SUPPLEMENTARY MATERIAL TO MOLECULAR ECOLOGY
REVIEW ARTICLE
Statistical methods in spatial genetics

Gilles Guillot*, Raphaël Leblois†, Aurélie Coulon†, Alain C. Frantz§

September 21, 2009

*Department of Informatics and Mathematical Modelling, Technical University of Denmark, Copenhagen, Denmark; gigu@imm.dtu.dk
†UMR7205 MNHN/CNRS, Département systématique et Evolution, Muséum National d’Histoire Naturelle, Paris, France
‡UMR 7179 MNHN/CNRS Département Ecologie et Gestion de la Biodiversité Muséum National d’Histoire Naturelle, Brunoy, France
§Department of Animal & Plant Sciences, University of Sheffield, United Kingdom
Figure I: Examples of simulated spatial patterns for population membership obtained by a constrained Voronoi tiling (induced by the sampling sites) and a Markov random field colouring. (a) locations of individuals here assumed to be regularly sampled over space; (b) Voronoi tessellation induced by the set of sampling sites; (c) Delaunay graph induced by the sampling sites: two sites are considered as neighbours if they belong to two Voronoi tiles sharing a common edge; (d) example of population memberships obtained by allocating colours with a Potts model with interaction parameter $\psi = 0.7$, neighbouring sites have a strong tendency to have similar colours; (e) same as (d) with $\psi = 0.6$, neighbouring sites have a mild tendency to have similar colours; (f) same as (d) with $\psi = 0.2$, the spatial pattern is very loose. The model is not defined in the continuum but only for sampling sites. A common rule of thumb to get nicer maps consists in extending the colour of each individual to the whole cell (not shown here).
Figure II: Same as Fig. I with irregularly spaced sampling sites.
Further examples of applications of clustering models

Genetic structure induced by habitat specialisation

Spatial genetic approaches have been used to show that ecological processes can have a strong impact on the amount of gene flow between populations. Indeed, some studies have found a good concordance between geographical boundaries of genetic clusters of a particular study organism and ecological factors such as climate and habitat types, as well as diet composition and distribution of prey. The majority of these studies have been performed on otherwise highly mobile carnivore species that typically disperse long distances and often exhibit low levels of genetic divergence between distantly separated sampling sites. For example, Sacks et al. (2004) and Sacks et al. (2008) have presented evidence of coyote (Canis latrans) genetic structure conforming to major habitat divisions. Sacks et al. (2008) analysed genetic profiles from more than 2000 coyotes from two adjacent ecoregions with differing levels of habitat heterogeneity. While the Desert-Prairie ecoregion (DPE) was relatively homogeneous and characterised by continuous tracts of gradually intergrading flora and fauna, the biologically diverse California Floristic Province (CFP) is a Mediterranean-type ecoregion known for its habitat heterogeneity. The authors plotted assignments obtained from Bayesian clustering methods on a map of the area to show that coyotes sampled from the CFP exhibited genetic cluster structure concordant with habitat sub-regions, while individuals sampled from dispersed sites in the DPE formed a large panmictic unit (Fig. III). The authors explained this result by natal habitat-specific dispersal, i.e. the tendency of coyotes to disperse preferentially into habitat similar to their natal one; a behaviour supported by a number of behavioural studies (Davis & Stamps 2004). Studying Salmon on the North Atlantic American coast, Dionne et al. (2008) identified several clusters characterised by different coastal distances and thermal regimes. Other applications to animal or plant aquatic species include (Alberto et al. 2008, Florin & Höglund 2008, Latch et al. 2008, E. Leclerc & Bernatchez 2008, Scott-McCairns & Bernatchez 2008) and (Pampoulie et al. 2009).

