MOLECULAR ECOLOGY

Molecular Ecology (2012) doi: 10.1111/j.1365-294X.2012.05598.x

The evolution of a highly speciose group in a changing environment: are we witnessing speciation in the Iberá wetlands?

MARIA JIMENA GÓMEZ FERNÁNDEZ,*† OSCAR E. GAGGIOTTI‡ and PATRICIA MIROL*†

*Museo Argentino de Ciencias Naturales ''Bernardino Rivadavia'', Angel Gallardo 470, C1405DJR, Ciudad de Buenos Aires, Argentina, †Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ciudad de Buenos Aires, Argentina, ‡Laboratoire d'Ecologie Alpine (LECA), CNRS UMR 5553, Université Joseph Fourier, 38041 Grenoble Cedex 09, France

Abstract

Delimiting species is very conflicting in the case of very young taxa that are in the process of diversification, and even more difficult if the species inhabit a heterogeneous environment. In this case, even population delimitation is controversial. The South American genus of subterranean rodents *Ctenomys* is highly speciose, with 62 species that appeared in the lapse of 3 Myr. Within the genus, the 'perrensi' group, formed by three named species and a group of forms of unknown taxonomic status, inhabits the Iberá wetland, in northern Argentina. Almost every locality shows a particular chromosomal complement. To understand the relationships and the evolutionary process among species and populations, we examined mitochondrial DNA sequences and microsatellite genotypes. We found an isolation-by-distance pattern with evidence of cluster-like behaviour of the system. The mitochondrial DNA network revealed two different groups, separated by one of the main rivers of the region. Clustering methods delimited 12 different populations and five metapopulation lineages that seem to be evolving independently. We found evidence of ancient migration among localities at the centre of the distribution but no signals of current migration among the 12 delimited clusters. Some of the genetic clusters found included localities with different chromosomal numbers, which points to the existence of gene flow despite chromosomal variation. The evolutionary future of these five lineages is controlled by the dynamics of their habitat: if stable, they may become distinct species; otherwise, they may collapse into a hybrid swarm, forming a single evolving metapopulation.

Keywords: *Ctenomys*, fragmented habitat, metapopulation lineages, speciation

Received 17 August 2011; revision received 14 March 2012; accepted 23 March 2012

Introduction

Speciation in a heterogeneous environment, where connections between populations fluctuate over time, may lead to uncertain boundaries between species. In these kinds of environments, natural populations are dynamic in time and space: they change in size, density and location, and they also can join with others or split into several populations (Hey & Machado 2003). This non-equilibrium dynamics is further increased by current habitat fragmentation, mainly due to anthropic disturbance. Thus, the study of evolutionary processes driving patterns of species or population divergence has to consider the geographic context of such divergence, otherwise interpretations could fail to recognize historical population associations (Knowles & Carstens 2007b). Moreover, in the case of recently diverged taxa, populations may have not been isolated long enough to accumulate differences, in which case the most probable gene tree may not match the actual species tree (Degnan & Rosenberg 2006). Indeed, the boundaries between species are not clearly reflected in a gene tree until the ancestral polymorphism has been fully sorted

Correspondence: Maria Jimena Gómez Fernández, Fax: +541149824494; E-mail: mjgfernandez@yahoo.com.ar

(Knowles & Carstens 2007a); a process that can take a significant amount of time.

A typical case of uncertainty on species and even population boundaries occurs in the *Ctenomys* groups living in the Iberá marsh, in the province of Corrientes in northeastern Argentina. The wetland occupies most of the province (more than 14 000 km^2) and consists of a vast mosaic of marshes, swamps and lagoons. Permanent or temporary water bodies dominate 90% of the wetland and in the remaining dry areas, there is a predominance of sandy soils, forming elongated sandy hills (Ferrati *et al.* 2005). The wetland is surrounded by three rivers, the Paraná Alto River, the Paraná Medio River and the Uruguay River, with no surface connection to any of them. Alongside the Brasilian marshland, they form the second largest wetland in the world, part of a hydrographic system in which there is a highly diverse subtropical ecosystem.

It is on those sandy hills where subterranean rodents of the genus *Ctenomys* reside (Reig *et al.* 1990; Cook & Lessa 1998; Lessa & Cook 1998; Lessa 2000; Castillo *et al.* 2005). The genus is highly speciose: it arose during the late Miocene or early Pliocene (Reig *et al.* 1990; Verzi 2002) and diversified into 62 living species, representing 45% of all species of subterranean rodents (Lacey *et al.* 2000). Tuco-tucos have limited dispersal capacity (Busch *et al.* 2000) and they are commonly grouped in small demes with low genetic variation and high interpopulation divergence (Wlasiuk *et al.* 2003). They also have small effective population sizes, socially structured mating systems and are territorial. In the Iberá, variations in the intensity and spatial extent of flooding events determine the availability of suitable habitat for tuco-tucos: because of their fossorial way of life, they need well-drained and aerated soils. Hence, their patchy populations can be connected or isolated at different times depending on the frequency and intensity of rainfall in a given year.

Three named species inhabit the wetland: *Ctenomys roigi* and *Ctenomys perrensi*, located at the northwest of the Iberá next to the Paraná River, and *Ctenomys dorbignyi*, which is found forming two separate nucleus, north and south of the distribution. Together with several other forms of the genus, whose taxonomical status has not yet been determined, they constitute the 'perrensi' group. The populations of the unnamed forms are located at the centre and west of the province. Previous studies on the 'perrensi' lineages showed that the three named species differ in their chromosomal (2n) and fundamental (FN) numbers, being 2n/FN = 48/76 in *C. roigi*, 2n/FN = 50/80 in *C. perrensi* and 2n = 70/80 in *C. dorbignyi*. The rest of the 'perrensi' complex exhibits 2n ranging from 42 to 66, and FN from 72 to 80 (Ortells 1995; García *et al.* 2000; Giménez *et al.* 2002). This high karyotypic variability, together with the distribution of their populations in isolated demes, suggests that chromosomal evolution is an ongoing active and recurrent process in these genomes. On the basis of chromosomal numbers and geographic distribution, Giménez *et al.* (2002) postulated the existence of two new species: species 'α' constituted by populations located at the northern part of the wetland and with the lowest chromosomal/fundamental numbers 42–46/72–74, and species 'β' at the centre of the distribution and with chromosomal/fundamental numbers 54–66/80. However, a molecular phylogeny based on *cytochrome b* sequences failed to detect any monophyletic groups among the established and newly proposed species (Giménez *et al.* 2002). This result is consistent with other studies of the whole genus that demonstrate a lack of differentiation among different lineages of the whole genus found when using several mitochondrial and nuclear intron sequences (Lessa & Cook 1998; Mascheretti *et al.* 2000; Castillo *et al.* 2005; Parada *et al.* 2011).

A more recent study based on 16 microsatellite loci (Mirol *et al.* 2010) uncovered a complex scenario in which *C. dorbignyi* is formed by two evolutionary separated lineages, *C. roigi* is a well-defined entity that contributes genes to a group of *Ctenomys* sp. populations and *C. perrensi* is separated into two different lineages, each one associated with a group of derived populations. The authors propose the existence of eight groups in the area: 1—*C. roigi*, 2—'Manantiales' group, 3—*C. perrensi* North, 4—*C. perrensi* South, 5—*C. dorbignyi* North, 6—*C. dorbignyi* South, 7—'San Miguel' group and 8—'Tacuaritas' group (Fig. 1 in Mirol *et al.* 2010). This study suggests hybridization among the named species, but could not fully resolve the pattern

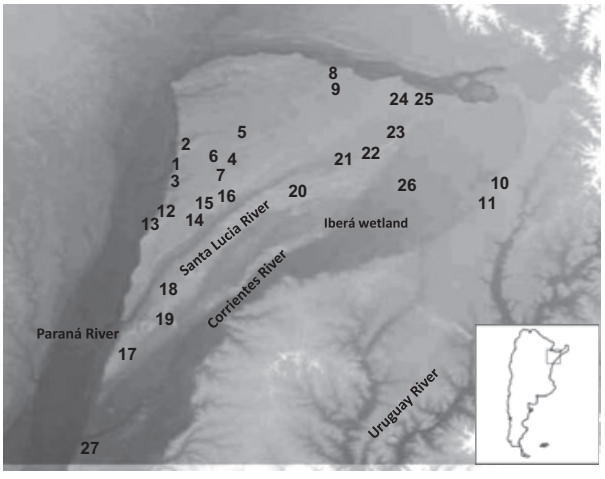

Fig. 1 Map of Corrientes Province showing the sampled localities of *Ctenomys*. The shading corresponds to a digital elevation model where it is possible to appreciate the extension of the lagoons in the area.

and dynamics of the evolutionary processes shaping the taxonomy of the group. Moreover, as no mitochondrial DNA was analysed, there were no insights into the historical events that shaped the actual distribution of genetic variation.

A speciose genus such as *Ctenomys*, which has the highest number of species among all subterranean rodents, and the fragmented and spatially and temporally changing nature of the wetlands habitat, provides a unique opportunity to assess the effects of heterogeneous habitats on the genetic structure of populations. Combining geographic patterns of genetic variation with the genealogical history and spatial distributions of gene lineages can greatly improve the understanding of historical and contemporary factors that influence the genetic compositions of populations and, ultimately, speciation (Hewitt 2001). Thus, we use this approach to study the genetic structure in this geographically subdivided group. More specifically, we have two objectives: (i) to explore genetic structure and past and present patterns of gene flow in the species complex and (ii) to investigate if differences in chromosomal and fundamental numbers could represent incipient barriers to gene flow. To this end, we combine analyses of mtDNA and microsatellite data to integrate different spatial and temporal perspectives. A phylogeographic approach allows inferences about tuco-tucos regional history and the processes that generated the observed patterns of molecular diversity. In addition, the analysis of spatially explicit genetic data allows us to infer species and population numbers and to describe the contemporary patterns of gene flow in relation to geography. Together, these analyses can provide the historical information, the evolutionary context and the present dynamics necessary to understand speciation in the group.

