differentiation among the 27 populations of T. infestans was examined using overall and pairwise comparisons of multilocus  $F_{ST}$  (using the estimator  $\theta$ );  $\theta$  values ranged from 0.012 between the populations of Palo Santo and La Esperanza (12 and 14, both from Formosa province) to 0.500 between the populations of Caucete (4, San Juan province) and El Zapallo (15, Corrientes province) (Table 2). The overall  $\theta$ value of 0.146 is significantly different from zero (P = 0.001), as are all pairwise  $\theta$  values (P < 0.05 and P < 0.01, FSTAT permutation procedure), except between the populations of Palo Santo (12) and La Esperanza (14). There was a significant association between geographical distance and genetic differentiation among sites,

Location	1	7	ŝ	4	5	9	~	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	4 2	2	
1																										
7	0.066																									
ю	0.074	0.029	I																							
4	0.195	0.139	0.147																							
Ŋ	0.109	0.087	0.091	0.206																						
6	0.094	0.076	0.078	0.206	0.025																					
7	0.083	0.064	0.062	0.197	0.026	0.019																				
8	0.197	0.139	0.162	0.281	0.049	0.079	0.073																			
6	0.202	0.146	0.157	0.262	0.047	0.087	0.076	0.038																		
10	0.305	0.213	0.231	0.359	0.133	0.168	0.135	0.102	0.070																	
11	0.194	0.163	0.170	0.294	0.057	0.086	0.083	0.025	0.026	0.110	I															
12	0.138	0.109	0.127	0.231	0.031	0.050	0.040	0.024	0.042	0.139	0.030															
13	0.167	0.139	0.152	0.279	0.046	0.060	0.052	0.024	0.048	0.136	0.024	0.015														
14	0.122	0.109	0.111	0.232	0.031	0.038	0.041	0.037	0.048	0.158	0.027	0.012 (	0.017													
15	0.390	0.297	0.316	0.500	0.217	0.234	0.209	0.157	0.173	0.134	0.197	0.196 (	).174 (	0.250												
16	0.192	0.111	0.133	0.279	0.121	0.118	0.121	0.170	0.171	0.238	0.198	0.146 (	0.175 (	0.151 (	0.360											
17	0.111	0.093	0.091	0.202	0.081	0.071	0.051	0.129	0.127	0.231	0.144	0.093 (	).119 (	) 260.0	0.289 (	0.155										
18	0.154	0.119	0.123	0.234	0.051	0.083	0.077	0.077	0.061	0.152	0.093	0.049 (	) 770.0	0.070.0	0.283 (	).177 (	0.119									
19	0.115	0.106	0.093	0.227	0.089	0.055	0.064	0.173	0.173	0.260	0.176	0.123 (	).152 (	0.103 (	0.410 (	).162 (	0.109 (	0.165								
20	0.170	0.128	0.134	0.271	0.082	0.109	0.098	0.084	0.105	0.223	0.113	0.074 (	).087 (	0.075 (	0.346 (	).181 (	0.129 (	0.099	0.205							
21	0.116	0.101	0.104	0.217	0.052	0.058	0.051	0.086	0.086	0.188	0.095	0.050 (	).070 (	0.048 (	0.288 (	).138 (	).087 (	0.072	0.117 (	0.075						
22	0.121	0.112	0.105	0.223	0.119	0.090	0.073	0.158	0.151	0.261	0.166	0.105 (	).127 (	080.0	0.351 (	).187 (	0.095 (	0.146	0.115 (	0.147 0	.057					
23	0.210	0.138	0.156	0.282	0.086	0.099	0.096	0.108	0.131	0.183	0.164	0.101 (	).147 (	0.129 (	0.260 (	).164 (	0.133 (	0.110	0.172 (	0.172 0	.119 0	.164	I			
24	0.264	0.218	0.236	0.397	0.135	0.139	0.153	0.160	0.175	0.233	0.211	0.151 (	0.176 (	0.161 (	0.302 (	).231 (	0.182 (	0.202	0.241 (	0.214 0	.173 0	.210 0.	127 -	1		
25	0.182	0.161	0.179	0.297	0.159	0.132	0.134	0.238	0.235	0.331	0.270	0.193 (	).233 (	0.195 (	).429 (	).222 (	0.148 (	0.227	0.180 (	0.254 0	.180 0	.170 0.	154 0.7	22	I	
26	0.166	0.122	0.127	0.230	0.080	0.109	0.100	0.112	0.110	0.193	0.142	0.100 (	).124 (	0.107 (	).262 (	).151 (	).109 (	0.098 (	0.175 (	0.104 0	.072 0	.133 0.	100 0.	[61 0.1	65	
27	0.138	0.074	0.096	0.240	0.155	0.120	0.113	0.205	0.201	0.304	0.224	0.163 (	).183 (	0.154 (	).364 (	).199 (	).151 (	0.166	0.169 (	0.154 0	.137 0	.136 0.	202 0.3	284 0.1	60	Ó



