

Colonization Process of the Brazilian Common Vesper Mouse, *Calomys expulsus* (Cricetidae, Sigmodontinae): A Biogeographic Hypothesis

FABRÍCIA FERREIRA DO NASCIMENTO*, LUCIANA G. PEREIRA*, LENA GEISE, ALEXANDRA M. R. BEZERRA, PAULO S. D'ANDREA, AND CIBELE R. BONVICINO

From the Genetics Division, Instituto Nacional de Câncer, Rua André Cavalcante, 37, 4 andar, 20231-050, Rio de Janeiro, RJ, Brazil (Nascimento and Bonvicino); the Laboratório de Biologia e Parasitologia de Mamíferos Silvestres Reservatórios, Instituto Oswaldo Cruz, Fiocruz, Avenida Brasil, 4365, 21045-900, Rio de Janeiro, RJ, Brazil (Pereira, D'Andrea, and Bonvicino); the Laboratório de Mastozoologia, Departamento de Zoologia, IB, Universidade do Estado do Rio de Janeiro, Rua São Francisco Xavier 524, 220559-900, Rio de Janeiro, RJ, Brazil (Geise); and Departamento de Zoologia, Universidade de Brasília, 70910-900, Brasília, DF, Brazil (Bezerra).

*These authors contributed equally to this work.

Address correspondence to Cibele R. Bonvicino at the address above, or e-mail: cibelerb@inca.gov.br.

Abstract

Riverine barriers have been associated to genetic diversification and speciation of several taxa. The Rio São Francisco is one of the largest rivers in South America, representing the third largest river basin in Brazil and operating as a geographic barrier to gene flow of different taxa. To evaluate the influence of the Rio São Francisco in the speciation of small rodents, we investigated the genetic structure of *Calomys expulsus* with phylogenetic and network analyses of cytochrome *b* DNA. Our results suggested that *C. expulsus* can be divided into 3 subpopulations, 2 on the left and another one on the right bank of this river. The time of divergence of these subpopulations, using a Bayesian framework, suggested colonization from the south to the north/northeast. Spatial analysis using a clustering method and the Monmonier's algorithm suggested that the Rio São Francisco is a biogeographic barrier to gene flow and indicated that this river may play a role in the incipient speciation process of these subpopulations.

Key words: *Calomys*, cytochrome *b*, Phyllotini, phylogeography, Rio São Francisco, South America

Natural geographic barriers have been associated to genetic diversification and speciation of different taxa (Miller et al. 2006; Rodriguez-Robles et al. 2008). In South America, most studies have been focused on the Amazonas river basin (Patton et al. 2000; Cheviron et al. 2005; Funk et al. 2007), the largest basin in the world (Meade et al. 1991), although other watercourses have also been shown to be effective barriers for several vertebrates (Rodrigues 2003; Passoni et al. 2008). The Rio São Francisco is one of the longest rivers of South America, extending for 2830 km, and its basin is the third largest in Brazil, accounting for 7.6% of the country's territory and encompassing an area of approximately 645,000 km² (Godinho and Godinho 2003; Kohler 2003). This river irrigates several regions of different morphoclimatic domains and climate settings of 6 territories (States of Minas Gerais, Bahia, Pernambuco, Sergipe, Alagoas, part of Goiás, and

Distrito Federal). These regions extend from the humid Atlantic Forest to the drier domains of Caatinga and Cerrado (Godinho and Godinho 2003; IBAMA 2007), representing a geographic barrier for several taxa (Rodrigues 2003; Coutinho-Abreu et al. 2008; Passoni et al. 2008).

The genus *Calomys*, comprising Sigmodontinae species of the tribe Phyllotini, is widely distributed across South America, in Venezuela, Colombia, Peru, Bolivia, Paraguay, southeastern Brazil, Uruguay, and Argentina (Musser and Carleton 2005). Different species of *Calomys* inhabit seasonally dry tropical environments, from Cerrado to Caatinga (Almeida et al. 2007). The complex taxonomic arrangement of this genus is a consequence of the morphological similarities between congeneric species, a situation that is common in other Sigmodontinae genera. Hershkovitz (1962), based on morphological characters,

reported the first taxonomic review of *Calomys* with only 4 species, *Calomys lepidus*, *C. sorellus*, *C. callosus*, and *C. laucha*. Recently, karyological, taxonomic, and phylogenetic studies showed that a considerably higher diversity of this genus has been underestimated. Olds (1988) and Salazar-Bravo et al. (2001) recognized 10 *Calomys* species, albeit with different taxonomic arrangements, whereas other authors recognized 12 or 13 species (Musser and Carleton 2005; Almeida et al. 2007, respectively).

Steppan and Sullivan (2000) and Haag et al. (2007) postulated that *Calomys* is a monophyletic genus with high levels of genetic differentiation that took place during a long period of intrageneric divergence. Molecular phylogenetic studies divided *Calomys* in 2 groups, one with lowland species and another with highland species (Salazar-Bravo et al. 2001). Subsequently, phylogenetic analysis based on a higher number of species showed that the large-bodied lowland species (*C. fecundus*, *C. expulsus*, *C. venustus*, *C. tocantinsi*, *C. callosus*, *C. cerqueirai*, and an undescribed species from Beni) were included in a separate clade (Almeida et al. 2007; Bonvicino et al. 2010). *Calomys expulsus* has also been placed as the most basal lineage of the large-bodied group (Almeida et al. 2007).

Calomys shows an interesting pattern of geographic distribution and is one of the few Sigmodontinae genera restricted to dry habitats (Almeida et al. 2007). Furthermore, small-bodied size species are mainly found in open vegetation (strict or mixed grassland and savanna habitats), whereas large-bodied species, including *C. expulsus*, can also be found in forest formations (Almeida et al. 2007). *Calomys expulsus* is endemic to Brazil, occurring from Caatinga to Cerrado in the east and central-west regions, respectively. This species was the first to diverge from the large-bodied stock, probably by an initial invasion of a new habitat (forest formations) associated to change in body size (Almeida et al. 2007; Cordeiro-Estrela P, personal communication). These features and the fact that *C. expulsus* is the most ancient lineage among large-bodied species are indicative of the presence of structured populations.

To investigate whether the Rio S o Francisco has been acting as a barrier to *C. expulsus* dispersal and to identify structured populations, we analyzed mitochondrial cytochrome *b* (*cyt b*) DNA sequences from specimens collected along a wide range of its geographic distribution, both at the left and right bank of this river. Our results suggested that specimens could be subdivided in 3 subpopulations and that colonization occurred from the south to the north and northeast of Brazil. Our analyses also confirmed that the Rio S o Francisco represents a geographic barrier to *C. expulsus* subpopulations.

