Restricted Gene Flow in *Calomys musculinus* (Rodentia, Muridae), the Natural Reservoir of Junin Virus

M. B. Chiappero and C. N. Gardenal

From the Cátedra de Bioquímica y Biología Molecular, Facultad de Ciencias Médicas, Universidad Nacional de Córdoba, Casilla de Correo 35, Sucursal 16, 5016 Córdoba, Argentina.

Address correspondence to C. N. Gardenal at the address above, or e-mail: ngardenal@biomed.fcm.unc.edu.ar.

Abstract

Random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) markers were used to evaluate the relative contribution of gene flow as a determinant of the population genetic structure of the wild rodent *Calomys musculinus* (the reservoir of Argentine hemorrhagic fever [AHF]) in central Argentina. One hundred eighty-seven individuals from 13 populations (9 of them from the endemic zone of AHF and 5 from areas outside it) were analyzed using 78 polymorphic RAPD loci. Genetic variation within each population was high; each individual was characterized by a unique RAPD phenotype. *C. musculinus* populations showed a moderate to high genetic subdivision and a random pattern of differentiation. Populations separated by the same geographic distance showed very different degrees of genetic divergence. The results indicate that populations of *C. musculinus* have colonized their present ranges relatively recently and differentiation by genetic drift has proceeded faster than homogenization by gene flow at the macrogeographic scale analyzed (10–700 km).

*Calomys musculinus* is a wild murid rodent widely distributed throughout most of Argentina. It is an opportunistic species, characterized by a high reproductive rate and the ability to colonize a wide variety of habitats, particularly disturbed ones like crop fields (Crespo et al. 1970; de Villafañe et al. 1988). Studies on *C. musculinus* populations are of particular interest because this species is the natural reservoir of Junin virus, the causative agent of Argentine hemorrhagic fever (AHF), a severe human disease endemic in part of the phytogeographic region called the “humid pampa” (Sabattini and Contigiani 1982; Sabattini et al. 1977). This region comprises the east-central provinces of Argentina, where agricultural and cattle raising activities are located. Originally this territory was a treeless prairie, but at present most of it is subdivided by crop fields and livestock pastures, separated by “border” habitats such as roads, streams, fence lines, and railroad rights-of-way (Ellis et al. 1997). Several authors proposed that these changes reduced the number of medium-size predators and favored rodent populations, particularly species such as *C. musculinus*, capable of taking advantage of unstable habitats (Bilenca 2000; Crespo 1966; Crespo et al. 1970). This species occupies “border” habitats throughout the year. From these habitats, it colonizes the fields when crops offer a good cover and become established there until harvest and plowing destroy the habitat, causing a high mortality and the dispersion of individuals back to the border habitats (de Villafañe et al. 1988; Ellis et al. 1997).

A remarkable characteristic of AHF has been the progressive extension of its endemic area, which was originally centered around the city of Junin (250 km west from Buenos Aires) and now comprises nearly 150,000 km² in four Argentinean provinces (Enríe and Feuillade 1998; Maiztegui et al. 1986). Studies on reservoirs are an essential part of any approach to understanding endemic disease. Knowledge of the genetic structure of reservoir populations may give insights into the factors responsible for the maintenance and spread of the virus in natural populations, particularly when the current endemic area includes only a small part of the geographic distribution of the reservoir, as is the case with AHF and *C. musculinus* (Mills and Childs 1998).

In a previous study using allozymic markers, Chiappero et al. (2002) analyzed 11 populations of *C. musculinus* from the humid pampa, 9 of them within the endemic zone of AHF and 2 outside of it. In agreement with previous studies in populations from another phytogeographic region (Gardenal and Blanco 1985; Gardenal et al. 1990), remarkably high levels of allozyme polymorphism have been found. Low (but statistically significant) genetic differentiation among populations, not correlated with geographic distance, was also observed. The pattern obtained suggested a more influential...
role of genetic drift relative to gene flow in determining the genetic structure of *C. musculinus* populations. In structural genes such as those coding for allozymes, mutations that create nonfunctional enzymes are eliminated by purifying selection. Therefore functional allelic variants accumulate slowly and, as a consequence, recent events in the history of populations are difficult to distinguish from older ones. Homogenizing selection acting on these loci may also maintain a genetic similarity not related to the levels of gene flow among populations (Bossart and Ashley Prowell 1998). Random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) markers, in contrast, usually reveal polymorphisms in noncoding regions of the genome and therefore are probably not subject to natural selection. Consequently they have a higher mutation rate and variation accumulates faster than in structural genes (Haymer 1994). In this study we used RAPD markers to evaluate the relative contribution of gene flow as a determinant of the genetic structure of *C. musculinus* populations from central Argentina.