**Figure 1** Isolation by distance among sampling sites. Pairwise estimates of  $\theta/(1-\theta)$  are plotted against the corresponding natural logharitm of the geographical distances between sites (Rousset, 1997). (a) All sampling site pairs considered (Mantel r = 0.405, P = 0.0005). (b) Sampling site pairs at distances ranging from 0–150 km (Mantel r = 0.163, P = 0.51).

suggesting isolation by distance (Mantel r = 0.405, P = 0.0005, Figure 1a). However, considering different geographical scales, no significant correlation between genetic and geographical pairwise distance was found among populations pairs situated from 0–150 km (Mantel r = 0.163, P = 0.51, Figure 1b).

The results of the factorial correspondence analysis are given in Figure 2, which present the projection of the populations on the plane defined by the first two axes of the analysis. The distribution of the populations along the first axis separated most of the populations belonging to localities from the west of Argentina (provinces of La Rioja (1), San Juan (4), Catamarca (21, 22, 23, 25, 26, 27) and the west of Córdoba province (2, 3)) from those belonging to the provinces of Santa Fe (17, 18), Salta (19), Santiago del Estero (5–8), Formosa (11–14), Chaco (9, 10), Corrientes (15), La Pampa (16) and two localities of Catamarca province (20, 24). The position of populations along the first correspondence axis reflected their geographical position from east to west along a longitudinal transect, as indicated by the significant correlation between the axis values and the longitudinal geographical coordinates of the different localities (r = 0.68, P = 0.0005). The negative part of the first axis clustered 17 populations distributed in a wide geographical range of Argentina; they were closely situated in the plane, except the populations from El Nochero (17, Santa Fe province) and Salvador Mazza (19, Salta province). On the other hand, the majority of the western populations placed at the positive end of the first axis were distributed more separately in the plane, suggesting a higher degree of genetic differentiation among the populations of this region with exception of two populations from Catamarca province, Saujil (23,



**Figure 2** Positions of 27 populations of *Triatoma infestans* in multivariate space defined by factorial correspondence analysis axes 1 and 2. Numerical code for populations corresponds to Table 1.

Tinogasta) and Medanitos (26), which were found superimposed. On the positive side of the first axis, the populations were separated along the second correspondence axis; El Jardín (1, La Rioja province), Caucete (4, San Juan province) and Chancaní and Sauce Arriba (2 and 3, Córdoba province) were distributed over the positive part of the second axis and the populations belonging to Catamarca province were on the negative part of this axis.