Materials and Methods

Sample Collection, DNA Extraction, Polymerase Chain Reaction (PCR) Assays, and Sequencing

Twenty-three *C. expulsus* were collected in 9 localities, in 5 Brazilian states along the left and right banks of the Rio S o

Francisco, mainly in regions of the Cerrado and Caatinga morphoclimatic domains (Figure 1a). DNA was isolated from livers preserved in 100% ethanol following Sambrook et al. (1989). *Cyt b* DNA (1143 bp) was PCR amplified using primers L14724 (Irwin et al. 1991) and CIT-REV (Casado et al. 2010). Amplicons were purified with GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, Brazil) and sequenced using the PCR primer L14724 and internal primers CitbAot (Menezes et al. 2010) and CB-in1 (Cassens et al. 2000). Sequencing was carried out using an ABI Prism 377 and 3730 platforms. Electropherograms were manually checked using BIOEDIT (version 7.0.8.0; Hall 1999) and CHROMAS (version 1.45; MacCarthy 1998).

Alignment, Test of Substitution Saturation, and Genetic Distance Estimation

We analyzed 23 *cyt b* sequences generated in this study and 49 *cyt b* GenBank sequences, 35 *C. expulsus*, 9 from other *Calomys* species, 1 *Auliscomys micropus*, 1 *Phyllotis magister*, 1 *Eligmodontia tytus*, 1 *Akodon montensis*, and 1 *Sigmodon hispidus*, which were manually aligned using BIOEDIT. These last 5 species were used as outgroups in phylogenetic trees. GenBank accession numbers are listed in Table 1 and Supplementary Table 1.

The DNA substitution saturation index (I_{ss} ; Xia et al. 2003), calculated with DAMBE (version 5.2.5; Xia and Xie 2001), was used for detecting phylogenetic signals. Two analyses were performed, one with *C. expulsus* with outgroup data and another with only *C. expulsus*. Results were compared with a critical I_{ss} value ($I_{ss,c}$) following Xia and Lemey (2009).

Pairwise genetic distances were estimated with the modified Log-Det for closely related taxa (Tamura and Kumar 2002) using MEGA (version 4; Tamura et al. 2007). The Hasegawa–Kishino–Yano (HKY) (Hasegawa et al. 1985) plus a proportion of invariable sites and gamma-distributed substitution rates (HKY + I + Γ) was selected as the DNA substitution model with MODELGENERATOR (version 0.84; Keane et al. 2006) using the Bayesian information criterion for phylogenetic reconstructions (see below).

Phylogenetic Reconstruction and Divergence Time Estimates

Phylogenetic reconstructions based on maximum likelihood (ML) were carried out with PHYML (version 3.0; Guindon and Gascuel 2003). The tree topology space was searched using the best of Nearest Neighbor Interchange and Subtree Pruning and Regrafting algorithms from 5 random starting trees generated by the BioNJ algorithm (Guindon and Gascuel 2003; Guindon et al. 2010). Branch support was calculated with the approximate likelihood ratio test (aLRT) with SH-like interpretation, a procedure that is conservative and accurate as bootstrapping but less computationally intensive (Anisimova and Gascuel 2006; Guindon et al. 2010).

Due to the scarcity of fossil records of *C. expulsus*, calibrations were carried out with the same constraints of

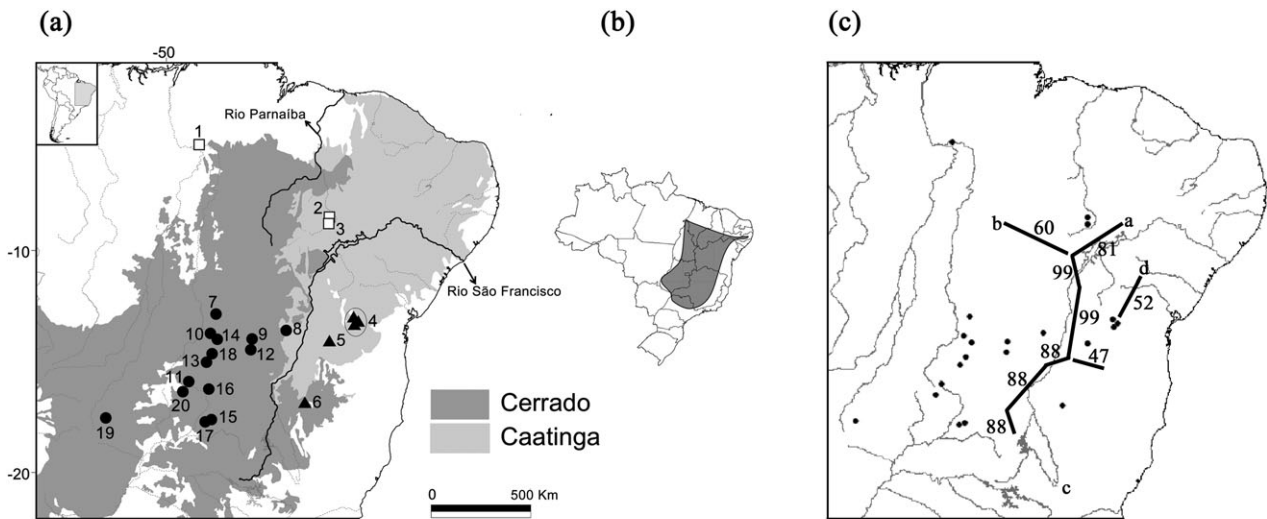


Figure 1. (a) Location of sites of collection of *Calomys expulsus*. See Supplementary Table 1 for a list of sites. The 3 subpopulation clusters identified by GENELAND are depicted with \square = group A, \bullet = group B, and \blacktriangle = group C. (b) Map showing geographic distribution of *C. expulsus*. (c) Biogeographic boundaries determined by the Monmonier's algorithm (a–d). Bootstrap support values are shown for each segment of each boundary.

Almeida et al. (2007) for estimating the time of divergence of specimens from the banks of the Rio São Francisco. The constrained nodes were 1) divergence of *Sigmodon* from *Oryzomayalia* (Steppan et al. 2004) estimated around 5–13 Ma based on fossil data of *Prosigmodon oroscoi* from Mexico and Phyllotini from the Montehermosan formation (Reig 1978) and previous molecular estimates (Smith and Patton 1999; Stepan et al. 2004); 2) cladogenesis of *C. expulsus*, with a lowest estimate of 0.2 Ma; and 3) divergence of Phyllotini from Akodontini tribes with a lowest estimate of 3.5 Ma (for details, see Almeida et al. 2007).