### Materials and Methods

Ten localities in central Argentina were sampled (Figure 1): San Nicolás (*n* = 12), San Pedro (*n* = 15), Zárate (*n* = 18), Uranga (*n* = 11), and Melo (*n* = 15) are located inside the endemic zone of AHF; the remaining populations, Maciel (*n* = 8), Oliveros (*n* = 6), Laguna Larga (*n* = 17), Molinari (*n* = 18), and Donovan (*n* = 16) lie outside the endemic area. Individuals from another three localities inside the endemic zone (Alcorta, *n* = 18; J. B. Molina, *n* = 13; and Pergamino, *n* = 14), analyzed in Chiappero and Gardenal (2001), were also included. Animals were captured with Sherman live traps, sacrificed by inhalation of methoxyfluorane, and a small piece of kidney or liver was preserved in 90% ethanol.

Optimization of the RAPD technique, selection of primers, and testing for RAPD marker heritability were reported by Chiappero and Gardenal (2001). DNA extraction and PCR procedures were performed as described in that study. RAPD profiles were obtained using four primers (A01, A02, A06, and B01; Biodynamics, Buenos Aires, Argentina). A negative control containing all reagents except genomic DNA was included in each reaction. PCR products were separated on 0.6% agarose gels with 1% Synergel (Diversified Biotech, Newton Centre, MA) stained with ethidium bromide.

DNA from approximately one-third of the individuals from each population was amplified twice on different days to test for reproducibility. One or two individuals previously run on another gel were included in each new gel to facilitate the identification of bands across gels. Bands analyzed included those used in Chiappero and Gardenal (2001) and novel ones detected for the first time in this study. The presence or absence of each band was scored and converted into allele frequencies (populations are in Hardy-Weinberg equilibrium, as previously measured using allozymes), assuming that (1) two alleles segregate in each locus, one dominant that amplifies and one recessive that does not; (2) comigrating bands are the products of the same loci; and (3) recessive alleles are identical in state. Since null alleles can arise from more than one mutation at any of the two priming sites of a RAPD locus, the last assumption is probably not met. But, as de Wolf et al. (1998) note, by treating each RAPD locus as diallelic, the recessive phenotype could be considered a pool of genotypes and therefore statistical tests will tend to be conservative. Allele frequencies were estimated using equation [2a] in Lynch and Milligan (1994), correcting for the downward bias in the estimation of the null allele frequency. These authors pointed out that equation [2a] gives a better estimate of *q* than the square root of the null homozygotes whenever *Nq*² > 3, where *N* is the number of analyzed individuals. Therefore they suggest using bands whose observed frequency is less than 1 – (3/*N*) to calculate unbiased population genetic parameters.

Allele frequencies and mean expected heterozygosity (*H*₂) were calculated using the Tools for Population Genetic Analysis program (Miller 1998). Levels of variability were also assessed by the nucleotide diversity index *π*, which measures the average number of nucleotide differences per site between two randomly chosen DNA sequences in a population, using the method of Borowsky (2001).

Genetic subdivision of populations was estimated using three different methods. Weir and Cockerham’s (1984) theta (*θ*), estimator of *F*₂, which corrects for unequal sample sizes and finite number of populations sampled, and Lynch and Milligan’s (1994) *F*₂, which takes into account the dominant nature of RAPDs. The significance of *F*₂ was obtained using the formula *χ*² = 2*N* *F*₂, with degrees of freedom equal to the number of subpopulations minus one. Calculations were performed using RAPDST (Black 1997). The presence/absence of fragments was scored, disregarding any assumptions about its genetic control, using the AMOVA approach of Excoffier et al. (1992). The method computes the percentage of total genetic variation attributable to among-population variation and to within-population variation, and *θ*₂, an analogue of *F*₂. This analysis was developed for restriction fragment length
polymorphism (RFLP) haplotypes, but it can be adapted to RAPD markers by considering the presence/absence of bands of each individual as its phenotype, and thus the variance components computed correspond to phenotypic rather than genotypic variation. For this analysis, a matrix of interindividual distances was constructed as \( D = 1 - S \), where \( S \) is Nei and Li’s (1979) similarity index, which was used to perform the AMOVA analysis with the WINAMOVA 1.55 program (Excoffier 1993). Significance testing was performed through 1000 permutations under the hypothesis of absence of population subdivision.