Using the program STRUCTURE version 2 (Pritchard et al., 2000), seven different genetic clusters could be identified (Figure 3). Groupings obtained with K > 7 were not repeatable between different runs, although the estimated probability of the data slightly improved with higher K (see supplementary Figure 1). Moreover, from K > 7 the variation of the probability among the replicates increased suggesting instability among the replicates and empty groups were obtained frequently, a phenomenon that happens when *K* is higher than the number of 'real' groups. Thus, the point of inflection was clearly at K = 7. As suggested by the Bayesian approach, there are six distinct, almost homogeneous clusters (A, B, C, D, E and F) distributed across specific geographical locations. Although 11 populations correspond to the six different, nearly homogeneous clusters (more than 80% of ancestry shared among individuals), 16 populations were more heterogeneous presenting a mix of different clusters. Most of the first 11 populations form three of the almost homogeneous clusters (A, B and C) detected in the provinces of San Juan (4), La Rioja (1) and Catamarca (23-27). In Catamarca, the samples belonging to the localities of Medanitos (26), Fiambalá (24) and Saujil (23, Tinogasta) form the cluster A; the samples from San José Banda (20), Santa Rosa (21) and Saujil (22, Pomán) share with this cluster around 37,26 and 61% of the ancestry, respectively. In the same province, the populations belonging to Copacabana (25) and Huillapima (27) form a separate cluster (B). On the other hand, the populations from Caucete (4, San Juan province) and El Jardín (1, La Rioja province) form a third almost homogeneous cluster (C), sharing the two populations analyzed from the west of Córdoba province (Chancaní, 2; Sauce Arriba, 3) a high level of the ancestry (71 years 74%, respectively) with this cluster. The populations belonging to the other three almost homogeneous clusters are geographically distant; Salvador Mazza (19, Salta province), Siete Árboles (10, Chaco province) and El Zapallo (15,

Molecular phylogeography of Triatoma infestans AR Pérez de Rosas et al



**Figure 3** Pie chart of combined genetic ancestries of all individuals sampled in each population, as obtained by STRUCTURE (Pritchard *et al.*, 2000). The seven different colors correspond to seven different genetic clusters, which are indicated with A, B, C, D, E, F and G. In italic the capital of Argentina as point of reference. Numerical code for populations corresponds to Table 1. A full color version of this figure is available at the *Heredity* journal online.

Corrientes province) and Santa Rosa (16, La Pampa province) form the clusters D, E and F, respectively. The seventh cluster (G) was found to contribute in different proportions to populations that presented a higher degree of heterogeneity, principally those from localities belonging to the provinces of Formosa (11–14), Chaco (9) and Santiago del Estero (5–8).

Comparison of allelic richness ( $R_A$ , number of alleles corrected for sample size) and  $F_{ST}$  values (a measure of the variance of gene frequencies) was performed between the group of the heterogeneous populations and the group of populations that correspond to the almost homogeneous clusters. The results revealed a significant highest level of allelic richness in the first group ( $R_A = 5.04$  and  $R_A = 3.81$ , respectively; P = 0.01) and of genetic differentiation in the second group ( $\theta = 0.084$  and  $\theta = 0.225$ , respectively; P = 0.001).

Results of the first-generation migrant test are shown in Table 3. The first-generation migrant test identifies immigrants within each population and the most likely source of these immigrants. A total of 41 putative immigrants were identified originating from 17 of the localities and migrating to 24 of the localities. The localities within Santiago del Estero and Formosa were the places of origin of the majority of the individuals identified as immigrants in distant localities in Argentina;

Molecular phylogeography of Triatoma infestans AR Pérez de Rosas et al

 $\overline{0}$ 

0 0 0 0 0 1 0 0 0 35

lab	le 3	Resu	lts of	first-	gene	ratior	i mig	rant i	test fo	or 27	popu	llatio	ns of	Iriato	oma 11	nfesta	ns us	ing C	enec	lass 2	2.0 (Pi	ry et	al., 2	004)			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	23	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	30	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	$\overline{0}$	2	34	0	0	$\overline{0}$	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	$\overline{0}$	0	0	29	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
6	0	0	1	0	1	34	0	0	0	0	0	1	0	0	0	0	0	$\overline{0}$	0	0	0	0	0	0	0	0	0
7	0	0	$\overline{0}$	0	$\overline{0}$	0	23	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	$\overline{0}$	0	0	0	35	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	1	1	0	71	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	$\overline{0}$	0	0	0	27	0	$\overline{0}$	1	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	1	0	0	0	0	0	34	1	$\overline{0}$	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	$\overline{0}$	0	0	0	0	0	1	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	$\overline{0}$	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	23	0	0	0	1	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	22	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	10	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	1	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	1	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	17	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	1	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	1	0
26	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	65	0

