Interactions between environmental factors can hide isolation by distance patterns: a case study of Ctenomys rionegrensis in Uruguay

Marcelo J. Kittlein$^1$ and Oscar E. Gaggiotti$^{2,*}$

$^1$Departamento de Biología, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, Funes 3250 4to piso, (7600) Mar del Plata, Argentina

$^2$Laboratoire d’Ecologie Alpine, UMR CNRS 5553, Université Joseph Fourier, BP 53, 38 041 Grenoble Cedex 09, France

Identifying the factors responsible for the structuring of genetic diversity is of fundamental importance for biodiversity conservation. However, arriving at such understanding is difficult owing to the many factors involved and the potential interactions between them. Here, we present an example of how such interactions can preclude us from arriving at a complete characterization of the demographic history and genetic structure of a species. Ctenomys rionegrensis is a species with restricted dispersal abilities and, as such, should exhibit an isolation by distance (IBD) pattern, which previous studies were unable to uncover. It was therefore concluded that this species underwent a recent population expansion. Using a novel hierarchical Bayesian method, we show that the inability to detect the IBD pattern is due to the interaction between elevation and geographical distance. We posit that populations in low areas suffer periodic floods that may reduce local population sizes, increasing genetic drift, a process that masks the effect of distance on genetic differentiation. Our results do not refute the possibility that the populations of C. rionegrensis underwent a recent population expansion but they indicate that an alternative scenario described by a metapopulation model at or near migration-drift equilibrium cannot be excluded either.

Keywords: genetic structure; isolation by distance; population expansion; Bayesian methods

1. INTRODUCTION

For many decades, population geneticists have been interested in quantifying the relative contribution of different evolutionary forces in shaping the genetic variation observed among living organisms. Genetic variation is structured in space, being influenced by evolutionary factors such as genetic drift, selection, migration and mutation, which in turn must operate within the historical and biological context of each species. Understanding the processes responsible for the structuring of genetic variation is of fundamental importance in the fields of biodiversity conservation, improvement of agricultural crops and breeds, and identification of disease genes in humans and economically important species. Arriving at such understanding, however, is difficult owing to the many interactions that may exist between demographic and evolutionary processes and between the different environmental factors influencing genetic structuring. In particular, it is frequently difficult to distinguish scenarios of recent population expansion from those of near equilibrium between migration and genetic drift. It is generally assumed that these two scenarios have contrasting effects on the pattern of spatial apportionment of genetic variation (Matocq et al. 2000). For example, in the case of species with restricted dispersal abilities that are at migration-drift equilibrium, we expect to observe a correlation between the genetic distance between pairs of populations and the geographical distance that separates them, this is the so-called isolation by distance (IBD) pattern. Its absence can be interpreted as evidence for a non-equilibrium situation where the species has recently colonized its current distribution (Slatkin 1993). This conclusion is, however, valid only for distant locations. Indeed, when the range expansion is rapid with a single ancestral population giving rise to all extant demes at the same time $t$ in the past, it is still possible to detect IBD among nearby populations (cf. fig. 2, Slatkin 1993). Nevertheless, when range expansion is more gradual, the IBD pattern among nearby demes may become too weak to be detected. This would be the case if, for example, there is a stepwise colonization process where at time $t$ in the past a single population gives rise to one neighbouring deme and all subsequent colonizations originate from the most recently colonized deme only (Good’s model in Slatkin 1993). In this latter case, differentiation between the ancestral deme and the descendant demes will increase as the age of the descendant demes decreases (cf. fig. 8, Slatkin 1993). Thus, it may be possible to identify a gradual range expansion by looking at the effect that the distance from the ancestral population (a surrogate for the age of the descendant population) has on genetic differentiation. If it is unclear which of all the extant demes is the ancestral population, then it is possible to use latitude and longitude under the assumption that colonization progressed steadily from one extreme of the range to the other (e.g. from the southeast to the northwest), as Foll & Gaggiotti (2006) have done in the case of humans.

Another potential explanation for an apparent absence of IBD pattern that has not yet been evaluated is that an environmental factor that has not been taken into account. Identifying the factor responsible for the structuring of genetic diversity is of fundamental importance for...
consideration precludes the detection of the effect of geographical distance. For example, environmental perturbations that lead to periodic decreases in population size (e.g. floods, droughts, epidemics) can greatly increase the effect of genetic drift in a subset of the subpopulations. Thus, two nearby populations may become genetically differentiated despite being geographically close. Standard population genetics models do not allow us to uncover such effects, but recent developments in the field of statistical genetics have made this possible. Here, we present an example of a situation where the effect of an important environmental factor (elevation) hides the effect of distance on genetic differentiation and, therefore, precludes us from accurately characterizing the demographic history of a population when using standard methods.