The tree topology and divergence times were subsequently co-estimated using a Bayesian inference with BEAST (version 1.5.3; Drummond and Rambaut 2007). Divergence times were estimated using an uncorrelated

lognormal relaxed clock (Drummond et al. 2006) and a Yule tree prior. The monophyly of *C. expulsus*, the divergence of Phyllotini and Akodontini, and the divergence of *Sigmodon* and *Oryzomayalia* were decided a priori based on previous phylogenies (Almeida et al. 2007; Haag et al. 2007). Estimates of divergence times were calibrated using a uniformly distributed prior based on calibration constraints. Estimates of the posterior distribution were obtained by Markov chain Monte Carlo (MCMC) sampling with samples drawn every 2000 MCMC steps over a total of 50 000 000 steps. Acceptable mixing and convergence to the stationary distribution were checked with TRACER 1.5 (Rambaut and Drummond 2007) and the first 10% were discarded as burn-in. The maximum clade credibility tree (MCCT) was computed from the trees sampled from the posterior with TREEANNOTATOR (version 1.5.3) of the BEAST package.

Table 1 List of GenBank accession numbers of *Calomys* and outgroup species used in phylogenetic analyses

Species	GenBank	Reference
<i>Calomys</i> sp.	AY033156	Salazar-Bravo et al. (2001)
<i>Calomys cerqueirai</i>	DQ447276	Almeida et al. (2007)
<i>C. fecundus</i>	AF385592	Salazar-Bravo et al. (2001)
<i>C. venustus</i>	AY033176	Salazar-Bravo et al. (2001)
<i>C. tocantinsi</i>	DQ447278	Almeida et al. (2007)
<i>C. callosus</i>	DQ447282	Almeida et al. (2007)
<i>C. laucha</i>	AF385593	Salazar-Bravo et al. (2001)
<i>C. tener</i>	DQ447301	Almeida et al. (2007)
<i>C. lepidus</i>	AF159294	Yates and Anderson (2000)
<i>Phyllotis magister</i>	AF484214	Kuch et al. (2002)
<i>Anliscomys micropus</i>	AF108690	Smith and Patton (1999)
<i>Eligmodontia typus</i>	AF108692	Smith and Patton (1999)
<i>Akodon montensis</i>	AF184055	Geise et al. (2001)
<i>Sigmodon hispidus</i>	AF425227	Carrol DS and Bradley RD (unpublished data)

Intraspecific Analyses

NETWORK (version 4.5.1.6; available at <http://www.fluxus-engineering.com>) was used for reconstructing a median-joining (MJ) network (Bandelt et al. 1999) to evaluate subpopulation structure and geographic distribution patterns. MJ was calculated using variable sites only.

Distribution of pairwise differences or mismatch distributions (Slatkin and Hudson 1991; Rogers and Harpending 1992) were used for estimating demographic events with ARLEQUIN (version 3.5.1.2; Excoffier and Lischer 2010). Confidence intervals were calculated with 1000 bootstrap replicates and significance of $\alpha = 0.05$ (Schneider and Excoffier 1999). The sudden and spatial demographic expansion models were tested using bootstrap and the sum of square deviations between the observed

mismatch distribution and simulated data as a test statistic (test of goodness of fit; Schneider and Excoffier 1999; Excoffier 2004). ARLEQUIN was used for estimating indices of nucleotide (π) and haplotype (h) diversity (Nei 1987).

To distinguish models of population growth from the null hypothesis of constant population size, Fu's F_S (Fu 1997) and the R_2 neutrality tests (Ramos-Onsins and Rozas 2002) were estimated with ARLEQUIN and DNASP (version 5.10.01; Librado and Rozas 2009), respectively. The probability of neutrality tests was calculated based on 1000 replicates. Fu's F_S and the R_2 are most powerful for detecting sudden population growth or contraction (Ramos-Onsins and Rozas 2002; Ramirez-Soriano et al. 2008) and because Fu's F_S is recommendable for large populations and R_2 for small ones (Ramos-Onsins and Rozas 2002).

Spatial Analysis

The Mantel test included in the isolation by distance (IBD) web service (version 3.15; Bohonak 2002; Jensen et al. 2005) was used to analyze subpopulations according to the IBD model, postulating that populations separated by longer geographic distances would show higher genetic differences. Our sample was divided into 3 subpopulations according to phylogenetic and network data, and geographic distances were calculated using the most distant and closest points in the map (localities 1, 5, and 19, and 2, 5, and 8, respectively; Figure 1a).

The genetic structure of populations was analyzed with GENELAND (version 3.2.2; Guillot, Estoup, et al. 2005; Guillot, Santos, and Estoup 2005) using a Bayesian model based on the MCMC. The number of clusters was determined by an independent, 10 times run of MCMC, with K (number of populations) varying from 1 to 10. The number of MCMC interactions was set to 100 000 per run with a thinning of 100. We used the correlated frequency model setting the uncertainty on coordinates to 0 km, and the maximum number of nuclei in the Poisson-Voronoi tessellation was fixed at 300 (default option). Analyses were performed using variable sites only.

BARRIER (version 2.2; Manni et al. 2004) was used to compute the Monmonier's maximum-difference algorithm (Monmonier 1973) for identifying biogeographic boundaries or areas with largest genetic differences between population pairs (see Manni and Gu erard 2004).

The Monmonier's algorithm is supervised because the number of biogeographic boundaries to be computed must be specified in advance. A similar approach to Patten and Smith-Patten (2008) was followed by a priori selection of the maximum pairwise genetic distance between haplotypes. This cutoff was determined by the mean of the maximum and minimum pairwise genetic distances between haplotypes.

Bootstrap values for the computed boundaries were estimated from multiple genetic distance matrices with BARRIER. A set of 100 DNA bootstrapped sequences were generated with SEQBOOT using PHYLIP version 3.69 (Felsenstein 2004), and subsequently, 100 genetic distance matrices were calculated using DNADIST (PHYLIP). These

multiple matrices were used to assess the statistical significance of boundaries calculated by BARRIER (Manni et al. 2004). Only one sequence per haplotype was used to perform these calculations.