Ta

Rows indicate sampling site, columns indicate potential source of the insects. Bold underlined numbers are putative immigrants (P < 0.01).

while other localities as, for example, from Catamarca province, principally interchanged migrants among them, as well as among the populations from La Rioja and Córdoba provinces.

## Discussion

27

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Genetic drift and limited gene flow seem to have generated a substantial degree of genetic differentiation among populations of Triatoma infestans ( $\theta = 0.146$ , P = 0.001) as suggested by previous studies (Pérez de Rosas et al., 2007, 2008; Pizarro et al., 2008). This is not surprising, as the limited dispersal described for T. infestans (Schofield 1988) would tend to increase genetic differentiation through the process of genetic drift. Moreover, significant isolation by distance among sampling sites, with nearby sites apparently exchanging more genes than distant ones, was also confirmed (Figure 1a). However, considering distance ranges between population pairs smaller than 150 km, significant association between geographical distance and genetic differentiation among sites was not detected (Figure 1b). The results for sites in this smaller geographical scale, as well as for some sites in a bigger geographical scale (Figure 1a), which exhibited deviation from a pattern of isolation by distance, suggest that allele frequencies in each of these sites drift independently of the geographical distances separating them; probably factors other than geographical distance could be involved in generating the structure of populations. The majority of T. infestans populations studied belong to areas with different elapsed periods since the last insecticide treatment and only three are from areas that never received treatment (Pérez de Rosas et al., 2007, 2008). The possibility is that local differentiation by genetic drift enhanced by the reduction of population size by insecticide treatment could create significant structuring independent of distance. In this regard, the high degree of divergence between some geographically close populations may be because of a short time as the last insecticide intervention in at least one of the locations, which would limit the time available for gene flow to diminish differentiation. Therefore, as a consequence of insecticide treatment, it seems that genetic drift may have enhanced differentiation among populations (Pérez de Rosas et al., 2007).

The first-generation migrant assignment test suggests putative immigrants in 24 of the 27 populations (Table 3). However, it is important to distinguish between current and historical levels of gene flow. In this regard, the estimates of gene flow and migration rates via Wright's  $F_{ST}$  statistic (a measure of the variance of gene frequencies between populations) represent evolutionary averages, rather than contemporaneous patterns of dispersal (Bossart and Prowell, 1998; Palsboll, 1999), whereas assignment methods have the potential to provide direct estimates of real-time dispersal through the detection of individual immigrants (Paetkau et al., 2004). Considering that historical events can leave imprints on the distribution of genetic variability (Ibrahim et al., 1996) that can last for many generations (Nei et al., 1975), and the important role of genetic drift generating genetic differentiation among populations, it seems that the degree of current dispersion is not enough to produce a homogeneous metapopulation.