The existence of an isolation by distance pattern was tested following Rousset (1997). To visualize genetic relationships among populations, a principal components analysis (PCA) was performed using the Infostat software (Di Rienzo et al. 2003).

All population genetics parameters were calculated taking all fragments into account, and including only those with a frequency lower than \( 1 - (3/N) \). When a fragment did not meet this criterion in one population, it was excluded from the entire study.

**Results**

One hundred eighteen distinctive RAPD bands ranging from 0.31 to 2.5 kb, corresponding to an equal number of presumptive loci, were considered in the present study. Only five loci (A1-1.22, A2-1.46, A6-0.74, A6-0.70, and B1-0.85) were monomorphic in all the populations. Allele frequencies for polymorphic loci are available via e-mail upon request. Seventy-eight polymorphic loci met the exclusion criterion simultaneously in all populations.

There was considerable variation within each population since every individual was characterized by a unique RAPD phenotype. Nucleotide diversity and mean expected heterozygosity for each population are shown in Table 1. Values of the latter parameter were consistently higher when all RAPD bands were considered. \( H_e \) values obtained for the same populations using allozymes are also presented for comparative purposes.

All computed measures of population structure showed an important overall degree of subdivision in *C. musculinus* populations (Table 2). In each case the estimates made including fragments at high frequencies were higher than those made with the \( 1 - (3/N) \) exclusion criterion. Lynch and Milligan’s \( F_{ST} \) showed the largest difference between both estimates, while \( \phi_{ST} \) gave similar values regardless of the fragments considered. AMOVA analysis showed that 87.8% of variability is attributable to differences between individuals within populations, while 12.2% was explained by differences between populations when no exclusion criterion is applied. When high-frequency bands are excluded, similar values were obtained (89.1%–10.9%). Pairwise estimates of genetic differentiation showed a similar pattern to the overall estimates: values of \( \theta \) and \( F_{ST} \) without pruning of fragments were higher than those obtained with pruning; conversely, values of \( \phi_{ST} \) with and without pruning were very similar. Isolation by distance was tested using \( \Phi_{ST} \) after pruning the high-frequency bands. Figure 2 shows the correlation between \( \phi_{ST}/(1 - \phi_{ST}) \) and the lognormal of geographic distance. Populations studied are not at equilibrium between gene flow and genetic drift.

Principal components analysis was performed on the reduced dataset (after Lynch and Milligan’s exclusion criterion was applied) and its graphical representation is

---

**Table 1.** Levels of genetic variability detected using RAPD markers (measured as mean expected heterozygosity, \( H_e \), and nucleotide diversity, \( \pi \)) in 13 populations of *C. musculinus*. \( H_e \) and \( \pi \) were calculated considering all loci (without exclusion) and excluding those bands with a frequency greater than \( 1 - (3/N) \) in any population (with exclusion).

<table>
<thead>
<tr>
<th></th>
<th>RAPDs</th>
<th>With exclusion</th>
<th>Without exclusion</th>
<th>Allezymes*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \pi )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pergamino</td>
<td>8.2 \times 10^{-3}</td>
<td>7.8 \times 10^{-3}</td>
<td>0.190</td>
<td>0.206</td>
</tr>
<tr>
<td>San Nicolás</td>
<td>8.0 \times 10^{-3}</td>
<td>8.1 \times 10^{-3}</td>
<td>0.174</td>
<td>0.221</td>
</tr>
<tr>
<td>San Pedro</td>
<td>9.4 \times 10^{-3}</td>
<td>8.5 \times 10^{-3}</td>
<td>0.208</td>
<td>0.230</td>
</tr>
<tr>
<td>Zárate</td>
<td>7.6 \times 10^{-3}</td>
<td>8.1 \times 10^{-3}</td>
<td>0.179</td>
<td>0.231</td>
</tr>
<tr>
<td>Alcorta</td>
<td>9.0 \times 10^{-3}</td>
<td>9.1 \times 10^{-3}</td>
<td>0.175</td>
<td>0.220</td>
</tr>
<tr>
<td>J. B. Molina</td>
<td>8.9 \times 10^{-3}</td>
<td>8.6 \times 10^{-3}</td>
<td>0.172</td>
<td>0.211</td>
</tr>
<tr>
<td>Uranga</td>
<td>8.6 \times 10^{-3}</td>
<td>8.3 \times 10^{-3}</td>
<td>0.201</td>
<td>0.220</td>
</tr>
<tr>
<td>Maciel</td>
<td>7.7 \times 10^{-3}</td>
<td>6.8 \times 10^{-3}</td>
<td>0.158</td>
<td>0.102</td>
</tr>
<tr>
<td>Oliveros</td>
<td>10 \times 10^{-3}</td>
<td>9.7 \times 10^{-3}</td>
<td>0.183</td>
<td>0.216</td>
</tr>
<tr>
<td>Melo</td>
<td>6.6 \times 10^{-3}</td>
<td>7.2 \times 10^{-3}</td>
<td>0.150</td>
<td>0.189</td>
</tr>
<tr>
<td>Laguna Larga</td>
<td>9.0 \times 10^{-3}</td>
<td>8.8 \times 10^{-3}</td>
<td>0.193</td>
<td>0.236</td>
</tr>
<tr>
<td>Molinari</td>
<td>9.2 \times 10^{-3}</td>
<td>8.9 \times 10^{-3}</td>
<td>0.185</td>
<td>0.234</td>
</tr>
<tr>
<td>Donovan</td>
<td>8.4 \times 10^{-3}</td>
<td>8.2 \times 10^{-3}</td>
<td>0.168</td>
<td>0.206</td>
</tr>
</tbody>
</table>