Results

Cyt *b* DNA, comprising 1143 bp from 23 specimens, showed 17 haplotypes. These data, together with 35 *C. expulsus* sequences retrieved from GenBank, accounted for 35 haplotypes, some of which present in more than one locality (Supplementary Table 1). Furthermore, haplotype H17 (Supplementary Table 1) was the most frequent one and present in different localities of Goi as state. Very low saturation was observed in data sets with *C. expulsus* and outgroup sequences or with only *C. expulsus*, with I_{ss} significantly lower than $I_{ss,c}$ for a symmetrical tree topology (see Xia and Lemey 2009).

The ML topology was similar to the MCCT except for group C haplotypes. Because of the similarity of topologies, only the ML tree is shown with divergence dates estimated by Bayesian inference for the main *C. expulsus* groups (Figure 2). Both ML and MCCT showed that *C. expulsus* split in 2 well-supported clades designated A and B and several nonresolved lineages included in group C. Group A comprised haplotypes from the northern geographic range of the area under study (Figure 1a), whereas group B comprised haplotypes from the center-west of Brazil, on the left bank of the Rio S o Francisco (Figure 1a). Finally, group C haplotypes were distributed in the middle section of the right bank of the Rio S o Francisco (Figure 1a). This set comprised 10 haplotypes (H4–H13) that did not group in a single clade. Four of them (H8–9 and H11–12) grouped in a well-supported clade, whereas the relation of the remaining haplotypes (H4–7, H10, and H13) was not well defined in the ML tree (Figure 2). Furthermore, group A was more closely related to group B than to group C (Figure 2).

Analysis of all *C. expulsus* sequences showed 83 variable sites (69 transitions and 17 transversions), and genetic distance estimates ranged from 0.000 to 0.027. Intra- and intergroup genetic distance estimates were similar, as well as number of variable sites, transitions, and transversions between groups B and C. Nucleotide and haplotype diversity were slightly higher for group C (0.009 ± 0.005 and 0.97 ± 0.04 , respectively) than for group B (0.004 ± 0.002 and 0.90 ± 0.03 , respectively). Genetic distance estimates as well as nucleotide and haplotype diversity of group A were not comparable to other groups due to its small sample size.

Group A showed 3 haplotypes, each one in a different locality, in Cel. Jos e Dias and Jos e Costa in Piaui state, and Buriti in Tocantins state (localities no. 3, 2, and 1, respectively; Figure 1a). Group B showed 22 haplotypes, with localities showing more than one haplotype. Haplotype H17 was shared by specimens from 5 localities (no. 10, 13, 16, 18, and 19; Supplementary Table 1 and Figure 1a), up to 660 km apart from one another (from Serra da Mesa [no. 10] to Mineiros

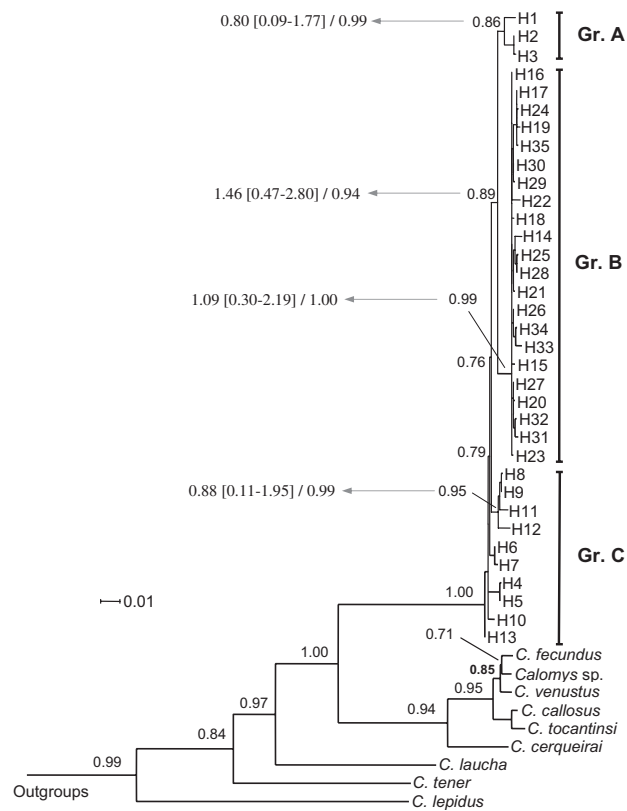


Figure 2. ML phylogeny of *Calomys expulsus*. Numbers close to branches are aLRT estimates. For clarity, only aLRT estimates for the main internal groups for *C. expulsus* are shown. Groups described in text are highlighted. Height and 95% highest posterior density interval (in brackets) of date estimates followed by their posterior probability are also shown for *C. expulsus* main groups.

[no. 19] municipality). Haplotype H16 occurred in 3 localities (no. 9, 12, and 13), as far as 50 km from one another (between Correntina [no. 9] and Mimoso de Goiás [no. 13] municipalities). Haplotype H26 occurred in 3 localities (no. 12, 15, and 17), up to 450 km apart from one another (between Mambai [no. 12] and Ipameri [no. 17] municipalities). Finally, group C showed 10 haplotypes across 3 localities in the states of Minas Gerais (Juramento [no. 6] municipality) and Bahia (Caetitê [no. 5] municipality and Chapada Diamantina [no. 4]).

MJ network (Figure 3) showed 2 well-defined groups (A and B) previously identified by phylogenetic reconstructions, whereas group C haplotypes were clustered in 2 subgroups more related to one another than to A or B haplotypes. Group A did not show any internal median vector but a high number of nucleotide substitutions were observed between haplotypes H2–3 and H1, which might be explained by the large distance between their localities (Figure 1a and Supplementary Table 1). Group B showed 7 median vectors and up to 3 nucleotide substitutions from one haplotype to another or median vector. The connection between haplotype H22 and H16 was exceptional in showing 5 nucleotide substitutions. Finally, group C showed

7 median vectors, one of which apparently playing a central role by connecting 2 smaller subgroups from the right bank of the Rio São Francisco, one with haplotypes H8–9 and H11–12 and another with H4–7, H10, and H13. Interestingly, both subgroups contained haplotypes from Caetitê (no. 5) and Chapada Diamantina (no. 4), indicating lack of geographic structure in group C (Supplementary Table 1; Figure 1a).

Estimates of the time of radiation of *C. expulsus* and of other large-bodied *Calomys* species (*Calomys* sp., *C. fecundus*, *C. venustus*, *C. tocantinsi*, *C. callosus*, and *C. cerqueirai*) were very similar (2.03 and 1.90 Ma, respectively). Divergence of *C. expulsus* was not synchronic, indicating that group A (approximately 0.80 Ma) is more recent than group B (approximately 1.09 Ma).