Pritchard et al. (2000) highlight the strengths of the Bayesian approach in evolutionary studies of genetic relationships among populations, compared with the summary statistics such as F<sub>ST</sub>. Both approaches suggest that populations of *T. infestans* are highly structured and are reinforced by the results of the factorial correspondence analysis. Three different nearly homogeneous clusters (A, B and C) were detected in the west of Argentina, which comprise seven *T. infestans* populations from the provinces of Catamarca (23–27), San Juan (4) and La Rioja (1), as well as two populations from the west of Córdoba province (2, 3) that share a high level of the ancestry with cluster C (Figure 3). These three groups of populations practically do not share ancestry among them, suggesting a lower level of gene flow and consequently a higher population differentiation by drift in this region. Besides, these populations belonging to localities geographically close to the Andean highlands almost do not share ancestry with the rest of the populations analyzed. Panzera et al. (2004) identified two chromosomic allopatric groups in T. infestans populations, named Andean and non-Andean. These authors detected differences in heterochromatin that have been considered the main cause of the variation in the DNA content between both groups, with  $\sim 30\%$ more hetrochromatin in the Andean insects. On the basis of the observation that Andean samples came from geographical regions generally above 1800 m, whereas non-Andean populations were mainly from localities below 500 m, these authors suggested that heterochromatin and DNA variation reflect adaptative genomic changes that contribute to the ability of T. infestans to survive and reproduce in environments with different altitudes; populations with large genome would be better adapted to highlands (Andean regions), whereas populations with smaller genomes would do better in lowlands (non-Andean regions). Moreover, phylogenetic inferences carried out, using nuclear rDNA (ITS-1 and ITS-2) sequences, supported the previously suggested origin of T. infestans in Bolivia highlands (Schofield, 1988; Giordano et al., 2005; Noireau et al., 2005) and two different dispersal lines, one throughout Andean regions of Bolivia and Perú, and the second in non-Andean lowlands of Chile, Paraguay, Argentina, Uruguay and Brazil (Bargues et al., 2006). Considering these findings, it is probable that the populations from the western area of Argentina mentioned above, most of which are located between 1000 and 1700 m, could have been established from the dispersal line of T. infestans that would have arrived to Argentina through the Andes. Presumably, the presence in this area of populations that share some degree of ancestry with eastern populations (for example, the two populations from Córdoba province) suggests a process of secondary contact between populations from the different dispersal lines.

On the other hand, probably most of the other populations analyzed may have derived from the dispersal line of *T. infestans* in non-Andean lowlands associated to human migratory movements. These populations, principally those from localities belonging to the provinces of Formosa (11–14), Chaco (9), Santiago del Estero (5–8) and Santa Fe (17, 18), which presented lower degree of genetic differentiation, showed higher heterogeneity (mix of different clusters) sharing different percentages of ancestry (Figure 3). The human internal migratory movement linked to regional economies, possibly associated with passive dispersal, would allow to maintain a higher genetic exchange between these

Heredity

populations of *T. infestans*; for example, the migratory movement of people from provinces as Santiago del Estero originated by the cotton crop in Chaco, Formosa and North of Santa Fe. Consistent with these observations, the results of the first-generation migrant test suggested that principally the localities belonging to Santiago del Estero province and those from Formosa province were the places of origin of the majority of the individuals identified as immigrants in different localities of Argentina (Table 3).

The almost homogeneous cluster detected in Santa Rosa (16, La Pampa province), unlike the other six, did not contribute >18% to any of the other populations (Figure 3). As the insects were captured in each place from several houses or peridomiciliary sites, except the sample from Santa Rosa that was obtained from an attic invaded by pigeons, the results obtained in this locality are compatible with a founder event. On the other side, the populations from three localities of Santiago del Estero province (5, Sabagasta; 6, Vaca Human; 7, Taco Totorayo), El Nochero (17, Santa Fe province) and Santa Rosa (21, Catamarca province) share a high degree of the ancestry with the geographical distant population from Salvador Mazza (19, Salta province). The analysis of antennal phenotypes in populations from Argentina, Paraguay and Bolivia was carried out by Hernández et al. (2008), showed the existence of two groups: (a) the group of populations analyzed from the Andean valleys of Bolivia, (b) the group that comprised the populations from the Gran Chaco and the intermediate valleys. In agreement with our results, within the second group a close relationship among samples from Salvador Mazza, Santiago del Estero and Catamarca provinces (the study did not include samples from Santa Fe province) was detected using the antennal phenotype data (Hernández et al., 2008).

In summary, it seems probable that the evolutionary history of *T. infestans* was marked by two independent events of colonization and secondary contacts. Comparisons of the populations analyzed in this study with populations of the Andean and non-Andean groups determined in previous studies (Panzera *et al.*, 2004; Bargues *et al.*, 2006) should be performed to test the hypothesis of two independent migration events of colonization in Argentina. This study, using microsatellite markers, provides a new approach for evaluating the validity of the different hypotheses concerning the evolutionary history of this species.