We use as case study a species of South American rodents of the genus *Ctenomys*, commonly known as tuco-tucos, which is the most speciose among all subterranean rodent genera (Reig et al. 1990; Cook & Lessa 1998; Lessa & Cook 1998; Castillo et al. 2005). Populations of *Ctenomys* are composed of semi-isolated demes, occupying patches of habitat where soil hardness and particle size provide suitable conditions for burrowing activities (Busch et al. 2000). Species-specific habitat requirements, such as those related to body size differences (see Lessa & Thaeler 1989) and other factors intrinsically associated with the species' biology (i.e. pelage crypticity, Krupa & Geluso 1989) and other factors intrinsically associated with the species' biology (i.e. pelage crypticity, Krupa & Geluso 1989), strongly influence the type of soil subterranean rodents use to build their burrows, affecting the connectivity among local populations and resulting in the genetic structuring of populations. *Ctenomys* shares with other subterranean rodents restricted dispersal capabilities, small local effective population numbers, high territoriality and, in some cases, socially structured mating systems (Reig et al. 1990; Lacey 2000; Massarini & de Freitas 2005).

Several lines of evidence suggest that the Rio Negro tuco-tuco, *Ctenomys rionegrensis*, has undergone a recent population expansion into its current Uruguayan distribution from a more restricted range (D’Elia et al. 1998; Wlasiuk et al. 2003). Wlasiuk et al. (2003) found no correlations between genetic distances based on allozymes, cytochrome *b* sequences or microsatellites with each other, as well as between any of these and geographical distances. Morphological differentiation appears to show a similar pattern (D’Anatro & Lessa 2006). Additionally, Wlasiuk et al. (2003) carried out several analyses that specifically test for signatures of population expansion using mtDNA and, although non-significant, they were consistent in suggesting such a scenario. Moreover, Fu’s test uncovered a significant excess of low-frequency haplotypes, which could be explained by either a departure from the neutral mutation model or a recent expansion. Pleistocene and Holocene marine transgressions may have played an important role in such demographic changes since it is known that these perturbations reached the area of the present distribution of *C. rionegrensis* in Uruguay (Alonso 1978; Sprechmann 1978). Thus, the prevailing view is that populations of *C. rionegrensis* have differentiated relatively recently and, therefore, there has been insufficient time for a pattern of IBD to establish. An alternative view that has not been evaluated is that the population may be at or near equilibrium and that the signature of differentiation can be associated to distance and other so far undiscovered factors; their interaction could then preclude us from detecting the pattern of IBD.

Recent advances in statistical genetics have led to the development of Bayesian approaches that allow us to investigate potential interactions between environmental factors involved in the genetic structuring of populations. Here, we use a recently developed method (Foll & Gaggiotti 2006) to try to disentangle the effects of demographic history and those of various environmental factors. Using this approach, we expand previous analyses of *C. rionegrensis* populations and uncover an IBD pattern that could be suggestive of a near equilibrium situation. Additionally, we show that genetic differentiation in this population is multifactorially determined.

2. MATERIAL AND METHODS

The genetic data used in our analyses consist of the 11 microsatellite loci published by Wlasiuk et al. (2003) for *C. rionegrensis* (see their appendix). These data were collected from eight sampling locations that cover the geographical range of this species in Uruguay (figure 1). In terms of environmental data, we used the area of each population’s patch, estimated from a LANDSAT ETM+ image of the region, distance between populations, and elevation above sea level obtained from a digital elevation model of the region (table 1). This latter factor was considered as a proxy for the effect of floods, which can lead to local population bottlenecks and increased genetic drift. We also considered latitude (noted y) and longitude (noted x) expressed using the Universal Transversed Mercator planar transformation in order to correct for the curvature of the Earth. The effect of distance was modelled using the average distance between each focal population and all other populations, which is a measure akin to connectivity.

We used the hierarchical Bayesian method of Foll & Gaggiotti (2006), as implemented in the programme *Geste*, to test for the effect that different environmental factors may have on the genetic structure of *C. rionegrensis*. This method estimates the genetic differentiation between each local
population and the overall metapopulation (as measured by a local population $F_{ST}$ value) using the approach first put forward by Balding & Nichols (1995) and relates it to environmental factors using a generalized linear model. For the moment, the software that implements this method allows for a maximum of two factors, thus we carried out multiple analyses, each considering two factors at a time. The consideration of two factors and their interaction leads to nine alternative models, and the method provides posterior probabilities for each one of them using a reversible jump Markov chain Monte Carlo (MCMC) approach. The model with the highest posterior probability is the one that best explains the data. In all cases, we used 10 pilot runs of 2000 iterations to obtain the parameters of the proposal distributions used by the MCMC. We also used an additional burn-in of $5 \times 10^4$ iterations and a thinning interval of 50. All estimates were obtained using a sample size of 20 000.