Materials and methods

Sampling strategy, DNA extraction and amplification

The study area is located in the Northeast of Argentina, in Corrientes Province (Fig. 1). It is a highly fragmented habitat because of the presence of the Iberá marsh and human disturbance (ranching, farming and forestry). The Iberá is one of the largest wetlands in South America, with most of its surface dominated by permanent or temporary water bodies delimited by dry areas with a predominance of sandy soils, sometimes forming elongated sandy hills. The gentle slope and the abundant vegetation in water bodies cause a slow drainage system, where the water flows gradually southwestward to drain towards the Paraná and the Uruguay rivers. Frequent rainfall, particularly during spring and autumn, recovers the water level of the marshes, which has remained stable in recent years, although with seasonal variations (Ferrati & Canziani 2005).

We obtained tissue samples (toe snips, preserved in 100% ethanol) from a total of 354 individuals from 27 localities, which were live trapped with Oneida Victor N° 0 snap traps, with a rubber cover to avoid injuring animals. This procedure does not affect survival or digging performance of the individuals (Mora *et al.* 2006). After collection of tissue samples for genetic analyses, animals were immediately released within the same burrow system where they had originally been captured. Tissue samples were collected from 1993 to 2003 and from 2007 to 2010. The first part of the data was formerly analysed by Mirol *et al.* (2010). The location of the sampling sites in both time periods was the same in all cases but one: Loreto A and Loreto B, separated by 1 km. Thus, we considered them as different sampling sites. Each individual location was recorded as Global Positioning System (GPS) coordinates. There was no data on individual location for the 1993–2003 samples; hence, random coordinates around the sampling site locations were generated for these individuals. Total DNA was extracted using a SDS-proteinase K-ClNa protocol (modified from Miller *et al.* 1988).

A fragment of 411 bp of the mitochondrial control region was amplified with primers TucoPro (Tomasco & Lessa 2007) and TDKD (Kocher *et al.* 1989). Polymerase chain reaction (PCR) amplifications were carried out in a 15-μl volume containing 30 ng DNA, 0.4 μM each primer, 0.2 mM dNTP, 1× Taq buffer (750 mm Tris–HCl, 200 mM $(NH_4)_2SO_4$, 0.1% (v/v) Tween 20), 2.25 mM $MgCl_2$ and 0.4 units of Taq polymerase (Fermentas). The cycling consisted of 94 °C for 5 min, followed by 30 cycles of 94 °C for 30 s, 52° for 30 s and 72 °C for 30 s, with a final extension of 72 °C for 5 min. Negative controls were included in all PCR runs to check for contamination. The products were visualized on 1.2% agarose gels and the size was verified with a 100-bp Ladder (Invitrogen). Amplicons were purified with shrimp alkaline phosphatase and exonuclease I (Amersham Biosciences). The purified PCR products were sequenced with a capillary sequencer ABI3100 (MACROGEN Inc., Korea), with the primers used for PCR.

Fifteen microsatellite loci described for *Ctenomys haigi* (Hai2, Hai4, Hai5, Hai6, Hai9, Hai10, Hai12, Lacey *et al.* 1999) and *Ctenomys sociabilis* (Soc1, Soc2, Soc3, Soc4, Soc5, Soc6, Soc7, Soc8, Lacey 2001) were amplified. Amplification was conducted on three multiplex panels of four microsatellites each and one of three microsatellites, which were intended to avoid marker overlap based on fluorescent labels and fragment sizes. Three loci failed to produce PCR products (Hai6, Hai7 and Soc4). PCR consisted in 12 μL volume containing 40 ng DNA, 0.2 mM dNTP, 1× Taq buffer (750 mm Tris–HCl,

200 mM $(NH_4)_2SO_4$, 0.1% (v/v) Tween 20), 2 mM $MgCl_2$ and 1.08 units of Taq polymerase (Invitrogen). Primer concentration varied from 0.35 to 0.66 μM. The cycling consisted of 94 °C for 5 min, followed by 35 cycles of 94 °C for 30 s, annealing temperature for 30 s and 72 °C for 30 s, with a final extension of 72 °C for 8 min. Annealing temperatures were 61 °C for Multiplex 1 (Soc5, Soc6, Hai12), 62 °C for Multiplex 2 (Soc3, Hai2, Hai5, Hai10), 60 °C for Multiplex 3 (Soc2, Soc8, Hai4) and 60.6 °C for Multiplex 4 (Soc1, Soc7). Genotypes were obtained with a MegaBace 1000 automated sequencer (GE Healthcare) and the patterns were scored and analysed using the MEGABACE FRAGMENT PROFILER 1.2 software (Amersham Biosciences).

Statistical analysis

Genetic variability. Mitochondrial DNA sequences were aligned with BIOEDIT version 7.0.9 (Hall 1999). ARLEQUIN 3.11 (Schneider *et al.* 1997) was used to identify haplotypes and determine the number of polymorphic sites, nucleotide diversity, pairwise distances between haplotypes and analysis of population genetic structure (AMOVA).

Relationships among haplotypes were studied with a network constructed with TCS 2.1 (Posada & Crandall 2001) using a 95% maximum parsimony connection criterion. The evolutionary history of the group was also studied using Tajima–Nei distances (Tajima & Nei 1984) among all sampled localities. We used MEGA 4 (Kumar *et al.* 2004) for haplotypes and DISPAN (Ota 1993) for microsatellites, in order to construct a phenogram of distances using NEIGHBOR in PHYLIP (Felsenstein 1991).

We used ARLEQUIN 3.1 to calculate standard genetic diversity indices for the microsatellite loci analysed: allele frequencies, observed and expected heterozygosities, linkage disequilibria for all pairs of loci and pairwise F_{ST} (Wright 1951; Weir & Cockerham 1984). We used the Markov Chain method of Guo & Thompson (1982) as implemented in GENEPOP 4.0 (Raymond & Rousset 1995; Rousset 2007) to evaluate deviations from Hardy–Weinberg equilibrium for all locus–population combinations and globally. We estimated allelic richness using FSTAT 2.9.3 (Goudet 1995) with 1000 resamplings of the data. Differentiation between populations and groups of populations was also assessed by analysis of molecular variance (AMOVA, Excoffier *et al.* 1992). Statistical significance of F_{ST} values was tested by 1000 permutations of genotypes among localities. This conservative procedure does not assume Hardy–Weinberg equilibrium and allows for linkage among loci.

Relationship between genetic and geographic patterns. First, we investigated the relationship between geographic and genetic distances using Mantel test and spatial autocorrelation analysis with the program AIS (Miller 2005). For microsatellites, the program calculates *D* based on Nei *et al.* (1983) distances for population frequency data, but applied to pairs of individuals rather than pairs of populations. For mitochondrial DNA, AIS calculates *D* as the percent/proportion of mismatched nucleotides between pairs of individuals. To do the autocorrelation analysis, we divided geographic distance into 15 distance classes of equal sample sizes. In this case, AIS calculates the statistic *Ay* as the average genetic distance between pairs of individuals that fall within each distance class, taking into account all loci. Thus, *Ay* takes on values of 1 when all individual are dissimilar and 0 when they are identical. We also evaluated the spatial pattern of covariance among individuals using Principal Component Analyses (PCA) of microsatellite genotypes (Patterson *et al.* 2006). The PCA and the spatial representation of its results were carried out using the R packages ADEGENET and FIELDS respectively. Finally, to detect sharp discontinuities in the distribution of allelic frequencies over the whole distribution of samples, we used the program BARRIER (Manni *et al.* 2004) to investigate the location of genetic barriers.

Population and species boundaries. We used STRUCTURE 2.3.3 (Pritchard *et al.* 2000) to estimate the number of populations and to assign individuals to them. This approach divides the sampled individuals into a number of genetic clusters (*K*) independently of locality information (i.e. based only on multilocus genotypic data), to minimize deviations from Hardy–Weinberg and linkage equilibrium within each cluster. The program also calculates the fractional membership of each individual in each cluster (*Q*). We let *K* vary between 1 and 30 and used the admixture model with correlated allele frequencies, as suggested by the software developers for closely related populations (Falush *et al.* 2003). We performed ten independent runs for each *K*, each one with 500 000 iterations as burn-in and 1×10^6 additional iterations. Selection of the most likely number of genetic clusters (*K*) was based on the methods of Pritchard *et al.* (2000) and Evanno *et al.* (2005). Chain convergence was checked by comparing values of parameters among independent runs with the same *K*.

To incorporate geographical coordinates into the assignment analyses, we employed the program TESS 2.3 (François *et al.* 2006). First, we tested the CAR and BYM admixture models, which differ in the prior used for admixture proportions (for a full description of the models, see François *et al.* 2006 and Durand *et al.* 2009), to choose the one that produced the lowest Deviance Information Criterion (DIC, a statistical measure of the model prediction capabilities). After choosing the best

model, the program was run 10 times for each K_{max}, with 500 000 sweeps, a burn-in of 20 000 sweeps and the interaction parameter $\psi = 0.6$. We varied K_{max} between 5 and 21 and plotted DIC vs. K_{max} to infer the number of clusters. More precisely, the number of clusters was inferred using the runs corresponding to the K_{max} value at which the DIC curve asymptoted.

Estimating contemporary and historical gene flow. LAMARC 2.1.3 (Kuhner 2006) was used to estimate historical migration, performing Bayesian analyses with five initial chains and two final chains with 100 000 recorded genealogies sampled every 20 steps, with a burn-in of 10 000. MODELTEST (Posada & Crandall 2001) was used to define the appropriate base-substitution model (HKY+I+Γ). We repeated the analyses five times, to verify the convergence of the Markov Chain Monte Carlo (MCMC) by visually inspecting and comparing the 95% percentiles for the most probable parameter estimates. The runs were combined using TRACER 1.5 (Rambaut & Drummond 2007), and only the estimates that gave effective sample sizes (EES) higher than 200 were accepted. We estimated the scaled population mutation rate (θ) and pairwise migration (*M*), and then calculated the per-generation average effective number of migrants dispersing into each population by multiplying the migration rate M by θ for each recipient population.

To make inferences about recent migration and compare them with the historical migration estimated with LAMARC, we analysed the microsatellite data using BAYESASS 1.3 (Wilson & Rannala 2003). We estimated recent migration rates (m) between clusters as identified by STRUCTURE and TESS. Burn-in and total number of MCMC iterations were 1 000 000 and 3 000 000 respectively, with a delta value of 0.15.