## Conflict of interest

The authors declare no conflict of interest.

## Acknowledgements

We thank Dr A Blanco for critical revision of the manuscript and O Fusco for technical assistance. Research was supported by the grants from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR), the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) of Argentina, and the Secretaría de Ciencia y Tecnología de la Universidad Nacional de Córdoba. A R Pérez de Rosas is a Fellow, and E L Segura and B A García are Career Investigators of CONICET.

## References

Bargues MD, Klisiowicz DR, Panzera F, Noireau F, Marcilla A, Perez R *et al.* (2006). Origin and phylogeography of the Chagas disease main vector *Triatoma infestans* based on nuclear rDNA sequences and genome size. *Infect Genet Evol* **6**: 46–62.

- Belkhir K, Borsa P, Chikhi L, Goudet J, Bonhomme F (2000). GENETIX 403, WindowsTM Software for Population Genetics. Laboratoire Génome, Populations, Interactions, CNRS UPR 9060, Université de Montpellier II, Montpellier, France.
- Bossart JL, Prowell DP (1998). Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends Ecol Evol* **13**: 202–205.
- Bruford MW, Wayne RK (1993). Microsatellites and their application to population genetic studies. *Curr Opin Genet Dev* **3**: 939–943.
- Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14: 2611–2620.
- García BA, Manfredi C, Fichera L, Segura EL (2003). Variation in mitochondrial 12S and 16S ribosomal DNA sequences in natural populations of *Triatoma infestans* (Hemiptera: Reduviidae). *Am J Trop Med Hyg* **68**: 692–694.
- García BA, Zheng L, Pérez de Rosas AR, Segura EL (2004). Isolation and characterization of polymorphic microsatellite loci in the Chagas' disease vector *Triatoma infestans* (Hemiptera: Reduviidae). *Mol Ecol Notes* **4**: 568–571.
- Giordano R, Pizarro Cortez JC, Paulk S, Stevens L (2005). Genetic diversity of *Triatoma infestans* (Hemiptera: Reduviidae) in Chuquisaca, Bolivia based on the mitochondrial cytochrome b gene. *Mem Inst Oswaldo Cruz* **100**: 753–760.
- Goudet J (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/popgen/softwares/fstat.htm. *Updated from Goudet* (1995).
- Goudet J, Raymond M, De Meeus T, Rousset F (1996). Testing differentiation in diploid populations. *Genetics* **144**: 1931–1938.
- Guinand B (1996). Use of a multivariate model using allele frequency distributions to analyse patterns of genetic differentiation among populations. *Biol J Linn Soc Lond* **58**: 173–195.
- Hernández L, Abrahan L, Moreno M, Gorla D, Catalá S (2008). Phenotypic variability associated to genomic changes in the main vector of Chagas disease in the southern cone of South America. *Acta Trop* **106**: 60–67.
- Ibrahim KM, Nichols RA, Hewitt GM (1996). Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity* 77: 282–291.
- Monteiro FA, Pérez R, Panzera F, Dujardin JP, Galvão C, Rocha D *et al.* (1999). Mitochondrial DNA variation of *Triatoma infestans* populations and its implication on the specific status of *T melanosoma*. *Mem Inst Oswaldo Cruz* **94**: 229–238.
- Nei M, Maruyama T, Chakraborty R (1975). The bottleneck effect and genetic variability in populations. *Evolution* **29**: 1–10.
- Noireau F, Rojas Cortez MG, Monteiro FA, Jansen AM, Torrico F (2005). Can wild *Triatoma infestans* foci in Bolivia jeopardize Chagas disease control effort? *Trends Parasitol* **21**: 7–9.
- Paetkau D, Slade R, Burden M, Estoup A (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Mol Ecol* 13: 55–65.
- Palsboll PJ (1999). Genetic tagging: contemporary molecular ecology. *Biol J Linn Soc* 68: 3–22.
- Panzera F, Dujardin JP, Nicolini P, Caraccio MN, Rose V, Tellez T et al. (2004). Genomic changes of Chagas disease vector, South América. Emerg Infect Dis 10: 438–446.