Mismatch distribution for group B showed a bimodal curve with a higher P value indicating a spatial expansion model (Figure 4a). On the other hand, group C showed a multimodal curve with similar P value for both sudden and spatial expansion models (Figure 4b). All observed data (Figure 4) fitted their respective simulated curves describing either a sudden or spatial expansion models. Mismatch distribution was not estimated for group A because only 3 specimens and 3 haplotypes were found.

Fu's F_S and R_2 statistical neutrality tests in group B showed an estimate of -10.93 ($P = 0.0$) and 0.05 ($P = 0.0$), respectively, suggesting population growth. On the other hand, Fu's F_S and R_2 tests in group C showed an estimate of -1.25 ($P = 0.24$) and 0.12 ($P = 0.12$), respectively, suggesting a constant population size.

The Mantel test did not reveal significant IBD using raw data ($P \approx 0.50$); in fact, similar P values were observed when genetic and/or geographic distances were log transformed (Bohonak 2002). On the other hand, all 10 runs with GENELAND resulted in a K estimate = 3, with population partitions identical with those observed in Figure 1a.

Using the mean of the maximum and minimum pairwise genetic distance, the cutoff stipulating the number of potential boundaries was 0.13. This allowed the identification of 4 barriers separating *C. expulsus* haplotypes (Figure 1c). However, only barrier 1 showed a high bootstrap value (from 88% to 99%) contrary to the 3 other likely “spurious” barriers with low bootstrap support (below 60%). BARRIER 1 coincided with the location of the Rio São Francisco (Figure 1c).

Discussion

Calomys expulsus comprised 3 groups, A, B, and C, the first 2 being well defined and monophyletic, comprising animals from the left bank of the Rio São Francisco. The arrangement of group C, with animals from the right bank of the Rio São Francisco, varied according to different analyses (Figures 2 and 3). These phylogenetic subdivisions were not coincident with different ecoregions where *C. expulsus* specimens were collected (Dinerstein et al. 1995). Group A included individuals collected both in the Mato Grosso Tropical Dry Forest and in Caatinga. Group B

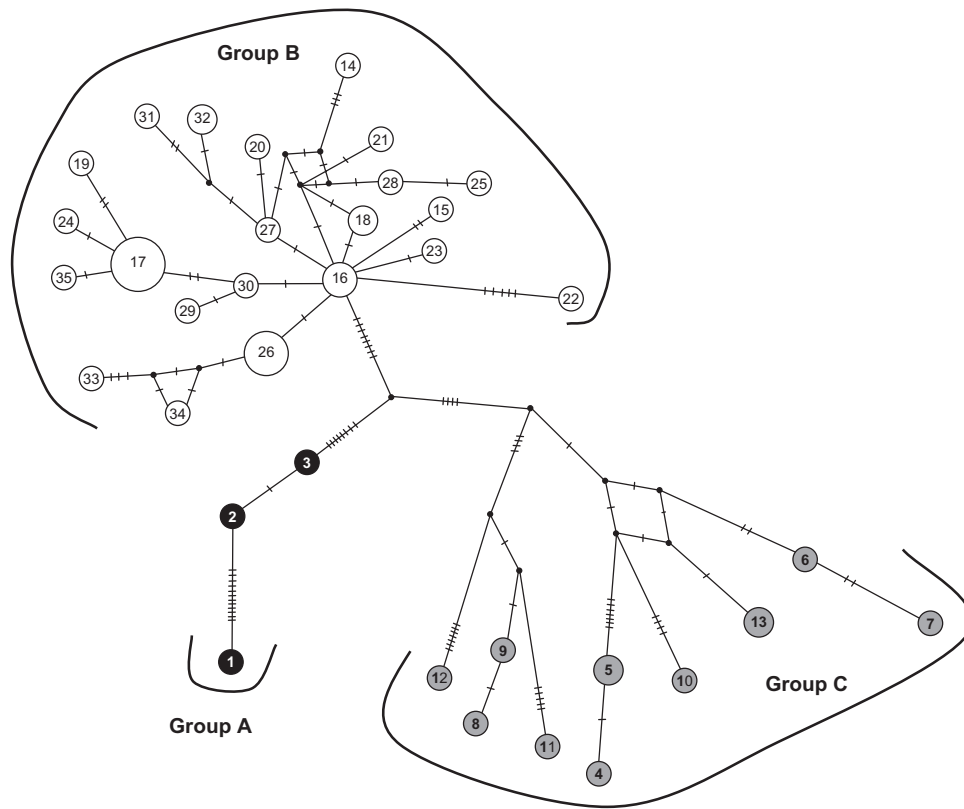


Figure 3. MJ network of *Calomys expulsus*. Numbers inside circles indicate haplotypes listed in Supplementary Table 1. White circles represent group A, black circles represent group B, and gray circles represent group C. Circle sizes correspond to number of individuals carrying a given haplotype. Black points represent median vectors. Small lines to connecting branches denote the number of nucleotide substitutions. Branches are not to scale.

included individuals from Cerrado and the Brazilian Northeastern Dry Forest, and group C included specimens from Cerrado, Brazilian Northeastern Dry Forest, and Caatinga (Figure 1a). Our analyses, however, showed structured populations of *C. expulsus* according to their geographic distributions, 2 on the left and another one at the right bank of the Rio São Francisco.

At the population level, networks are more appropriate than traditional phylogenetic reconstructions because they consider ancestral nodes, multifurcations, and reticulations (Crandall and Templeton 1993; Posada and Crandall 2001). A low level of genetic diversity is expected between populations, leading to lack of phylogenetic resolution using strictly dichotomous trees (traditional methods; Morrison 2005; Woolley et al. 2008). In this respect, MJ network data were more appropriate than ML, a reason why haplotypes from the right bank of the river were included in group C. Specimens representing this group were separated by several median vectors. This might be because some extant haplotypes have not been detected or, alternatively, due to lack of intermediate haplotypes (Bandelt et al. 1999).