---

* Data for allozyme genetic variability in Laguna Larga population are from Gardenal et al. (1990) while those for the remaining populations are from Chiappero et al. (2002).

* Data from Chiappero and Gardenal (2001).
shown in Figure 3. The first three principal components explained 22%, 17%, and 11% of the variation, respectively.

**Discussion**

Previous studies demonstrated that *C. musculinus* populations harbor remarkably high levels of genetic variability at structural allozymic loci. Two alternative explanations for that result were proposed. Important levels of polymorphism could be maintained in the species by selective mechanisms in the highly temporally and spatially variable agroecosystem habitats or they could be the consequence of large effective population sizes determined, at least in part, by high levels of genetic exchange between neighboring populations (Chiappero et al. 2002; Gardenal and Blanco 1985; Gardenal et al. 1990).

Mean heterozygosity values in *C. musculinus* estimated using RAPD markers were higher than those obtained in the same populations using allozymes. As Liu and Fournier (1993) pointed out, the dominant nature of RAPDs will produce opposed biases in the estimation of $H_e$. If there is more than one mutation that creates a recessive allele or it has a low frequency (so that the majority of its copies are carried by heterozygotes), $H_e$ will be underestimated. On the other hand, the only monomorphic loci that are recorded are those monomorphic for the dominant allele, resulting in an overestimation of $H_e$. To address this issue, Szmidt et al. (1996) and Isabel et al. (1995) compared heterozygosity levels in conifers calculated from RAPD genotypic information obtained from haploid macrogametophyte tissues and from RAPD phenotypes. Both articles report that values of heterozygosity calculated from phenotypes (including all loci) were underestimated with respect to the values obtained from genotypes. In both studies, the pruning of fragments with frequencies greater than $1 - 3/N$ led to a higher value of $H_e$, due to the exclusion of loci monomorphic for the dominant allele. However, $H_e$ was still biased, now upwards, with respect to the values inferred from haploid tissue. In *C. musculinus*, levels of heterozygosity are also higher when pruning of fragments was applied (Table 1). Following Szmidt et al. (1996) and Isabel et al. (1995), the most appropriate estimation of heterozygosity in *C. musculinus* should stand between the two values presented in Table 1.

Since RAPD markers are amplified mainly from non-coding regions of the genome, any selective mechanisms can be ruled out to explain the maintenance of most of the variability detected. According to the neutralist theory, this process depends on the relationship $4 Ne\mu_0/(4 Ne\mu_0 + 1)$, where $N_e$ is the effective size of the population and $\mu_0$ is the neutral mutation rate (Kimura 1983). Thus high levels of variability require large effective population sizes to counteract genetic drift. Chiappero et al. (2002) found that although *C. musculinus* populations suffer drastic decreases in density during winter and early spring, they retain an important number of low-frequency (between 1% and 5%) allozyme alleles. The authors argued that either these effective population sizes are not reduced enough to lose these alleles or, during high population density periods, they are recovered by migration from nearby populations. Results presented here would support the argument that large effective sizes (at least during favorable seasons) could be of key importance for variability maintenance in each population.