Results

Genetic variability

Mitochondrial DNA. We obtained mtDNA sequences from 192 of the 354 samples used for microsatellite analysis. There were 37 haplotypes, 25 were unique haplotypes and 12 were shared by two or three localities (Table 1).

Table 1 Sampling localities and mitochondrial variation of *Ctenomys* in Corrientes

Locality	Sp	*N*	Latitude	Longitude	H	P	π
1 Arroyo Pehuajó	Cr	8	−28.083	−58.833	1, 2	3	0.545
2 Costa Mansion	Cr	–	−28.016	−58.800	–	–	–
3 E. San Luis	Cr	9	−28.112	−58.863	1, 2	3	1.167
4 Mburucuyá	Cs	5	−28.097	−58.268	29, 30, 31	2	1.000
5 Manantiales	Cs	6	−27.956	−58.277	32	–	0
6 Loma Alta	Cs	3	−28.083	−58.324	29, 32	3	2
7 Pago Alegre	Cs	9	−28.144	−58.362	31	–	0
8 Mbarigüi	Cd	8	−27.574	−57.480	9	–	0
9 P. Angostura	Cd	10	−27.607	−57.469	9, 10	2	1.111
10 Tacuaritas	Cs	8	−27.966	−56.600	37	–	0
11 Contreras Cue	Cs	5	−28.019	−56.599	37	–	0
12 R. Ambrosio	Cp	9	−28.264	−58.885	7	–	0
13 C. 3 de Abril	Cp	10	−28.390	−58.893	6, 7, 8	2	0.733
14 Saladas S	Cs	6	−28.286	−58.684	35, 36	2	1.667
15 Saladas C	Cs	7	−28.239	−58.628	33, 34	6	3.429
16 Saladas N	Cs	–	−28.250	−58.518	–	–	–
17 Goya	Cp	8	−29.180	−59.200	3, 4, 5	4	1.179
18 San Roque	Cs	9	−28.654	−58.632	22, 23, 24, 25	6	3.278
19 Chavarría	Cs	9	−28.971	−58.583	24	–	0
20 Santa Rosa	Cs	8	−28.166	−58.016	29, 33	2	0.857
21 P. Caimán	Cs	9	−28.042	−57.679	16, 17, 18	6	1.333
22 San Miguel	Cs	9	−28.016	−58.684	16, 19, 20, 21	14	5.944
23 C. Laurel	Cs	9	−27.923	−57.489	15	–	0
24 Loreto B	Cs	6	−27.756	−57.272	14	–	0
25 Loreto A	Cs	10	−27.750	−57.255	14	–	0
26 San Alonso	Cs	6	−28.285	−57.412	26, 27, 28	4	1.333
27 Sarandicito	Cd	6	−30.166	−59.499	11, 12, 13	3	1.400

Sp, nominal species status; Cr, *Ctenomys roigi*; Cp, *Ctenomys perrensi*; Cd, *Ctenomys dorbignyi*; Cs, *Ctenomys* sp; *N*, number of individuals used for control region sequencing; H, code for the haplotypes found; P, number of polymorphic sites; π, nucleotide diversity.

The number of pairwise differences ranged from 1 to 14. Overall nucleotide diversity (π) was 1.08, and ranged from 0 to 5.95. Six localities (9, 15, 16, 18, 26 and 27) showed between one and two private substitution sites, whereas locality 22 showed four private substitutions. Pairwise F_{ST} values showed only 13 nonsignificant values (Table S1, Supporting information).

Microsatellites. Most of the localities were sampled two times, with a maximum time span of 8 years, which roughly corresponds to three generations (Busch *et al.* 1989). Although no significant genetic change is expected in such a short time, we performed an AMOVA analysis comparing allele frequencies to check if samples from the same locality but different years could be combined. There were no significant differences between the years of collection within localities ($P > 0.05$, Bonferroni corrected, in all cases), and therefore, the samples were combined. There were between 7 (Hai2 and Soc7) and 22 (Soc 6) alleles per locus. The mean number of alleles per locus was 13.50 ± 4.21, and the allelic richness within each locality varied between 1 and 3.65. Observed heterozygosity varied between 0.245 (for Hai2) and 0.617 (for Hai4) (Table S2, Supporting information) and its mean (±SD) was 0.448 ± 0.008. As previously reported (Mirol *et al.* 2010), there are two sampling sites, located east of the Iberá wetland, that show remarkable characteristics: all the individuals in locality 11 (Contreras Cue) and the majority of individuals of locality 10 (Tacuaritas) were homozygous for the 12 loci, with the exception of three heterozygotes for locus Soc1, Hai2 and Hai4 in Tacuaritas.

Locus Soc8 deviated from the Hardy–Weinberg proportions in 14 of the 27 sampling locations, probably because of null alleles, and was therefore excluded in all analyses in which Hardy-Weinberg (HWE) was assumed. Probability tests for genotypic linkage disequilibrium for each locus across all populations showed non-random associations in two comparisons Soc2-Hai2 ($P = 0.025231$) and Soc2-Hai10 ($P < 0.0001$).

The AMOVA results indicate that 34% of the variation was a result of among-locality differences ($F_{ST} = 0.34$, $P < 0.0001$). Most of the pairwise F_{ST} comparisons suggest significant differences between localities except for the case of Arroyo Pehuajó and Costa Mansion (localities 1 and 2).

Relationship between genetic and geographic patterns. Mantel tests were significant for mitochondrial haplotypes ($r = 0.3833$, $P < 0.001$) and microsatellite genotypes ($r = 0.3553$, $P < 0.001$) (Fig. S1, Supporting information). However, the correlation values were low, and seem to be mainly due to isolation by distance (IbD) at short distances, particularly in the case of microsatellite data. Figure S1 shows that for both data sets, the smallest geographic distance classes included almost the whole range of genetic distance values. The same pattern is obtained when running the autocorrelation analyses (Fig. 2a and b). In the case of microsatellites ($V = 0.081$, $P < 0.001$), there is a sharp increase in genetic correlation between the first and second distance classes and no further increases afterwards. For mtDNA ($V = 0.003$, $P < 0.001$), there is also a second increase between larger distance classes.

PC maps of the first four axes (Fig. S2) exhibit sinusoidal patterns similar to those described by Novembre & Stephens (2008) for samples simulated under several population structure scenarios. PC1 exhibits a clinal variation in the West–East direction, whereas the three other PCs exhibit a 'mound'-like shape. These patterns do not correspond to the theoretical expectations under isolation-by-distance scenarios (cf. Novembre & Stephens 2008; François *et al.* 2010), suggesting instead a discrete cluster-like population structure with asymmetric and nonhomogeneous migration patterns.

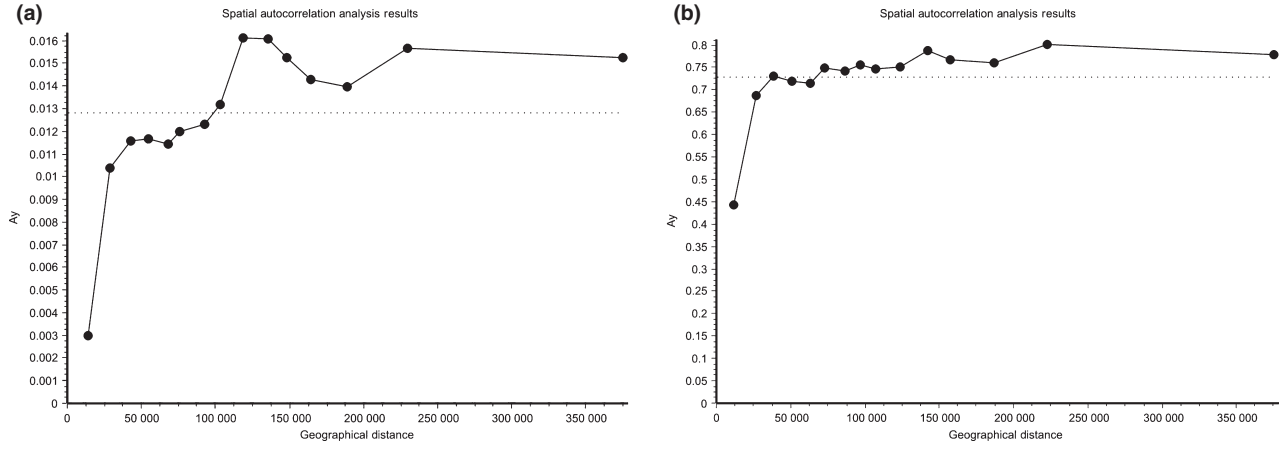

Fig. 2 Results of the spatial autocorrelation analysis (a) for mitochondrial DNA, (b) for microsatellites. Geographical distance in metres.

Population and species boundaries

Results of the STRUCTURE runs are shown in Fig. S3 (Supporting information). The log likelihood increases steadily without reaching a clear asymptote, but starts approaching a plateau at about $K = 17$. On the contrary, ΔK shows multiple peaks at $K = 2$, 5, 17, 20, 23, 26 and 29. Runs with $K = 2$ and $K = 5$ did not produce consistent results (cluster composition differed among independent runs). Runs with $K = 20$, 23 and 26 produced clusters where the ancestry coefficients of more than 50% of the individuals were smaller than 0.7. Only $K = 17$ produced consistent results (convergence of MCMC chains, with the same clustering of individuals and similar cluster membership coefficients) across the 10 independent runs (clusters 1–17, Fig. 3 and Table 2). An AMOVA analysis showed significant differences between the 17 clusters ($F_{ST} = 0.35$, $P < 0.0001$). Of these, nine inferred clusters (9–17) corresponded closely to predefined populations with membership coefficients ranging from 0.74 (population 21/cluster 11) to 0.92 (population 26/cluster 16). The eight remaining clusters comprised several populations but they all tended to group together nearby localities. Most populations associated with these clusters draw most of their ancestry from a single cluster. However, locality 18 had partial membership in multiple clusters, with the highest ancestry proportion (0.52) in cluster 8 and the second highest in cluster 9, which corresponded closely with a nearby population (19). This population may reflect admixture between neighbouring localities 17 and 19.