### Table 1. Environmental variables and posterior mode of population specific $F_{ST}$s for the eight populations of *C. rionegrensis* in western Uruguay.

<table>
<thead>
<tr>
<th>population</th>
<th>area (km$^2$)</th>
<th>distance (km)</th>
<th>elevation (m)</th>
<th>$F_{ST}$ (HPDI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Chapareí</td>
<td>0.13</td>
<td>24.06</td>
<td>14</td>
<td>0.258 (0.185; 0.369)</td>
</tr>
<tr>
<td>2. Abrojal</td>
<td>0.40</td>
<td>27.26</td>
<td>14</td>
<td>0.412 (0.301; 0.538)</td>
</tr>
<tr>
<td>3. La Guardia</td>
<td>0.08</td>
<td>38.47</td>
<td>26</td>
<td>0.687 (0.538; 0.813)</td>
</tr>
<tr>
<td>4. La Tabaré</td>
<td>0.07</td>
<td>44.62</td>
<td>11</td>
<td>0.479 (0.374; 0.595)</td>
</tr>
<tr>
<td>5. Las Cañas</td>
<td>0.09</td>
<td>39.09</td>
<td>6</td>
<td>0.601 (0.474; 0.743)</td>
</tr>
<tr>
<td>6. Arrayanes</td>
<td>1.58</td>
<td>28.76</td>
<td>19</td>
<td>0.185 (0.127; 0.272)</td>
</tr>
<tr>
<td>7. Mafalda</td>
<td>0.03</td>
<td>34.31</td>
<td>65</td>
<td>0.281 (0.199; 0.389)</td>
</tr>
<tr>
<td>8. Nuevo Berlín</td>
<td>0.10</td>
<td>26.99</td>
<td>5</td>
<td>0.984 (0.892; 1.000)</td>
</tr>
</tbody>
</table>

### Table 2. Posterior probabilities associated with the nine alternative models arising in the analysis that considers elevation above sea level and distance. ($\sigma^2$ is the posterior mode of unexplained variance associated to each model. Values in parenthesis correspond to the 95 per cent highest posterior probability interval. The model with the highest probability and the lowest variance is highlighted in italics).

<table>
<thead>
<tr>
<th>model</th>
<th>probability</th>
<th>$\sigma^2$ (HPDI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. constant</td>
<td>0.048</td>
<td>1.62 (0.40; 6.36)</td>
</tr>
<tr>
<td>2. elevation</td>
<td>0.161</td>
<td>1.22 (0.34; 5.51)</td>
</tr>
<tr>
<td>3. constant + elevation</td>
<td>0.036</td>
<td>1.50 (0.53; 6.89)</td>
</tr>
<tr>
<td>4. distance</td>
<td>0.040</td>
<td>1.51 (0.51; 6.17)</td>
</tr>
<tr>
<td>5. constant + distance</td>
<td>0.008</td>
<td>1.61 (0.66; 7.31)</td>
</tr>
<tr>
<td>6. elevation + distance</td>
<td>0.030</td>
<td>1.20 (0.43; 6.57)</td>
</tr>
<tr>
<td>7. constant + elevation + distance</td>
<td>0.006</td>
<td>1.51 (0.50; 7.15)</td>
</tr>
<tr>
<td>8. elevation + distance + elevation x distance</td>
<td>0.586</td>
<td>0.74 (0.20; 3.60)</td>
</tr>
<tr>
<td>9. constant + elevation + distance + elevation x distance</td>
<td>0.086</td>
<td>0.87 (0.17; 4.39)</td>
</tr>
</tbody>
</table>

### 3. RESULTS

We first investigated the hypothesis of a gradual population expansion. If this hypothesis is correct, we can assume a fission model in which successive founder events would lead to a gradual increase in genetic differentiation between local and ancestral populations as the distance between them (a proxy for deme age; see §1) increases. Such an effect can be uncovered by a GESTE analysis using latitude and longitude as factors (Foll & Gaggiotti 2006). Results show that posterior probabilities are fairly similar for models 1 (constant factor only) and 2 (longitude only) and somewhat higher for model 4 (latitude only). The overall posterior probability for models that include an effect of longitude only (2 and 3) is 0.27. The corresponding value for models that include latitude only (4 and 5) is 0.34. Finally, the posterior probability for models that include both factors (6–9) is 0.15, while that of the null model is 0.24. The effect of longitude is very weak (modal value for $\sigma^2 = -0.008$, highest posterior density interval (HPDI) $=(-1.17; 1.14)$) while that of latitude is somewhat stronger (modal value for $\sigma^2 = 0.24$, HPDI $=(-0.79; 1.52)$).