- Pérez de Rosas AR, Segura EL, Fichera L, García BA (2008). Macrogeographic and microgeographic genetic structure of the Chagas' disease vector *Triatoma infestans* (Hemiptera: Reduviidae) from Catamarca, Argentina. *Genetica* **133**: 247–260.
- Pérez de Rosas AR, Segura EL, García BA (2007). Microsatellite analysis of genetic structure in natural *Triatoma infestans* (Hemiptera: Reduviidae) populations from Argentina: its implication in assessing the effectiveness of Chagas' disease vector control programmes. *Mol Ecol* **16**: 1401–1412.
- Piccinali RV, Marcet PL, Noireau F, Kitron U, Gurtler RE, Dotson EM (2009). Molecular population genetics and phylogeography of the Chagas disease vector *Triatoma infestans* in South America. *J Med Entomol* **46**: 796–809.
- Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A (2004). GENECLASS2: a software for genetic assignment and first-generation migrant detection. *J Hered* **95**: 536–539.
- Pizarro JC, Gilligan LM, Stevens L (2008). Microsatellites reveal a high population structure in *Triatoma infestans* from Chuquisaca, Bolivia. *PLoS Negl Trop Dis* 2: e202.
  Pritchard JK, Stephens M, Donnelly P (2000). Inference
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Rannala B, Mountain JL (1997). Detecting immigration by using multilocus genotypes. Proc Natl Acad Sci USA 94: 9197–9201.
- Rice WR (1989). Analysing tables of statistical tests. *Evolution* **43**: 223–225.
- Richer W, Kengne P, Rojas Cortez M, Perrineau MM, Cohuet A, Fontenille D *et al.* (2007). Active dispersal by wild *Triatoma infestans* in the Bolivian Andes. *Trop Med Int Health* **12**: 759–764.
- Roderick GK (1996). Geographic structure of insect populations: gene flow, phylogeography, and their uses. *Annu Rev Entomol* **41**: 325–352.
- Rousset F (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.
- Schlötterer C, Pemberton J (1994). The use of microsatellites for genetic analysis of natural populations. In: Schierwater B, Streit B, Wagner GP, DeSalle R (eds) *Molecular Ecology and Evolution: Approaches and Applications*. Birkhauser Verlag, Basel, Switzerland. pp 203–214.
- Schmunis GA (1999). Iniciativa del Cono Sur. In: Proceedings of the Second International Workshop on Population Biology and Control of Triatominae, Santo Domingo de los Colorados, Ecuador. INDRE, Mexico City. pp 26–31.
- Schofield CJ (1988). Biosystematics of the Triatominae. In: Service MW (ed) *Biosystematics of Haematophagous Insects, Systematics Association Special*. Volume N° 37 Clarendon, Oxford. pp 285–312.
- Segura EL (2007). Presente y futuro de la enfermedad de Chagas. Su conocimiento y control. In: *La enfermedad de Chagas, a la puerta de los 100 años del conocimiento de una endemia americana ancestral.* Organización Panamericana de la Salud y Fundación Mundo Sano, Buenos Aires. pp 51–54.
- Segura EL, Sosa Estani S, Esquivel ML, Gomez A, Salomón OD (1999). Control de la transmisión de *Trypanosoma cruzi* en la Argentina. *Medicina* 59: 91–96.
- Segura EL, Torres AG, Fusco O, García BA (2009). Mitochondrial 16S DNA variation in populations of *Triatoma infestans* from Argentina. *Med Vet Entomol* **23**: 34–40.
- Weir BS, Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- Werman SD, Davidson EH, Britten RJ (1990). Rapid evolution in a fraction of the *Drosophila* nuclear genome. J Mol Evol 30: 281–289.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)