We can evaluate differences in the degree of individual admixture across localities by choosing an arbitrary threshold of $q \geq 0.9$ above which an individual is considered non-admixed. The percentage of individuals with mixed ancestry varied among localities between 0 (localities 10, 11 and 12) and 77% (locality 18, cluster 8). It is remarkable that only three localities had only non-admixed individuals, whereas the rest had individuals with varying degrees of admixture. Again, locality 18 (San Roque) had the highest proportion of admixed individuals, and the STRUCTURE result shows three sam-

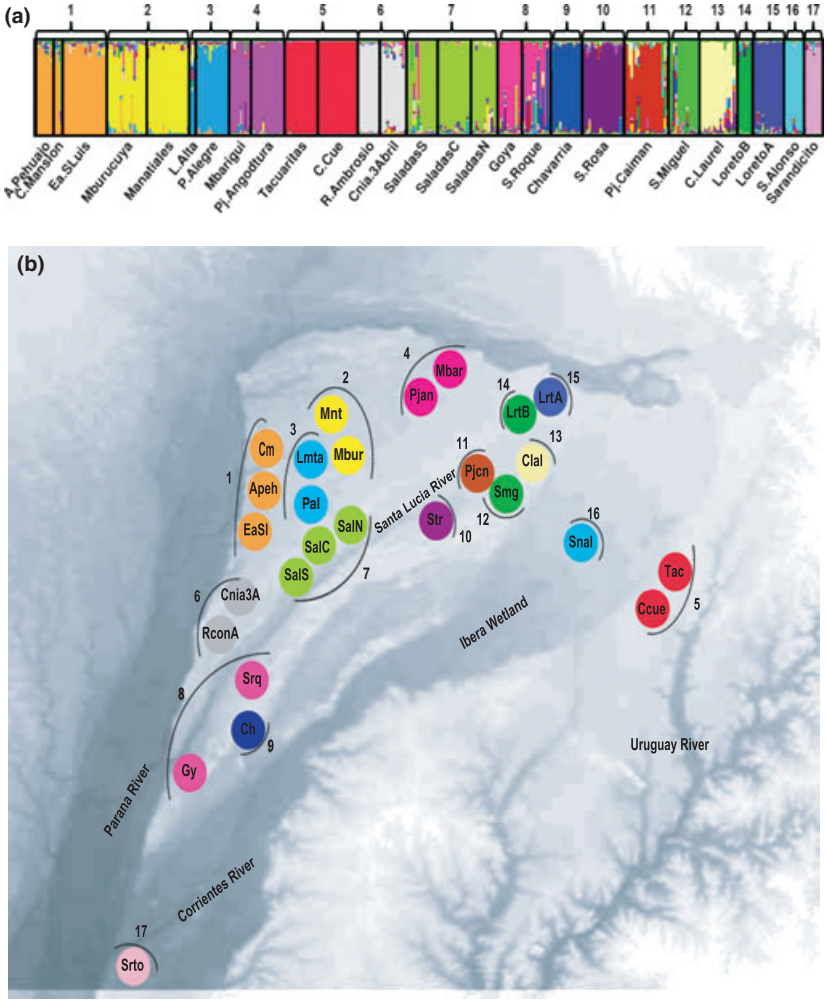

Fig. 3 (a) Proportional membership (q) of each individual in the $K = 17$ population clusters identified by STRUCTURE. Each animal is identified by a single vertical bar, and each cluster by a different colour. At the bottom are the locality names and at the top the cluster number. (b) Map showing the geographic location of each cluster.

Table 2 Clusters as determined by STRUCTURE (S) and TESS (T). Also shown, the clusters inferred by BAYESASS (B)

Locality	*N*	S	Q	F_{STL}	T	AR	B	2n/FN
1 Arroyo Pehuajó	8	1	0.887	0.408	A	3.478	A	48/76*
2 Costa Mansion	4	1	0.657		A		A	48/76*
3 Estancia San Luis	20	1	0.920		A		A	48/76†
4 Mburucuyá	18	2	0.774	0.418	B	3.818	B	51–53/76
5 Manantiales	19	2	0.908		B		B	?
6 Loma Alta	3	3	0.675	0.398	B		B	?
7 Pago Alegre	15	3	0.897		B		B	?
8 Mbarigüi	10	4	0.816	0.461	C	3.477	C	70/80*
9 Paraje Angostura	15	4	0.921		C		C	70/80*
10 Tacuaritas	15	5	0.967	0.923	E	1.194	E	42/72*
11 Contreras Cué	18	5	0.975		E		E	41/72*
12 Rincón de Ambrosio	10	6	0.941	0.288	I	4.493	I1	50/80*
13 Colonia 3 de Abril	12	6	0.832		I		I1	50/80*
14 Saladas S	14	7	0.793	0.332	I		I2	54–58/80*
15 Saladas C	15	7	0.906		I		I2	?
16 Saladas N	12	7	0.821		I		I2	54–56/80*
17 Goya	12	8	0.803	0.144	J	5.700	J	50/80*
18 San Roque	13	8	0.519		J		J	62/80*
19 Chavarría	14	9	0.870	0.215	K	4.786	J	56/80*
20 Santa Rosa	19	10	0.874	0.245	G	4.959	G1	66/80*
22 San Miguel	14	12	0.767	0.299	G		G2	44/72*
21 Paraje Caimán	20	11	0.735	0.248	H	4.975	H	46/74*
23 Curuzú Laurel	17	13	0.805	0.332	H		H	42/72*
24 Loreto B	8	14	0.871	0.317	H		D	42/72*
25 Loreto A	13	15	0.885	0.400	D	3.483	D	42/72†
26 San Alonso	9	16	0.923	0.498	F	2.935	H	42/74†
27 Sarandicito	7	17	0.791	0.301	L	4.083	L	70/80*

N, number of individuals used for microsatellite genotyping; Q, membership proportion according to STRUCTURE, F_{STL}, local F_{STs}; AR, allelic richness; 2n/FN, diploid and fundamental numbers

*As in Giménez *et al.* (2002).

†D. Caraballo, personal communication.

ples in this locality that could be considered as migrants, as their highest membership coefficient (above 0.7) corresponded to the geographically closer locality 19 (Chavarría, cluster 9). There was only one more migrant identified by STRUCTURE, in locality 2 that originated in locality 22 (cluster 12). Both localities are separated by 116 km, which makes the result highly suspicious. The lowest local F_{ST} (0.144, Table 2) corresponded to cluster 8 (localities 17 and 18, with 42% and 77% of admixed individuals respectively). The highest local F_{ST} (Table 2) corresponded to cluster 5 (0.923), composed of two almost monomorphic localities (10 and 11) and with a mean observed heterozygosity of 0.01.

The results of the TESS analyses showed that the best DIC for the BYM model (22020.5) was much larger than that for the CAR model (21529); therefore, we only present the results for the latter. The plot of DIC vs. K shows that the curve asymptotes at about $K = 17$ (Fig. 4a). The best three runs corresponded to $K_{max} = 19$, which inferred 12 distinct clusters. Increasing the K_{max} did not change this solution. The clusters, labelled A to L, are shown in Fig. 4b and c, and Table 2. An AMOVA analysis showed significant differences among the 12 clusters ($F_{ST} = 0.35$, $P < 0.0001$). TESS clusters A, C, D, E, F, J, K and L correspond to individual clusters identified by STRUCTURE (1, 4, 15, 5, 16, 8, 9 and 17 respectively). Differences between both analyses consisted of TESS cluster B, formed by clusters 2 and 3, TESS cluster G by 10 and 12, TESS cluster H by 11, 13 and 14 and TESS cluster I by 6 and 7. Four inferred clusters (D, F, K and L) correspond closely to predefined populations. Most of the remaining ones group together neighbouring localities but one, cluster G, groups together non-neighbouring populations (localities 20 and 22). Locality 20 had 12 of 19 individuals with $q < 0.9$. Those admixed individuals draw their ancestry not only from neighbouring clusters I, B and H, but also from the more distant cluster C. All individuals of locality 22 had $q < 0.9$ and had ancestry mainly in nearby cluster H, and also a substantial contribution from non-neighbouring clusters B, C, D and E.

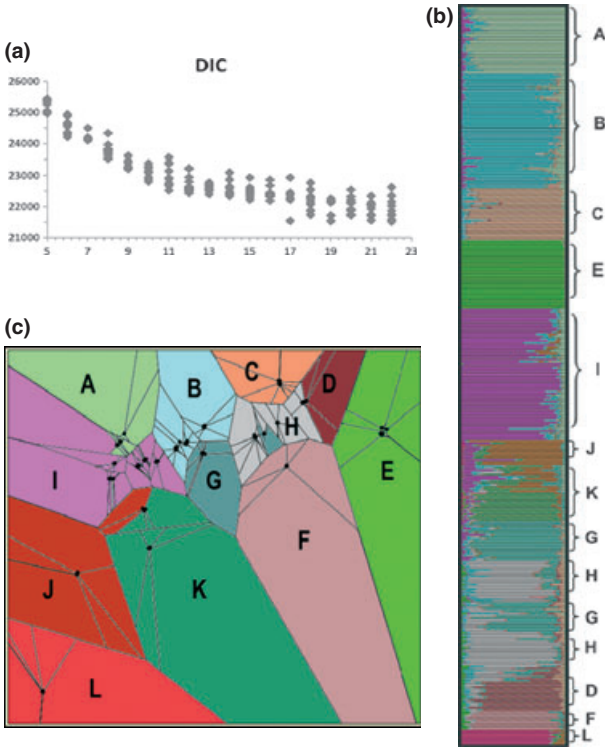

Fig. 4 (a) Plot of DIC values against *K*, the putative population clusters defined by TESS; (b) proportional membership of each individual in the 12 clusters; (c) map obtained with TESS of the 12 distinct clusters identified.