Despite lack of known barriers between groups A and B (Figure 1a), they appeared as sister clades in all phylogenetic analyses, suggesting a likely vicariant distribution. Group A

occurs in the lowlands of the Tocantins and the Parnaíba river valleys at altitudes as high as 300 m, whereas group B is found at higher altitudes (up to 1300 m) in the Central Plateau of Goiás state, suggesting that these groups are separated by altitude. This separation was detected by GENELAND analysis (Figure 1a) and the Monmonier's algorithm (barrier b; Figure 1c), although barrier b was detected with low bootstrap support (60%; Figure 1c). Separation of group C from A and B was supported by all analyses (Figures 1a, 2, and 3), suggesting that the Rio São Francisco is a geographic barrier to gene flow for *C. expulsus* subpopulations. A boundary (barrier a; Figure 1c) with high bootstrap support (from 81% to 99%) was placed by the Monmonier's algorithm at the same location of the Rio São Francisco.

The Bayesian clustering method of GENELAND usually overestimates genetic structure in samples characterized by IBD (Frantz et al. 2009; Guillot et al. 2009). However, the population structure of *C. expulsus* could not be explained by IBD, suggesting that the 3 clusters identified by GENELAND were not spurious. Thus, the use of the Monmonier's algorithm, which does not follow a clear statistical principle (Guillot et al. 2009), coupled with a clustering method suggested a strong differentiation between these 3 *C. expulsus* populations, corroborating the role of the Rio São Francisco as a geographic barrier to gene flow.

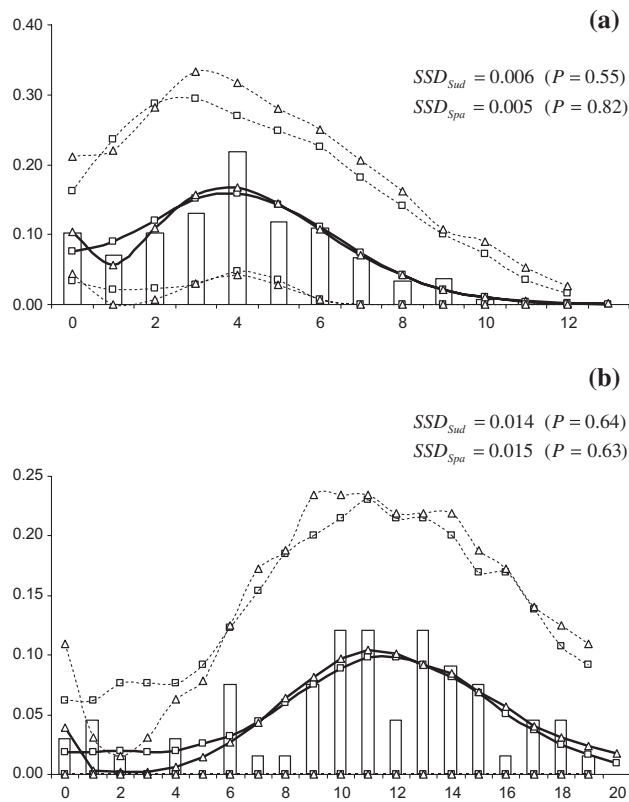


Figure 4. Mismatch distribution (a) for group B and (b) for group C. White bars: observed distribution; solid line and squares: simulated distribution under the demographic expansion model; solid line and triangles: simulated distribution under the spatial expansion model. Dashed lines: respective 95% and 5% percentiles. X axis: number of pairwise differences. Y axis: frequency of observed and simulated differences. SSD_{Sud} and SSD_{Spa} = sum of square deviations under the sudden (Sud) and spatial (Spa) demographic expansion models, respectively.

Calomys expulsus specimen OT5185 (GenBank: AY964030) from Ipameri (Goiás state) has been previously designated “*C. expulsus* variant 1” ($2n = 64$, FN = 66) because its chromosome complement differed from the standard karyotype for this species (Haag et al. 2007). However, this specimen shared the same H32 haplotype with MN72050 (a specimen with $2n = 66$, FN = 68, from Campo Alegre de Goiás, GO [locality no. 20; Supplementary Table 1 and Figure 1a]; Bonvicino CR, personal communication). Haag et al. (2007), analyzing 26 *C. expulsus* specimens from 3 localities in the left bank of the Rio São Francisco, found that haplotype from different localities were grouped in separate clusters. Conversely, in the present study, comprising a larger sample size and a more extensive geographic coverage, haplotypes were not restricted to localities (Figure 3). Mismatch distributions of both groups B and C suggested population expansions contrary to neutrality tests, suggesting expansion only for group B, whereas group C appeared to remain under constant

population size. Mismatch distribution of group B showed a considerably higher P value for the spatial expansion model (Figure 4), which, together with neutrality tests, suggested that this population from the left bank of the Rio São Francisco may be expanding its range and population size. This hypothesis was confirmed by divergence time estimates (see below).

Historical Colonization of *C. expulsus* and the Influence of the Rio São Francisco

The wide range of divergence time estimates for *C. expulsus* and outgroups suggested the need of additional paleontological data of Sigmodontinae to improve calibration points, as already pointed out by Almeida et al. (2007).

The ancestral form of the large-bodied *Calomys* species (split into 2 lineages that diverged simultaneously during the Pliocene [approximately 2.0 Ma]), one leading to *C. expulsus* and another leading to the other large-bodied species (*Calomys* sp., *C. fecundus*, *C. venustus*, *C. tocantinsi*, *C. callosus*, and *C. cerqueirai*).

The estimates of divergence of the different *C. expulsus* groups suggested that colonization occurred from south to north and northeast of its current range. Group A and part of group C probably migrated more recently (ca. 0.80 Ma) and are presently found in the most northern and eastern range of *C. expulsus*, whereas group B (1.09 Ma) is slightly older and more centrally distributed. In addition, lack of monophyletic arrangement of group C in the ML topology pointed to different and successive migrations to the right bank of the Rio São Francisco. This and analyses of mismatch distribution and neutrality tests suggested that *C. expulsus* is spatially and demographically expanding.

The emergence of *C. expulsus*, estimated approximately 2.0 Ma, during the Pliocene, occurred after the rise of the Rio São Francisco (estimated at 65 Ma; Valadão 1998). This scenario, together with the finding of the oldest *C. expulsus* subpopulation on the left bank of this river, suggested that the right bank was occupied later following migrations across regions south of its headwaters. Furthermore, the higher diversity observed in the left bank haplotypes also suggested that these migrations took place several times. The Rio São Francisco barrier may also explain the vicariance of endemic lizards restricted to each margin, such as *Amphisbaena bastata* and *A. ignatiana* (Rodrigues 2003), corroborating its relevant role in deterring gene flow for nonvolant vertebrates leading to the incipient speciation of *C. expulsus* subpopulations.