These results provide evidence for a gradual range expansion. The genetic signal is somewhat weak, which suggest that enough time may have elapsed since colonization to approach migration-drift equilibrium. If this is the case, *C. rionegrensis* may be described by a metapopulation model that is at or near migration-drift equilibrium. The absence of IBD pattern could be the result of an interaction between distance and some other undisclosed factor. Thus, we investigated the effect of habitat patch area and elevation and their interaction with the average distance among populations. We carried out analyses that considered all possible pairwise combinations of these three factors. Only the analysis that considered elevation and distance provided a strong and unambiguous result showing an effect on the genetic structuring (table 2). More specifically, the overall posterior probability for the models that considered both factors and their interaction (8 and 9) was 0.67; all other models have a much lower posterior probability. Models that include the effect of distance only have a posterior probability of 0.05, while those that include elevation only add up to 0.20. Thus, neither factor by itself explains genetic structuring well; instead, they act concurrently and in particular through their interaction. The very strong effect of the interaction is illustrated by the dramatic
that includes elevation and distance by focusing on the low-altitude populations.

is present but is hidden by the strong genetic drift affecting distance only. These results indicate that the IBD pattern 0.32, but it is still lower than that of models that include elevation. Finally, the probability of the null model increased to 0.16 and that of models that include only distance increased to 0.41. Additionally, the posterior probability of models specifically, the posterior probability of models that include elevation decreased to 0.045 (HPDI = (0.6621; 5.994)). This interaction indicates that distance manifests its effect only among populations at higher elevation because strong genetic drift at low elevation hides the IBD pattern (figure 2).

To verify this assertion, we carried out an additional analysis excluding the two lowest populations (Las Cañas and Nuevo Berlin, table 1). The results differ dramatically from those of the analysis that included all demes. More specifically, the posterior probability of models that include the interaction (8 and 9) becomes negligible (the overall posterior probability for them decreased to 0.045) while that of models that include only distance increased to 0.41. Additionally, the posterior probability of models that include elevation decreased to 0.16 and that of models that include both factors increased to 0.07. Finally, the probability of the null model increased to 0.32, but it is still lower than that of models that include distance only. These results indicate that the IBD pattern is present but is hidden by the strong genetic drift affecting low-altitude populations.

We can compare the results of the analysis that considers latitude and longitude with those of the one that includes elevation and distance by focusing on the estimates of the variance that remain unexplained by the models with the highest posterior probability. In the case of the latitude/longitude analysis, the variance that remains unexplained by the best model is 1.52 (HPDI = (0.52; 6.32)). On the other hand, the respective estimate for the elevation/distance analysis that includes all populations is 0.74 (HPDI = (0.2; 3.60)). The much lower variance of the latter suggests that the near equilibrium scenario is a better description of the demographic history of C. rionegrensis.

4. DISCUSSION

The results presented here illustrate very clearly the importance of considering the effect of interactions between environmental factors in studies of population genetic structure. In the specific case of the example we consider, the interaction between geographical distance and elevation above sea level hides the pattern of IBD that is expected in species with limited dispersal capabilities. Populations of C. rionegrensis in western Uruguay are located near rivers and, therefore, populations in low areas are potentially subjected to periodic floods that may reduce local population sizes and increase the effect of genetic drift; a process that masks the effect of distance on genetic differentiation. Figure 2 shows that the effect of distance does increase with elevation, a pattern that is captured by the interaction term of the generalized linear model implemented in the hierarchical Bayesian approach used for our analyses.

It is worth emphasizing that it is the consideration of the interaction between environmental factors that allow us to identify the effect of geographical distance on genetic differentiation. In this regard, we note that we carried out a multivariate Mantel test to estimate the correlation between pairwise genetic and geographical distances with difference in elevation partialled out and the results were non-significant. This highlights the usefulness of new Bayesian methods that allow us to consider complex models and to investigate the effect of factors that are population specific (such as elevation) and cannot be measured in terms of pairwise differences between populations.