The TESS analysis also showed, as STRUCTURE, that the individuals from population 18 had partial membership in several clusters indicating that it may reflect admixture between neighbouring localities. The same is true of population 24. Figure 5 shows the pattern of ancestry detected for each locality. When considering $q > 0.9$ as the individual threshold for acceptance of its membership to a genetic cluster, only 5 localities (3, 5, 9, 10 and 11) had all its individuals belonging to a single genetic cluster. The remaining 22 localities had individuals with varying degrees of admixture. Clusters A, B, C and I, corresponding respectively to the three named species *Ctenomys roigi*, *Ctenomys perrensi* and *Ctenomys dorbignyi*, are located between the Paraná and the Santa Lucía Rivers, and group the localities with the lowest percentage of admixed individuals. On the contrary, clusters D, G, H, J and K correspond to localities found between the Santa Lucía River and the Corrientes River which show the highest percentage of individuals with mixed ancestry. These include locality 18 and locality 24, which, as already mentioned, showed partial membership in many different clusters without a clear pertinence to any of them. Locality 26 (cluster F), at present very isolated and located in an island in the middle of

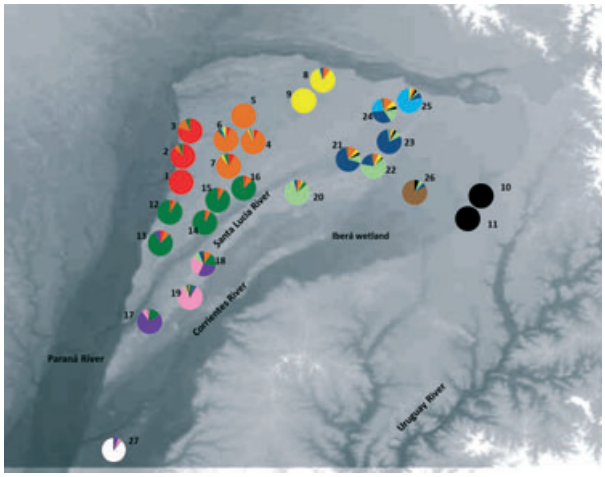

Fig. 5 Pattern of ancestry in the 27 sampled localities according to TESS. Colours represent the different clusters, and numbers the localities as in Table 1.

the Iberá wetland, shows partial membership to clusters located to the west (E) and to the east (G and H). These localities and their corresponding clusters showed the highest allelic richness (Table 2), which can be explained by the high proportion of mixed ancestry.

It is also interesting to explore what happens when the maximum number of clusters used as input is smaller than the best K obtained with the best runs. When K_{max} was set to 11 and 10, the solution was 10 clusters. From that on, to $K_{max} = 5$, the number of clusters always matched up K_{max}. The successive grouping of clusters was as follows: J+K+L; H+F; D+H+F and D+H+F+G. This result resembles very closely the eight groups within the 'perrensi' complex proposed in previous studies based on chromosomal constitution and *cytochrome b* sequences (Giménez *et al.* 2002; Mirol *et al.* 2010). This suggests a hierarchical population structure with individual populations grouped according to their chromosomal constitution. To further explore this possibility, we ran TESS 10 times with $K_{max} = 8$. The results of the different runs were inconsistent, although this inconsistency was always related to the grouping of localities of clusters D, G, H and F. When we set the K_{max} to 6, the ten runs produced always the same clustering of localities: A+B, I, C, J+K+L, D+H+F+G and E (Fig. 6). This structure is very congruent with the variation in chromosomal numbers: FN = 76, 2n = 48–56 for localities in the cluster A+B; FN = 80, 2n = 50–58 for localities in I; FN = 80, 2n = 70 for cluster C; FN = 80, 2n = 50–70 for J+K+L; FN = 72–74 2n = 42–46 for D+H+F+G as well as in cluster E. We interpret the separation of these last two clusters, which share fundamental and chromosomal numbers, as the consequence of the almost homozygous nature of localities 10 and 11 in cluster E. Running STRUCTURE with $K = 6$ produces the same result.

An AMOVA among these six groups showed that 63.82% of the variation was found within populations, 19.65% among groups and 16.53% among populations within groups. Although the three fixation indexes were significant (F_{ST} = 0.36, F_{SC} = 0.21, F_{CT} = 0.19, $P < 0.001$ in all cases), it is interesting that more variation is explained by differences among groups than among populations within groups.

Figure 7a illustrates a population neighbour-joining tree based on Nei's *D* (Nei *et al.* 1983). Three main clades, with bootstrap support between 56% and 72%, are observed: the first one includes *C. roigi*, the 'Manantiales' group and *C. dorbignyi* North (TESS clusters A, B and C), the second consists of *C. perrensi* North and South and geographically close localities (TESS clusters I, J and K) and the third one incorporates TESS clusters D, H, F and G. The two almost monomorphic localities 10 and 11 (TESS cluster E) and *C. dorbignyi* South (TESS cluster L) are part of the basal polytomy. Figure S4 (Supporting information) shows the location of the main genetic barriers as depicted by the BARRIER analysis. The barriers found determine 10 clusters, almost identical to

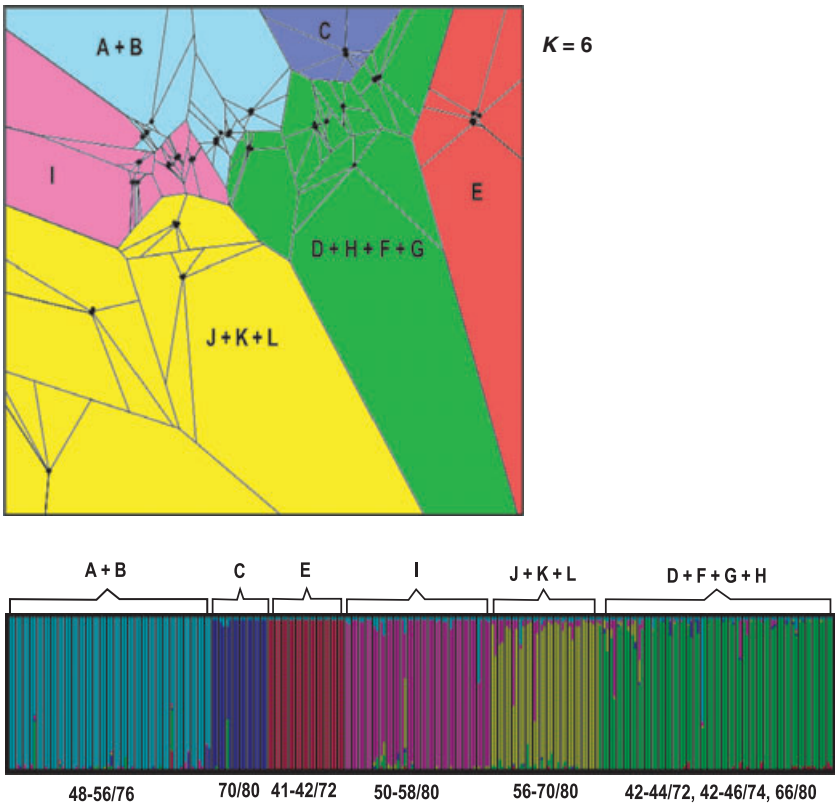

Fig. 6 Clusters obtained with TESS for $K = 6$. STRUCTURE rendered the same clustering solution. Diploid and fundamental numbers are indicated under each cluster.

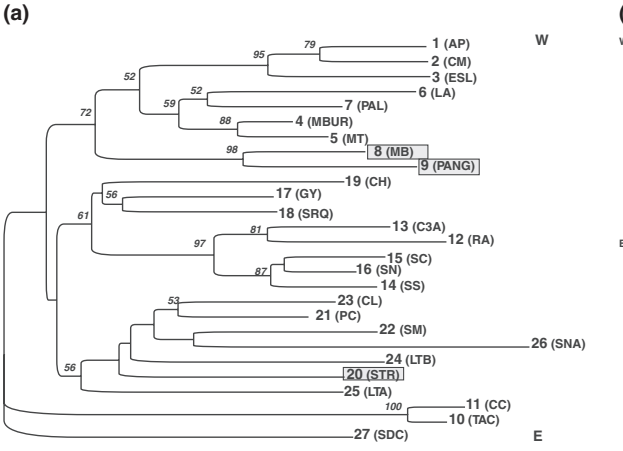

Fig. 7 (a) Neighbour-joining trees based on Tajima–Nei distances for microsatellites. Values on the nodes are bootstrap values over 1000 replicates. (b) Network of haplotypes obtained with TCS. The circles represent different haplotypes. E and W represent the East and West clades respectively, whereas the grey shadow shows the contrasting clade membership, according to both molecular markers, for localities 8, 9 and 20.

the ones obtained with TESS, except for the joining of J and K as well as G and H.

Finally, we used BAYESASS to determine recent migration based on allelic frequencies at the microsatellite loci. To do this, we first ran the program with 17 clusters according to STRUCTURE and then with 12 clusters according to TESS.

When we considered 17 populations, we found the following migration pattern: from cluster 2 to 3 (m = 0.245, CI: [0.138, 0.317]), from cluster 16 to 13 and 11 (m = 0.237 CI: [0.127, 0.313] and m = 0.252 CI: [0.156, 0.317] respectively), from cluster 14 to 15 (m = 0.216 CI: [0.100, 0.311]) and from cluster 8 to 9 (m = 0.213 CI: [0.099, 0.311]). However, all individuals identified as migrants were first-generation migrants, which suggests that STRUCTURE overestimates the number of clusters.

When we set 12 populations representing the TESS clusters, there were no first-generation migrants, but two individuals belonging to cluster J were assigned as second-generation migrants from cluster I and cluster B respectively.

Evolutionary history

We analysed the mtDNA of 192 individuals corresponding to the 27 sampled localities. In total, we obtained 37 haplotypes (Table 1). Most of the haplotypes were confined to one locality, with few exceptions. Furthermore, among these exceptions, only two haplotypes were found in different clusters as defined by the TESS analysis, both cases involving one locality: haplotype 29, found in localities 4 and 6 in cluster B and locality 20 in cluster G; and haplotype 33 found also in locality 20 from cluster G and in locality 15 from cluster I. Remarkably, those two were the only haplotypes found in the eight individuals sequenced from locality 20. The maximum number of pairwise differences between haplotypes was 14, with most haplotypes having 4–8 differences between them.