Funding

The Instituto Brasileiro do Meio Ambiente e Recursos Naturais Renováveis and Instituto Chico Mendes de Conservação da Biodiversidade; Conselho Nacional de Desenvolvimento Científico e Tecnológico, Ministério da Ciência e Tecnologia, Brazil (to L.G., C.R.B., fellowships to L.G.P., A.M.R.B., F.F.N., C.R.B., L.G.); Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de

Janeiro (FAPERJ), Secretaria da Ciência e Tecnologia, Brazil (to C.R.B.); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Ministério da Educação; Universidade do Estado do Rio de Janeiro, Brazil (to L.G.).

Supplementary Material

Supplementary material can be found at <http://www.jhered.oxfordjournals.org/>.

Acknowledgments

We are grateful to Dr R. Cerqueira (UFRJ, Brazil) for kindly donating samples for present work and to H.N. Seuánez (INCA, Brazil) and J.L. Patton (Museum of Vertebrate Zoology, U. C. Berkeley) for reviewing a previous version of the manuscript. We are thankful to the staff of the Genome Sequencing Platform PDTIS/FIOCRUZ, to Dr S.Y.W. Ho (ANU, Australia) for discussion on Bayesian analysis and divergence time estimates, and to Assistant Prof. Michael A. Patten (University of Oklahoma, USA) for helpful discussion on the Monmonier's algorithm.

References

Almeida FC, Bonvicino CR, Cordeiro-Estrela P. 2007. Phylogeny and temporal diversification of *Calomys* (Rodentia, Sigmodontinae): implications for the biogeography of an endemic genus of the open/dry biomes of South America. *Mol Phylogenet Evol.* 42:449–466.

Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst Biol.* 55: 539–552.

Bandelt HJ, Forster P, Rohlf A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 16:37–48.

Bohonak AJ. 2002. IBD (Isolation by Distance): a program for analyses of isolation by distance. *J Hered.* 93:153–154.

Bonvicino CR, Oliveira JA, Gentile R. 2010. A new species of *Calomys* (Rodentia: Sigmodontinae) from Eastern Brazil. *Zootaxa.* 2336:19–25.

Casado F, Bonvicino CR, Nagle C, Comas B, Manzur TD, Lahoz MM, Seuánez HN. 2010. Mitochondrial divergence between 2 populations of the Hooded Capuchin, *Cebus (Sapajus) cay* (Platyrrhini, Primates). *J Hered.* 101:261–269.

Cassens I, Vicario S, Waddell VG, Balchowsky H, Van Belle D, Ding W, Fan C, Mohan RS, Simoes-Lopes PC, Bastida R, et al. 2000. Independent adaptation to riverine habitats allowed survival of ancient cetacean lineages. *Proc Natl Acad Sci USA.* 97:11343–11347.

Chevron ZA, Hackett SJ, Capparella AP. 2005. Complex evolutionary history of a Neotropical lowland forest bird (*Lepidothrix coronata*) and its implications for historical hypotheses of the origin of Neotropical avian diversity. *Mol Phylogenet Evol.* 36:338–357.

Coutinho-Abreu IV, Sonoda IV, Fonseca JA, Melo MA, Balbino VQ, Ramalho-Ortigao M. 2008. *Lutzomyia longipalpis* s.l. in Brazil and the impact of the São Francisco River in the speciation of this sand fly vector. *Parasit Vectors.* 1:16.

Crandall KA, Templeton AR. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics.* 134:959–969.

Dinerstein E, Olson DM, Graham DJ, Webster AL, Primm SA, Bookbinder MP, Ledec G. 1995. A conservation assessment of the terrestrial ecoregions of Latin America and the Caribbean. Washington (DC): WWF and World Bank.

Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.

Excoffier L. 2004. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Mol Ecol.* 13:853–864.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10:564–567.

Felsenstein J. 2004. PHYLIP (Phylogeny Inference Package) version 3.6. Seattle (WA): Department of Genome Sciences, University of Washington Distributed by the author

Frantz AC, Cellina S, Krier A, Schley L, Burke T. 2009. Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *J Appl Ecol.* 46:493–505.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics.* 147:915–925.

Funk WC, Caldwell JP, Peden CE, Padial JM, De la Riva I, Cannatella DC. 2007. Tests of biogeographic hypotheses for diversification in the Amazonian forest frog, *Physalaemus petersi*. *Mol Phylogenet Evol.* 44:825–837.

Geise L, Smith MF, Patton JL. 2001. Diversification in the genus Akodon (Rodentia: Sigmodontinae) in southeastern South America: mitochondrial DNA sequence analysis. *J Mammal.* 82:92–101.

Godinho AL, Godinho HP. 2003. Breve visão do São Francisco. In: Godinho AL, Godinho HP, editors. *Águas, peixes, e pescadores do São Francisco das Minas Gerais.* Belo Horizonte (Brazil): PUC Minas. p. 15–24.

Guillot G, Estoup A, Mortier F, Cosson JF. 2005. A spatial statistical model for landscape genetics. *Genetics.* 170:1261–1280.

Guillot G, Leblois R, Coulon A, Frantz AC. 2009. Statistical methods in spatial genetics. *Mol Ecol.* 18:4734–4756.

Guillot G, Santos F, Estoup A. 2005. Geneland: a computer package for landscape genetics. *Mol Ecol Notes.* 5:708–711.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 52:696–704.

Haag T, Muschner VC, Freitas LB, Oliveira LFB, Langguth AR, Mattevi MS. 2007. Phylogenetic relationships among species of the genus *Calomys* with emphasis on South American lowland taxa. *J Mammal.* 88:769–776.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 22:160–174.

Hershkovitz P. 1962. Evolution of Neotropical cricetine rodents (Muridae), with special reference to the Phyllotini group. *Fieldiana Zool.* 46:1–524.

IBAMA2007. Programa de revitalização da bacia hidrográfica do São Francisco: estatística de desembarque pesqueiro; censo estrutural da pesca 2006. Brasília (Brazil): IBAMA.

Irwin DM, Kocher TD, Wilson AC. 1991. Evolution of the cytochrome *b* gene of mammals. *J Mol Evol.* 32:128–144.

Jensen JL, Bohonak AJ, Kelley ST. 2005. Isolation by distance, web service. *BMC Genet.* 6:13.