The data presented here show a moderate to high degree of subdivision in *C. musculinus* populations (Wright 1978), which is by far larger than that found using structural genes ($\theta = 0.02$; Chiappero et al. 2002). The high degree of differentiation observed with RAPDs is due to frequency differences, because only four bands unique to one population were detected. An important part of the variability found in the populations analyzed would be ancestral, given that it is common to all populations. Results obtained show that genetic differentiation is not correlated with geographic distance, which confirms that these populations in central Argentina are not at equilibrium between gene flow and genetic drift.

Available paleontological data support a “recent colonization” hypothesis. Several authors report that the pampean fauna from the late Pleistocene and much of the Holocene has no modern analogues, suggesting that extant mammal communities have emerged in the last few thousand years (Tonni and Cione 1997; Tonni et al. 1999). Several authors proposed that before the settlement of agriculture in the

### Table 2. Estimates of population subdivision in *C. musculinus*

<table>
<thead>
<tr>
<th>Bands with frequency &gt; $1 - (3/N)$</th>
<th>Included</th>
<th>Excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theta (Weir and Cockerham 1984)</td>
<td>0.129***</td>
<td>0.078**</td>
</tr>
<tr>
<td>$F_{ST}$ (Lynch and Milligan 1994)</td>
<td>0.168***</td>
<td>0.092***</td>
</tr>
<tr>
<td>$\varphi_{ST}$ (Excoffier et al. 1992)</td>
<td>0.122***</td>
<td>0.109***</td>
</tr>
</tbody>
</table>

** $P < .01$; *** $P < .001$.
humid pampa, opportunistic species like *C. musculinus* were less abundant than stronger competitors such as *Akodon* and *Oligoryzomys*. The transformation of large surfaces from stable natural pastures to unstable crop fields created favorable conditions for *C. musculinus*, allowing populations to increase in size and to expand (Bilenca 2000; Crespo 1966). Pardiñas et al. (2000) analyzed rodent species in a 1830-year-old fossil owl pellets sample from the lower Chubut River (Patagonia) and compared it with the composition of modern pellets collected at the same locality. They found that in the upper Holocene, the modern species of micromammals were present in the lower valley of the Chubut River and species of the *Calomys* genus represented only 5% of the total. In contrast, *C. musculinus* constituted 95% of a modern pellet sample. The authors proposed that the change was due to the establishment of agroecosystems along the river.

After range expansion, the newly formed populations have a homogeneous genetic constitution. But if genetic exchange thereafter is low, genetic drift will be more influential than gene flow, and every population will differentiate at random with respect to all others, producing in influential than gene flow, and every population will exchange thereafter is low, genetic drift will be more along the river.

![Figure 3](http://example.com/figure3.png)

**Figure 3.** Populations of *C. musculinus* arranged along the first three axes of a PCA based on RAPD bands with a frequency less than 1 – 3/N.

> Results of epidemiological studies on AHF are consistent with current low levels of genetic exchange among *C. musculinus* populations. Mills et al. (1991) found that the prevalence of infection in *C. musculinus* populations can be high in a given locality and very low or zero in adjacent populations (the separation among populations in that study ranged from 10 to 160 km). García et al. (2000) analyzed the genetic relationships among wild-type strains of Junin virus involved in epidemics across a span of 35 years from the center and edges of the endemic area. Their study showed no correlation between genetic divergence and the geographic origin of the samples.

The availability of favorable habitat for *C. musculinus* in central Argentina, its apparent continuity, and its annual colonization and evacuation of crop fields are situations that can lead to increased opportunities for gene flow among populations. However, the genetic differentiation observed clearly indicates that gene flow is restricted, at least at the geographic scale of the present study.

**Acknowledgments**

The authors thank Drs. A. Blanco and M. S. Sabattini for their comments and suggestions. The authors are also grateful to Dr. E. Caviedes Vidal (National University of San Luis, Argentina), who kindly provided the specimens from Donovan, to Dr. G. E. Calderón and her field assistants of the National Institute of Human Viral Diseases (Pergamino, Argentina), who captured the samples from endemic areas, and to J. Arroyo and Dr. G. Theiler for their collaboration in obtaining samples from Molinari. This work was supported, in part, by grants from the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), the Agencia Córdoba Ciencia, the ‘Ramon Carrillo-Arturo Ohtaita’ Fellowship from the Subsecretaría de Investigación y Tecnología (Ministerio de Salud), all of Argentina. C.N.G. is career investigator of the CONICET and M.B.C. is a fellow of the Antorchas Foundation.

**References**


Received August 23, 2002 
Accepted July 31, 2003

Corresponding Editor: William S. Modi

Chiappero and Gardenal • Gene Flow in Calomys musculinus

495