Haplotypes 1 and 2 were shared between the three localities corresponding to *C. roigi*, which were identified as a unique gene pool according to the TESS and STRUCTURE analyses. Haplotype 7 was shared between the two localities of *C. perrensis* North, haplotype 9 between the two localities of *C. dorbignyi* North, haplotype 14 between localities 24 and 25, haplotypes 31 and 32 between localities of cluster J, haplotype 24 between localities 18 and 19, haplotype 16 between localities 21 and 22 and finally haplotype 37 was the only one found in localities 10 and 11, recognized as a unique gene pool in all analyses.

Figure 7b shows the network of haplotypes, where all haplotypes appear closely related with no clear structure. Most connections among haplotypes included only one substitution. However, it is possible to differentiate two groups of haplotypes that are closer among themselves than with the others, and this could be related to geography. The first group is constituted by haplotypes found in populations at the north-west of the sampled region, between the Paraná and the Santa Lucía Rivers ('West' group). The second ('East' group) groups together the localities further east and north of the Santa Lucía River, including the two disjunct nucleus of *C. dorbignyi* (locality 27 in the south and localities 8 and 9 in the north), locality 26 is an island in the middle of the wetland and localities 10 and 11, the only areas with presence of the genus between the wetland and the Uruguay River. However, there is an exception: locality 20 (Santa Rosa), located at the east of the Santa Lucía River, is included in the first group. Note that the only two haplotypes found in this sampling site were shared with localities 4, 7 and 15.

Overall nucleotide diversity was 0.0113 in the East group, and per sampled site ranged between 0.0013 and 0.0083. In the West group, the overall nucleotide diversity was 0.0145, with a per locality diversity between 0.0027 and 0.0144. Differentiation between both groups was also tested by AMOVA. The largest proportion of the variance is explained by the variation among populations within the groups (53.09%), whereas the variation between groups was 30.25%, with an F_{ST} of 0.82060 ($P < 0.0001$).

We used the program LAMARC to infer ancestral migration and spreading of mitochondrial forms. We ran analyses with $K = 12$ according to the TESS results, and with $K = 17$ according to STRUCTURE results. Tables 3 and S3 (Supporting information) show the per generation number of immigrants between each pair of populations. Waples & Gaggiotti (2006) suggested that evolutionary units can be identified using $Nm = 1$ as the threshold for the transition between drift and gene flow regimes. Thus, we describe historical migration patterns by focusing on migration into populations for which $Nm > 1$. For $K = 12$ populations, Nm was above the threshold in the following cases: into cluster H from G and F; into cluster G from H and I; into cluster F from H: into cluster I from B and G and finally into cluster J from A and K. It is remarkable that migration seems to occur only between geographically contiguous clusters at the center of the studied distribution, while most of the marginal clusters (C, D, E and L) did not show any detectable migration.

When we ran LAMARC with $K = 17$, according to the STRUCTURE results, we found more substantial migration but the overall pattern did not change. This is because the extra migration occurred among localities that constituted different clusters in STRUCTURE but were included in the same cluster in TESS. For example clusters 2 and 3 in STRUCTURE showed high migration rate

Table 3 Bidirectional estimates of migration using $K = 12$ as obtained in TESS. Cluster names as in Table 2 and Figure 5. Values reported are estimated using LAMARC from the cluster along the column into the cluster along the line. Reported values are absolute number of migrants. In bold, $Nm > 1$

	A	B	C	D	E	F	G	H	I	J	K	L
A	.	0.62	0.10	0.12	0.12	0.14	0.14	0.22	0.18	0.36	0.14	0.12
B	0.74	.	0.14	0.12	0.16	0.30	0.82	0.03	0.62	0.32	0.34	0.36
C	0.16	0.22	.	0.24	0.18	0.26	0.30	0.40	0.32	0.24	0.24	0.40
D	0.01	0.01	0.01	.	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01
E	0.03	0.03	0.03	0.03	.	0.05	0.04	0.06	0.03	0.03	0.04	0.06
F	0.34	0.54	0.36	0.32	0.50	.	0.96	**1.90**	0.70	0.62	0.56	0.62
G	0.16	0.84	0.26	0.22	0.30	0.40	.	**2.66**	**1.26**	0.14	0.30	0.44
H	0.22	0.28	0.34	0.52	0.30	**1.04**	**3.06**	.	0.16	0.30	0.16	0.30
I	0.38	**1.12**	0.26	0.18	0.26	0.42	**1.92**	0.18	.	0.36	0.48	0.68
J	**1.06**	0.86	0.90	0.60	0.46	0.86	0.40	0.54	0.60	.	**2.44**	0.78
K	0.02	0.04	0.02	0.02	0.02	0.04	0.03	0.02	0.03	0.08	.	0.02
L	0.16	0.20	0.24	0.16	0.42	0.32	0.42	0.48	0.70	0.38	0.26	.

per generation between them ($4Nm = 2.38$), and both belonged to cluster B according to TESS. Tables 3 and S3 show that clusters G/12, F/16 and J/8—TESS and STRUCTURE names respectively—are the ones that received the largest number and variety of migrants. They are also the ones including the localities with the highest mitochondrial variability (Table 1).

Discussion

Our study investigates the genetic structure of a species complex of the subterranean rodent *Ctenomys* inhabiting the Iberá wetland, a highly dynamic landscape in northern Argentina. To understand the relationships among species and populations, and the evolutionary processes operating in the group, we examined mitochondrial DNA sequences and microsatellite genotypes of representatives of the different forms in the area. In what follows, we first describe the historical and current population structure and then draw conclusions about the potential for ongoing speciation in this system.

Historical and current population structure

The mitochondrial analysis performed here showed a strong geographical structuring of the localities, where most of them presented restricted and unique haplotypes, but with some evidence of IbD. Although the network of haplotypes (Fig. 7b) indicates close relationships, with most haplotypes differing in only one or two substitutions, it is possible to distinguish two groups in relation to geography, the West and East groups. Pairwise comparisons of mtDNA F_{ST} estimates and moderate nucleotide diversity indicated low levels of differentiation within these groups. Ancestral migration, as detected by LAMARC, seems to have had considerable influence in this structuring, including migration between East and West groups (Table 3). Remarkably, migration occurred only among localities at the centre of the distribution.

In terms of nominal species, the West group contains *Ctenomys roigi* and *Ctenomys perrensi*, whereas the East group includes *Ctenomys dorbignyi*. The two groups corresponded well with the barrier constituted by the Santa Lucía River, with one exception: locality 20 (Santa Rosa) is located to the east of the river, but groups with the localities to the west. The inclusion of Santa Rosa in the West clade is clear: there are two haplotypes in the locality, both of them shared with localities of the West clade. However, microsatellite data contradict the position of this locality within the system of sampling sites (see below).

The current genetic population structuring revealed by microsatellites differs to a large extent from the historical structuring inferred from mtDNA. Spatially explicit multivariate methods provide no clear evidence of IbD, suggesting instead a discrete cluster-like population structure with asymmetric and nonhomogeneous migration patterns. This result agrees with the sharp increase in genetic correlation between the first and second classes of the autocorrelation and no further increases afterwards, which fits a model of isolation with migration in a metapopulation.

The results obtained using TESS and STRUCTURE differed: 12 genetic clusters in the first case and 17 in the second. The extra five clusters defined by STRUCTURE were subdivisions of clusters in TESS. Moreover, according to BAYESASS, there was no detectable migration among the 12 TESS clusters, whereas all detectable migration for $K = 17$ occurred among STRUCTURE clusters that were grouped together in a single TESS cluster. Given these results, we concluded that STRUCTURE is overestimating the true number of popula-

tions, and adopted 12 genetic clusters as the most appropriate description of the genetic structure of this system. These clusters have a close relationship with geography in all cases but one: cluster G is formed by two noncontiguous localities (20 and 22). In many other cases, the distance among localities within the same cluster is far higher than what it can be accounted for based on dispersal. However, we cannot exclude the existence of unsampled localities that could 'fill in' the gaps and produce a more homogeneous picture.

The 12 inferred populations partially coincide with those identified by a previous study based on a smaller data set (Mirol *et al.* 2010). It is worth noting that the former study inferred eight clusters; six of them coincide with the results of this study and the other two are split each one into two here. The two additional clusters we found (clusters D and F) corresponded to newly sampled localities.

Our analyses uncovered remarkable differences between mtDNA and microsatellite genetic structuring and dynamics. As mentioned before, mitochondrial DNA results suggest substantial historical migration among localities in the centre of the *Ctenomys* distribution. However, microsatellite results indicate an absence of recent migration among the 12 populations. The results of the BARRIER analysis agreed with the different clustering studies performed (Fig. S3, Supporting information). Additionally, they do not indicate the West–East differentiation that we found based on mitochondrial DNA. Overall, these results suggest a metapopulation structure in which the connectivity of the system was recently reduced by ongoing fragmentation probably as a result of various anthropic disturbances, such as farming and forestry, and the rise in water level of the entire wetland as a result of the construction almost 30 years ago of the Yacyreta-Apipé hydroelectric dam.

Some results deserve a special mention. First, among the 27 localities that constitute the 12 genetic clusters, San Roque and Loreto B had all of their individuals with partial membership in several clusters, and could represent 'sink' localities, located in regions with higher connection within the system and where individuals originated in different genetic clusters interbreed.

Second, there seems to be evidence of the existence of founder events within the metapopulation system. The two localities that form cluster E, Tacuaritas and Contreras Cue, showed the majority of the microsatellites monomorphic for alleles commonly found in the other sampled sites. This result points out to a founder event of recent occurrence from a single source. As these two localities are the only ones at the northeast of the Iberá, our hypothesis is that a few individuals could have crossed the wetland from the west using 'embalsados'—floating isles formed by vegetation, sediments and soil, which move from side to side on the wetland carrying animals. The same hypothesis could be applied to cluster F. Locality San Alonso is isolated in an island in the middle of the Iberá basin, and, as in the case of the 'Tacuaritas' group, it could have been originated by the movement of 'embalsados' from the west. However, genetic variability in this locality is within the range found in the whole system, with three different haplotypes and an allelic richness of 2.37, thus it seems more probable in this case that the founding groups comprised individuals from more than one source.