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO. 2006. Assessment of methods for amino acid matrix selection and their use on

- empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol.* 6:29.
- Kohler HC. 2003. Aspectos geocológicos da bacia hidrográfica do São Francisco (primeira aproximação na escala 1:1 000 000). In: Godinho HP, Godinho AL, editors. *Águas, peixes e pescadores do São Francisco das Minas Gerais*. Belo Horizonte (Brazil): PUC Minas. p. 25–35.
- Kuch M, Rohland N, Betancourt JL, Latorre C, Steppan S, Poinar HN. 2002. Molecular analysis of an 11,700-year-old rodent midden from the Atacama Desert, Chile. *Mol Ecol.* 11:913–924.
- Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 25:1451–1452.
- MacCarthy C. 1998. Chromas version 1.45 (32-bit). Southport (Australia): School of Health Science, Griffith University.
- Manni F, Guérard E. 2004. Barrier version 2.2 user's manual. Paris (France): Population genetics team, Musée de l'Homme.
- Manni F, Guérard E, Heyer E. 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Hum Biol.* 76:173–190.
- Meade RH, Rayol JM, da Conceição SC, Natividade JRG. 1991. Backwater effects in the Amazon River basin of Brazil. *Environ Geol.* 18:105–114.
- Menezes AN, Bonvicino CR, Seuanes HN. 2010. Identification, classification and evolution of Owl Monkeys (*Aotus*, Illiger 1811). *BMC Evol Biol.* 10:248.
- Miller MP, Haig SM, Wagner RS. 2006. Phylogeography and spatial genetic structure of the Southern torrent salamander: implications for conservation and management. *J Hered.* 97:561–570.
- Monmonier MS. 1973. Maximum-difference barriers: an alternative numerical regionalization method. *Geogr Anal.* V:245–261.
- Morrison DA. 2005. Networks in phylogenetic analysis: new tools for population biology. *Int J Parasitol.* 35:567–582.
- Musser GG, Carleton MD. 2005. Family Muridae. In: Wilson DE, Reeder DM, editors. *Mammal species of the world: a taxonomic and geographic reference*. 3rd ed. Baltimore (MD): Johns Hopkins University Press. p. 1189–1531.
- Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Olds N. 1988. A revision of the genus *Calomys* (Rodentia: Muridae). New York: City University of New York.
- Passoni JC, Benozzati ML, Rodrigues MT. 2008. Phylogeny, species limits, and biogeography of the Brazilian lizards of the genus *Eurolephosaurus* (Squamata: Tropiduridae) as inferred from mitochondrial DNA sequences. *Mol Phylogenet Evol.* 46:403–414.
- Patten MA, Smith-Patten BD. 2008. Biogeographical boundaries and Monmonier's algorithm: a case study in the northern Neotropics. *J Biogeogr.* 35:407–416.
- Patton JL, da Silva MNF, Malcon JR. 2000. Mammals of the Rio Juruá and the evolutionary and ecological diversification of Amazonia. *Bull. Am. Mus. Nat. Hist.* 244:1–306.
- Posada D, Crandall KA. 2001. Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol Evol.* 16:37–45.
- Rambaut A, Drummond AJ. 2007. Tracer v1.4 [Internet]. [cited 2010 June 2]. Available from: URL <http://beast.bio.ed.ac.uk/Tracer>
- Ramirez-Soriano A, Ramos-Onsins SE, Rozas J, Calafell F, Navarro A. 2008. Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics.* 179:555–567.
- Ramos-Onsins SE, Rozas J. 2002. Statistical properties of new neutrality tests against population growth. *Mol Biol Evol.* 19:2092–2100.
- Reig OA. 1978. Roedores cricetidos del Plioceno superior de la provincia de Buenos Aires (Argentina). *Publ Mus Munic Cienc Nat Mar del Plata "Lorenzo Scaglia"*. 2:164–190.
- Rodrigues MT. 2003. Herpetofauna da Caatinga. In: Leal IR, Tabarelli M, Silva JMC, editors. *Ecologia e conservação da Caatinga*. Recife (Brazil): Editora Universitária da Universidade Federal de Pernambuco. p. 181–236.
- Rodríguez-Robles JA, Jezkova T, Leal M. 2008. Genetic structuring in the threatened "Lagartijo del Bosque Seco" (*Anolis cooki*) from Puerto Rico. *Mol Phylogenet Evol.* 46:503–514.
- Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol.* 9:552–569.
- Salazar-Bravo J, Drago J, Tinnin DS, Yates TL. 2001. Phylogeny and evolution of the neotropical rodent genus *Calomys*: inferences from mitochondrial DNA sequence data. *Mol Phylogenet Evol.* 20:173–184.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press.
- Schneider S, Excoffier L. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics.* 152:1079–1089.
- Slatkin M, Hudson RR. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics.* 129:555–562.
- Smith MF, Patton JL. 1999. Phylogenetic relationships and the radiation of Sigmodontine rodents in South America: evidence from cytochrome *b*. *J Mammal Evol.* 6:89–128.
- Steppan S, Adkins R, Anderson J. 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Syst Biol.* 53:533–553.
- Steppan SJ, Sullivan J. 2000. The emerging statistical perspective in systematic biology: a reply to Mares and Braun. *J Mammal.* 81:260–270.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol.* 24:1596–1599.
- Tamura K, Kumar S. 2002. Evolutionary distance estimation under heterogeneous substitution pattern among lineages. *Mol Biol Evol.* 19:1727–1736.
- Valadão RC. 1998. *Evolução de longo-termo do relevo do Brasil oriental (desnudação, superfícies de aplanamento e soerguimentos crustais)* [PhD]. Salvador (BA): Universidade Federal da Bahia. p. 230.
- Woolley SM, Posada D, Crandall KA. 2008. A comparison of phylogenetic network methods using computer simulation. *PLoS One.* 3:e1913.
- Xia X, Lemey P. 2009. Assessing substitution saturation with DAMBE. In: Salemi M, Vandamme A-M, Lemey P, editors. *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. 2nd ed.. Cambridge (UK): Cambridge University Press. p. 615–630.
- Xia X, Xie Z. 2001. DAMBE: software package for data analysis in molecular biology and evolution. *J Hered.* 92:371–373.
- Xia X, Xie Z, Salemi M, Chen L, Wang Y. 2003. An index of substitution saturation and its application. *Mol Phylogenet Evol.* 26:1–7.
- Yates TL, Anderson S. 2000. A new genus and species of phyllotine rodent from Bolivia. *J Mammal.* 81:18–36.

Received June 7, 2010; Revised January 11, 2011;
Accepted January 26, 2011

Corresponding Editor: Warren Johnson