The structuring of neutral genetic variation in natural populations is influenced by its demographic history. Thus, the analysis of genetic data can in principle allow us to distinguish between alternative hypotheses concerning the processes that have led to current spatial patterns of genetic structuring. However, achieving this goal is not straightforward because the differences between the genetic signatures left by alternative demographic processes are sometimes subtle. As explained in §1, a scenario of migration-drift equilibrium in species with limited dispersal will lead to a clear IBD pattern (Slatkin 1993). However, an alternative non-equilibrium scenario can also lead to IBD but only among nearby populations. This small-scale IBD can be discernible in the case of species that underwent a rapid range expansion, but it would be very difficult to uncover if the range expansion was gradual. Nevertheless, in this latter case, we expect to observe a progressive genetic differentiation between the ancestral deme and the descendant demes as the distance between them (a surrogate for the age of the descendant population) increases.
Our analyses suggest that the observed genetic structure of *C. rionegrensis* could be described by a metapopulation model that is approaching migration-drift equilibrium after having undergone a gradual range expansion. However, the alternative scenario of a rapid recent expansion cannot be completely excluded because the geographical scale considered in this study (64 km between the two most distant populations) may be too small, in which case a rapid population expansion could also have led to the observed IBD pattern. In this regard, we note that at least one of the many population expansion tests carried out by Wlasiuk *et al.* (2003) is significant, and although the others had a poor fit, they all were consistent in suggesting population expansion. Nevertheless, whether the population expansion of *C. rionegrensis* in western Uruguay is recent or not, our analysis clearly demonstrates that this species exhibits an IBD pattern. Our estimates of population-specific *F*<sub>ST</sub> can be used to characterize the genetic distinctiveness of each local population. The larger the local *F*<sub>ST</sub> is, the more genetically different this population is from the metapopulation as a whole. Poor connectivity and/or small effective sizes lead to high *F*<sub>ST</sub>. The two populations located in low elevation areas (Nuevo Berlin and Las Cañas) have the largest and third largest *F*<sub>ST</sub>’s, most probably due to the fact that, as already mentioned, they may have small population sizes due to recurrent flooding. The distinctiveness of these two populations at the molecular level is also manifested at the morphological level since they are only composed of individuals with melanic pelage. This correspondence between high population-specific *F*<sub>ST</sub> and melanic colour fixation gives additional support to the hypothesis that drift alone is responsible for the fixation of this coat type.

The lack of detection of IBD pattern in species with low dispersal capabilities is not uncommon (e.g. Dixon *et al.* 2007; Florin & Hoglund 2007). As mentioned before, a commonly invoked explanation for this observation is that the species is not yet at the migration-drift equilibrium because it has recently undergone a spatial expansion. Other examples of this situation are studies on pocket gophers (Alvarez-Castañeda & Patton 2004), other tuco-tucos (Mora *et al.* 2006) and mustelids (Congdon *et al.* 2000). Another invoked explanation is that Euclidean distance does not take into account the effect of landscape structure. Thus, there has been a growing interest in using landscape ecology measures based on the least cost path (LCP) model (Coulon *et al.* 2004), to better explain observed patterns of genetic differentiation (Spear *et al.* 2005; Broquet *et al.* 2006; Vuilleumier & Fontanillas 2007). LCP is supposed to reflect relative movement difficulties through different habitat types and, therefore, requires the assumption of hypotheses on how the effect of different landscape attributes can limit or facilitate dispersal. Another problem of LCP measures is that they cannot incorporate the effect of independent parallel pathways connecting sampling locations (McRae 2006).

To overcome this problem, McRae (2006) has proposed a distance metric based on circuit theory. However, as is the case for LCP-based metrics, the resistance metric he proposed incorporates the effect of many environmental factors simultaneously and, therefore, does not allow us to evaluate the relative importance of each one of them separately. Thus, although LCP and resistance metric approaches can be useful for descriptive purposes, they do not allow us to test hypothesis about the factors responsible for the genetic structuring of natural populations.

Examples such as the one we present here illustrate the fact that although it is sometimes useful to present equilibrium and non-equilibrium scenarios as mutually exclusive, in reality most species represent intermediate situations along a continuum that extends between these two extreme scenarios. The use of novel hierarchical Bayesian methods in conjunction with more traditional approaches allows us to fully characterize the complex demographic and evolutionary histories of species such as the one we analyse here.

This research was supported by grant A05B01 from the programme ECOS-Sud (France/Secyt-Argentina). Additional funding was provided by grants from CONICET (PIP-5844) and FonCyT (PICT-38361) to M.J.K. and from Fond National de la Science to O.E.G. (ACI-ImpBio-2004–ADGP). M.J.K. is a researcher from CONICET, Argentina. Enrique Lessa and an anonymous reviewer provided very useful comments that greatly improved the original manuscript.

REFERENCES


Sprechmann, P. 1978 The paleoecology and paleogeography of the Uruguayan coastal area during the neogene and quaternary. Zitteliana 4, 3–72.