Third, genetic cluster G is formed by localities that differ sharply in their karyotypes; locality 20 (Santa Rosa, 2n/FN = 66/80) and locality 22 (San Miguel, 2n/NF = 44/72). Santa Rosa is phylogenetically related to the Manantiales group and the Saladas group (West clade, 2n/NF = 50–58/76–80) with which it shares haplotypes. The connections with them were probably cut off because of an expansion of the marsh in between, leading to new contacts with eastern groups. It is remarkable that nowadays these two localities with very different chromosomal constitution represent a unique genetic cluster, providing the most striking evidence that gene flow between different karyotypic forms is not impeded.

Fourth, according to their mitochondrial DNA, localities 8 and 9 group with *C. dorbignyi* South and other localities to the east of the Santa Lucia river, but their microsatellites suggested a closer relationship with the geographically proximate localities of the West group (2n/FN = 50–58/76–80). This is very surprising, as individuals inhabiting those two localities belong to the nominal species *C. dorbignyi*, based on both morphology and karyotype.

Taken together, the above results indicate that the *Ctenomys* species complex inhabiting the Iberá wetlands behaves as a highly dynamic metapopulation with an asymmetric and nonhomogeneous migration pattern. Such migration pattern leads to a hierarchical population structure. In this regard, we obtained a very consistent result running TESS and STRUCTURE with $K = 6$ (Fig. 7). The six genetic clusters obtained fit very well with previous studies based on *cytochrome b* sequences and especially with chromosomal numbers (Giménez *et al.* 2002). Furthermore, the AMOVA analysis indicated that the proportion of variation explained by differences among groups was larger than that explained by differences among populations within groups. In view of the particular characteristics of monomorphic cluster E, which is probably the result of a recent founder event, and shows the same chromosomal number than populations in cluster D+H+F+G, we are inclined to join it to this last group. Thus, we conclude that the current genetic structuring of

this system is hierarchical, with populations grouping into five groups that differ in chromosomal constitution.

Ongoing speciation

The species concept has generated longstanding disagreement among biologists. However, as Hey *et al.* (2003) argued, species-related research can proceed without suffering from the ambiguity of the species concept; instead of drawing lines of demarcation, we can show the complexity of patterns and processes that evolution has created. Nevertheless, delimiting species is important for the understanding of those evolutionary patterns and processes: the species boundary will define the limits within or across which they will operate (Sites & Marshall 2004). For this reason, in the discussion that follows, we adopt the metapopulation lineage concept (de Queiroz 1998, 2005): species are separately evolving metapopulation lineages, where a metapopulation is a set of connected subpopulations and a lineage is a population extended through time; that is, an identifiable biological unit on an independent evolutionary trajectory (Reeves & Richards 2011). A species is not necessarily an entire metapopulation lineage, but one of many segments that make up such a species-level lineage.

This concept eliminates the problem of which property—reproductive isolation, monophyly, occupancy of a distinct niche or adaptive zone, etc.—is necessary for the definition of species, while considering them as contingent properties that the metapopulation may acquire during the course of its existence.

However, even this less restrictive concept of species presents problems in the case of very young taxa that are in the process of diversification. The 'perrensi' group of *Ctenomys* is a very challenging taxon in this respect. It belongs to a genus that underwent an explosive speciation during the Pleistocene, originating over 62 living species (Reig *et al.* 1990). Three of those species were described in the Iberá wetland: *C. dorbignyi*, *C. perrensi* and *C. roigi*, together with many forms of uncertain taxonomic status (Ortells 1995). Altogether, chromosomal numbers vary from 42 to 70, and every particular location seems to show a particular karyotype. Their existence is associated with a highly fragmented and fluctuating habitat: the Iberá wetland changes in time and space, altering the availability of suitable habitat for the rodents, making populations to become connected or isolated at different times and in different areas. In this context, not only defining species, but also delimiting populations becomes difficult.

There is an extensive chromosomal variability described in the genus *Ctenomys*. Within the 'perrensi' group, it has been proposed that by independent events, principally of Robertsonian type, the diploid number was reduced from an originally fixed and widespread 2n = 70 karyotype, now only present in relict populations such as the two *C. dorbingyi* nucleus (Giménez *et al.* 2002). This species is the only one whose distributional range continues outside the Iberá region, to the south. The reduction in chromosomal number would have occurred in small semi-isolated demes, where the stochastic fixation of chromosomal rearrangements may be the mechanism that generates this variability (Ortells *et al.* 1990; Massarini *et al.* 1991; Ortells 1995; García *et al.* 2000; Giménez *et al.* 2002). Our mtDNA analyses identified two groups: the West group was formed by localities with diploid numbers 48, 50, 54, 56, 58, 62 and 66, and fundamental numbers 76 and 80, whereas the East group included *C. dorbignyi* with 2n = 70, FN = 80 and a pool of localities with the lowest diploid (42–46) and fundamental (72–74) numbers. The two disjoint cores of *C. dorbignyi* at the north and south of the distribution suggest a more widespread distribution of the species in the past. It is then possible to imagine a dual pattern of spreading and fixation of forms, where gradually lower chromosomal numbers were fixed at local populations to the west, originating the clade with chromosomal numbers over 50, whereas a drastic reduction occurred to the East, originating individuals with only 42 chromosomes.

Given these results, the question arises as to whether or not variability in chromosomal numbers is acting as a reproductive barrier. One potential explanation for the high rate of speciation in the genus is that chromosomal variation could have triggered speciation through fixation in small and inbred demes. However, heterozygosity for Robertsonian (fusion and fission) rearrangements does not have strong negative meiotic effects (except for the case of monobrachial fusions; Baker & Bickham 1986), and will have essentially no effect on flow of neutral genes, except for those closely linked to the rearrangement (Rieseberg 2001 and references therein). Whether or not Robertsonian changes are the main source of variation in chromosomal numbers in the 'perrensi' group, one should not expect differences in karyotype to act as strong genetic barriers.

The analyses of microsatellite loci provide an alternative explanation. They uncovered a hierarchical genetic structure with 12 genetic clusters grouped into five higher level clusters that are congruent regarding molecular markers, chromosomal constitution and geography. Although populations within groups tend to be similar in terms of chromosomal diploid and fundamental numbers, there is clear evidence of substantial gene flow among localities that differ sharply in this respect. In particular, the Santa Rosa and San Miguel populations differ in their karyotype, but both are included in cluster G (see previous sections). This sug-

gests that chromosomal rearrangements do not represent a barrier to gene flow in this group. Instead, differences in chromosomal constitution among groups seem to be the result of an ongoing isolation with migration process. Additional evidence for this type of nonequilibrium dynamics is provided by the results of spatially explicit multivariate methods that uncovered short-range IbD, more evident with mtDNA than with microsatellite data. Indeed, for species with restricted dispersal, such as subterranean rodents, short-range IbD is indicative of a rapid range expansion (Slatkin 1993; Kittlein & Gaggiotti 2008).

Despite a history of common ancestry, the existence of unresolved mitochondrial ancestral polymorphisms and the finding of what seems to be a very low level of gene flow, the five groups seem to be behaving as separately evolving metapopulation lineages that are undergoing a speciation process. We believe that these incipient species are at the early stages of differentiation, just when lineage discovery and delimitation are neither straightforward nor precise (Shaffer & Thomson 2007). The probability of these lineages to acquire the contingent properties characteristic of species depends to a great extent on the stability of their habitat. The habitat theory of speciation (Vrba 1992), which has already been cited to explain the high rate of speciation in the genus (Lessa & Cook 1998), could be easily applied to the 'perrensi' group: speciation does not occur unless forced by changes in the physical environment, and clades of species whose resources tend to disappear—as in this case—during their histories have high rates of speciation and extinction (Vrba 1992). However, if the environmental barriers that help separate populations and prevent gene flow are ephemeral, even a high rate of evolution cannot lead to speciation (Magurran 1998). Ultimately, the evolutionary future of these five lineages is tightly controlled by the dynamics of their habitat: if this is stable, they will become distinct species; otherwise, they will collapse into a hybrid swarm, forming a single evolving metapopulation.

Acknowledgements

This study was supported by Agencia Nacional de Promoción Científica y Técnica (PICT 38361 and PICT 1551) and Consejo Nacional de Investigación, Ciencia y Técnica, Argentina (PIP 5776 and International Cooperation CNRS-CONICET). We thank Vanina Raimondi, Diego Caraballo, Fernando Mapelli and Matías Mora for their incalculable help during fieldwork and also all the members of the Instituto de Genética Veterinaria Ing. Fernando Noel Dulout (Facultad de Ciencias Veterinarias, UNLP) for their excellent work and assistance.

Author contributions

MJGF performed research, collected samples, analysed the data and wrote the paper; OEG analysed the data and wrote the paper and PMM designed research, collected samples, analysed the data and wrote the paper.

References

Baker RJ, Bickham JW (1986) Speciation by monobrachial centric fusions. *Proceedings of the National Academy of Sciences*, **83**, 8245–8248.

Busch C, Malizia AI, Scaglia OA, Reig OA (1989) Spatial distribution and attributes of a population of *Ctenomys talarum* (Rodentia: Octodontidae). *Journal of Mammalogy*, **70**, 204–208.

Busch C, Antinuchi CD, del Valle JC *et al*. (2000) Population ecology of subterranean rodents. In: *Life Underground: The Biology of Subterranean Rodents* (eds Lacey EA, Cameron G, Patton JL), pp. 183–226. The University of Chicago Press, Chicago, IL.

Castillo AH, Cortinas MN, Lessa EP (2005) Rapid diversification of South American tuco-tucos (*Ctenomys*; Rodentia, Ctenomyinae): contrasting mitochondrial and nuclear intron sequences. *Journal of Mammalogy*, **86**, 170–179.

Cook JA, Lessa EP (1998) Macroevolutionary patterns of diversification in subterranean rodents. The case of the South American tuco-tucos (genus Ctenomys, Rodentia: Octodontidae). *Evolution*, **52**, 1521–1527.

Degnan JH, Rosenberg NA (2006) Discordance of species trees with their most likely gene trees. *PLoS Genetics*, **2**, 762–768.

Durand E, Jay F, Gaggiotti OE, François O (2009) Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution*, **26**, 1963–1973.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Felsenstein J (1991) *PHYLIP (phylogeny inference package), version 3.4*. Department of Genetics, University of Washington, Seattle, WA.

Ferrati R, Canziani GA (2005) An analysis of water level dynamics in Esteros del Iberá wetland. *Ecological Modelling*, **186**, 17–27.

Ferrati R, Canziani GA, Ruiz Moreno D (2005) Esteros del Ibera: hydrometeorological and hydrological characterization. *Ecological Modelling*, **186**, 3–15.

François O, Ancelet S, Guillot G (2006) Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics*, **174**, 805–816.

François O, Currat M, Ray N, Han E, Excoffier L, Novembre J (2010) Principal component analysis under population genetic models of range expansion and admixture. *Molecular Biology and Evolution*, **27**, 1257–1268.

García L, Ponsà M, Egozcue J, García M (2000) Cytogenetic variation in *Ctenomys perrensi* (Rodentia, Octodontidae). *Biological Journal of the Linnean Society*, **71**, 615–624.

Giménez MD, Mirol PM, Bidau CJ, Searle JB (2002) Molecular analysis of populations of *Ctenomys* (Caviomorpha, Rodentia) with high karyotypic variability. *Cytogenetic and Genome Research*, **96**, 130–136.

Goudet J (1995) FSTAT Version 1.2: a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.

Guo S, Thompson E (1982) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

Hall TA (1999) BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hewitt GM (2001) Speciation, hybrid zones and phylogeography—or seeing genes in space and time. *Molecular Ecology*, **10**, 537–549.

Hey J, Machado CA (2003) The study of structured populations—new hope for a difficult and divided science. *Nature Review Genetics*, **4**, 535–543.

Hey J, Waples RS, Arnold ML, Butlin RK, Harrison RG (2003) Understandong and confronting species uncertainty in biology and conservation. *Trends in Ecology and Evolution*, **18**, 597–603.

Kittlein MJ, Gaggiotti OE (2008) Interactions between environmental factors can hide isolation by distance patterns: a case study of *Ctenomys rionegrensis* in Uruguay. *Proceedings of the Royal Society B*, **275**, 2633–2638.

Knowles LL, Carstens BC (2007a) Delimiting species without monophyletic gene trees. *Systematic Biology*, **56**, 887–895.

Knowles LL, Carstens BC (2007b) Estimating a geographically explicit model of population divergence. *Evolution*, **61**, 477–493.

Kocher TD, Thomas WK, Meyer A *et al.* (1989) Dynamics of mitochondrial-DNA evolution in animals—amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences, USA*, **86**, 6196–6200.

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*, **22**, 768–770.

Kumar SK, Tamura K, Nei M (2004) MEGA 3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.

Lacey EA (2001) Microsatellite variation in solitary and social tuco-tucos: molecular properties and population dynamics. *Heredity*, **86**, 628–637.

Lacey EA, Maldonado JE, Clabaugh JP, Matocq M (1999) Interspecific variation in microsatellites isolated from tuco-tucos (Rodentia: Ctenomyidae). *Molecular Ecology*, **8**, 1753–1768.

Lacey EA, Patton JL, Cameron GN (2000) *Life Underground: The Biology of Subterranean Rodents*. The University of Chicago Press, Chicago and London.

Lessa EP (2000) The evolution of subterranean rodents: a synthesis. In: *Life Underground: The Biology of Subterranean Rodents* (eds Lacey EA, Patton JL, Cameron GN), pp. 389–415. The University of Chicago Press, Chicago and London.

Lessa EP, Cook JA (1998) The molecular phylogenetics of tuco-tucos (genus *Ctenomys*, Rodentia: Octodontidae) suggests an early burst of speciation. *Molecular Phylogenetics and Evolution*, **9**, 88–99.

Magurran A (1998) Population differentiation without speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **353**, 275–286.

Manni F, Guérard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by Monmonier's algorithm. *Human Biology*, **76**, 173–190.

Mascheretti S, Mirol PM, Giménez MD, Bidau CJ, Contreras JR, Searle JB (2000) Phylogenetics of the speciose and chromosomally variable rodent genus *Ctenomys* (Ctenomyidae, Octodontoidae), based on mitochondrial cytochrome *b* sequences. *Biological Journal of the Linnean Society*, **70**, 361–376.

Massarini AI, Barros MA, Roig VG, Reig OA (1991) Banded karyotypes of *Ctenomys mendocinus* (Rodentia, Octodontidae) from Mendoza, Argentina. *Journal of Mammalogy*, **72**, 194–198.

Miller PM (2005) ALLELES IN SPACE: computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity*, **96**, 722–724.

Miller S, Diker D, Polesky H (1988) A Simple Salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acid Research*, **6**, 1215.

Mirol P, Giménez MD, Searle JB, Bidau CJ, Faulkes CG (2010) Population and species boundaries in the South American subterranean rodent *Ctenomys* in a dynamic environment. *Biological Journal of the Linnean Society*, **100**, 368–383.

Mora MS, Lessa EP, Cutrera AP, Kittlein MJ, Vassallo AI (2006) Phylogeography of the subterranean rodent *Ctenomys* australis in sand-dune habitats: evidence of population expansion. *Journal of Mammalogy*, **87**, 1192–1203.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution*, **19**, 153–170.

Novembre J, Stephens M (2008) Interpreting principal components analyses of spatial population genetic variation. *Nature Genetics*, **40**, 646–649.

Ortells MO (1995) Phylogenetic analysis of G-banded karyotypes among the South American subterranean rodents of the genus *Ctenomys* (Caviomorpha: Octodontidae), with special reference to chromosomal evolution and speciation. *Biological Journal of the Linnean Society*, **54**, 43–70.

Ortells MO, Contreras JR, Reig OA (1990) New *Ctenomys* karyotypes (Rodentia, Octodontidae) from north-eastern Argentina and from Paraguay confirm the extreme chromosomal multiformity of the genus. *Genetica*, **82**, 189–291.

Ota J (1993) *DISPAN Software. Genetic Distance and Phylogenetic Analysis*. University Park, Pennsylvania.

Parada A, D'Elía G, Bidau CJ, Lessa EP (2011) Species groups and the evolutionary diversification of tuco-tucos, genus *Ctenomys* (Rodentia: Ctenomyidae). *Journal of Mammalogy*, **92**, 671–682.

Patterson N, Price AL, Reich D (2006) Population structure and Eigenanalysis. *PLoS Genetics*, **2**, 190.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, **16**, 37–45.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

de Queiroz K (1998) *Endless Forms: Species and Speciation*. Oxford University Press, Oxford, UK.

de Queiroz K (2005) Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6600–6607.

Rambaut A, Drummond AJ (2007) *TRACER v1.4*. [http://beast.bio.ed.ac.uk/tracer].

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Reeves PA, Richards CM (2011) Species delimitation under the general lineage concept: an empirical example using wild North American hops (Cannabaceae*: Humulus lupulus*). *Systematic Biology*, **60**, 45–59.

Reig OA, Busch C, Ortells MO, Contreras JR (1990) An overview of evolution, systematics, population biology, cytogenetics, molecular biology and speciation in *Ctenomys*. In: *Evolution of Subterranean Mammals at the Organismal and Molecular Levels* (eds Nevo E and Reig OA), pp. 71–96. Alan R. Liss, NewYork.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution*, **16**, 351–358.

Rousset F (2007) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Schneider S, Kueffer JM, Roessli D, Excoffier L (1997) ARLEQUIN 1.1. A software for population genetic data analysis. URL: http://anthropologie.unige.ch/arlequin.

Shaffer HB, Thomson RC (2007) Delimiting species in recent radiations. *Systematic Biology*, **56**, 896–906.

Sites Jr JW, Marshall JC (2004) Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 199–227.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.

Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution*, **1**, 269–285.

Tomasco I, Lessa EP (2007) *The Quintessential Naturalist: Honoring the Life and Legacy of Oliver P. Pearson*. University of California Publications in Zoology Series, Berkeley, California.

Verzi DH (2002) Patrones de evolucion morfologica en Ctenomyinae (Rodentia: Octodontidae). *Mastozoolgıa Neotropical*, **9**, 309–328.

Vrba ES (1992) Mammals as a key to evolutionary theory. *Journal of Mammalogy*, **73**, 1–28.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for indentifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–1191.

Wlasiuk G, Garza JC, Lessa EP (2003) Genetic and geographic differentiation in the Rio Negro tuco-tuco (*Ctenomys rionegrensis*): inferring the roles of migration and drift from multiple genetic markers. *Evolution*, **57**, 913–926.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

The authors share a common interest in evolutionary biology. This paper was part of the PhD research of M.J.G.F., at the Museo Argentino de Ciencias Naturales 'Bernardino Rivadavia', whose interest include the use of molecular tools to understand evolutionary and ecological processes. P.M.M. supervised this work and is broadly interested in molecular ecological approaches to conservation of native species of Argentina. O.G.'s research focuses on developing theory and statistical methods aimed at bridging the gap between population ecology, population genetics and evolution.

Data accessibility

DNA sequences: GenBank accessions JQ686014–JQ686050. Microsatellite data and complete sequences data in the Supporting information section.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Correlation between geographic distance and genetic distance among all individuals based on (a) mitochondrial DNA and (b) microsatellites.

Fig. S2 Spatial representation of the individual-based PCA analysis.

Fig. S3 (a) Log likelihood values and (b) ΔK values as a function of K, the number of putative population clusters according to STRUCTURE.

Fig. S4 Results of the BARRIER analysis: a Delaunay triangulation and the position of the first 9 genetic barriers computed on a F_{ST} distance matrix between populations.

Table S1 F_{ST} values between sampling sites based on mitochondrial DNA (below the diagonal).

Table S2 Genetic variability in the sampling localities of *Ctenomys* analysed.

Table S3 Bidirectional estimates of migration using $K = 17$ as obtained in STRUCTURE.

Appendix S1 Genotypic data.

Appendix S2 Sequence data.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.