Genetic structure induced by habitat fragmentation

Habitat fragmentation is the progressive shrinkage and isolation of habitat patches (Andrén 1994). It can be due to the existence of narrow linear elements (e.g. roads, rivers) that divide otherwise continuous patches of habitat, to the presence of (non-linear) areas of non-suitable landscape elements, or both. The increase of isolation of habitat patches may have important consequences on population genetic structure. Indeed, depending on the abilities of organisms to move through the matrix separating habitat patches, gene flow directions and quantities can be deeply modified, which can in turn affect the genetic structure of the species. Coulon et al. (2006) tested the hypothesis that linear elements affect roe deer (Capreolus capreolus) effective dispersal and contribute to determining population genetic structure in that species. They sampled individuals in
Figure III: Modal assignments of individual coyotes to different clusters using the (a) STRUCTURE algorithm (K=6) and the (b) spatially explicit GENELAND (K=8) algorithm. In the STRUCTURE-based assignments, coloured circles and asterisks indicate individuals with $>80\%$ and $<80\%$, respectively, of their ancestry estimated to be from a single genetic cluster. The thin lines illustrate the boundaries between habitat sub-regions in the California Floristic Province ecoregion, which is itself separated from the Desert-Prairie ecoregion by the thick line. Blow ups illustrate sharp genetic divisions in coyote clusters corresponding to dramatic changes in the landscape with no significant physical barriers. Samples in the circle were from the short grass prairie from southeastern Colorado, a subregion of the Desert-Prairie ecoregion. (Caption and figures are reprinted from (Sacks et al. 2008).)
south-western France, in an area bisected by a linear zone containing a fenced highway, several
canals and the Garonne River, which were hypothesised as being potential barriers to roe deer
dispersal movements. Analyses using GENELAND inferred the existence of two genetic units. The
boundary between the two groups coincided with the area including the highway, the canals and
the river, suggesting that the combined presence of these linear elements acts as a moderator of
gene flow in roe deer. Riley et al. (2006) showed that linear elements can even affect the genetic
structure of very mobile species such as carnivores. They sampled bobcats (*Lynx rufus*) and coyotes
(*Canis latrans*) in California, in an area bisected by a congested and fenced 10-12 lane road built
in 1949. Bayesian clustering implemented in STRUCTURE revealed the existence of genetic clus-
ters whose limits coincided with the location of this freeway. Complementary analyses confirmed
that this road constrains gene flow in the two species: $F_{ST}$ values among clusters separated by
the highway were 2 to 9 times higher than those among clusters in the same side of the highway
and separated by similar or higher distances. The percentage of related individuals was also much
higher in clusters on the same side than in those on opposite sides of that road. Assignment tests
implemented with STRUCTURE suggested surprisingly high percentages of migrants across the
freeway (given the previous results): 3.4% per generation for bobcats, and 9.1% per generation
for coyotes. Based on simulations, the authors estimated that an effective migration rate of 0.5%
per generation or less was required to get the $F_{ST}$ values they observed. They interpreted this
discrepancy as an indication of a low reproductive success of migrants crossing the highway. See
also (Strasburg 2006) for further discussion. An example of a study on the effects of fragmentation
triggered by the presence of (non-linear) areas of non-suitable landscape elements can be found
in (Hanson et al. 2008). These authors studied the effects of such fragmentation on the genetic
structure of a tropical tree, *Dipteryx panamensis*. They compared the genetic structure of adult
trees (dating from before fragmentation occurred) and of their progeny in four types of habitats
representing a gradient of fragmentation level: continuous forest, forest fragments, pastures and
isolated pastures. Paired t-tests on $G_{ST}$ estimates among progeny and among adults revealed a sig-
nificant increase in differentiation among the progeny in isolated pastures, and no difference in the
three other habitat types. This test supported the author’s hypothesis that reduced out-crossing
and smaller breeding populations in isolated pastures would cause increased genetic structure in
progeny. Moreover, the comparison of spatial auto-correlation patterns in the four habitat types
showed that the magnitude and the scale of auto-correlation increase as fragmentation levels in-
crease. This result is indicative of longer pollen dispersal distances in more fragmented habitats, a
result also supported by paternity analyses.
Using spatial statistics methods to investigate disease spread

Understanding factors responsible for spatial heterogeneity in the distribution of wildlife diseases can be challenging. Spatial genetics potentially offers an efficient way of understanding where and why hosts are moving across the landscape, which may contribute to understanding present and future spread of disease, help to determine risks to domestic animals and humans, and designing optimal surveillance and control programs. Ultimately, integrating population genetic and environmental data with spatial disease patterns could help to discern the complex gene-environment interactions that result in spatial patterns of disease incidence (Sloan et al. 2009).

Some authors have tried to correlate landscape features with genetic discontinuities in an attempt to identify physical barriers to host movement and disease spread. Blanchong et al. (2008) found a negative correlation between the prevalence of chronic wasting disease in Wisconsin whitetailed deer (Odocoileus virginianus) and the genetic differentiation of their various study areas from the core area of disease origin. Genetic differentiation was greatest, and disease prevalence lowest, in study areas separated by a river from the area of disease origin. The authors took this as evidence that the river reduced deer flow and disease spread. Investigating spatial genetic structure in the raccoon (Procyon lotor), Cullingham et al. (2009) identified two geographically coherent genetic clusters each side of the Niagara River, but not the St Lawrence. This difference in permeability was consistent with the occurrence of rabies across the latter but not the former feature. However, the authors also reported independent evidence of raccoons moving across the putative Niagara barrier and they speculated that the control actions in place had been necessary to avoid the spread of the disease. Indeed, Rees et al. (2008) simulated the expansion of the raccoon population across the Niagara River using an individual-based, spatially explicit model that included information on individual mitochondrial haplotypes. The authors estimated genetic diversity measures assuming different proportions of individuals successfully crossing the river (i.e. different permeability values) and compared these to the ones obtained from the empirical data. They concluded that the Niagara River inhibited only one out of two raccoon crossings, illustrating how gene flow measures can be used to calibrate the effect of a linear feature as dispersal barrier. Spatial genetics approaches may also be used to reconstruct disease-spread and the underlying mechanism. Analysing phylogenetic structure and timing and location of consecutive outbreaks, Walsh et al. (2005) found evidence that the Zaire strain of the Ebola virus has recently spread across parts of central Africa in a wave-like fashion. The authors reconstructed the directional spread of the virus by identifying the putative path that maximised the correlation between spatial and genetic distances. Specifically, they found evidence for a change in direction of the spread, the largest river in the region having contained the spread for a few years. Using an isolation-by-distance-based, intra-population assignment method, Pope et al. (2007) confirmed not only that badger (Meles meles) dispersal increased after proactive culling in the context of bovine tuberculosis control, but showed that this was due to an increase
in medium- and long-distance dispersal. Also, individuals infected with the causative agent of the disease moved further than healthy individuals. These results helped to explain why the incidence of tuberculosis in cattle may increase after culling operations.
<table>
<thead>
<tr>
<th>Program</th>
<th>Spatial method(s)</th>
<th>Indiv. or pop.-level</th>
<th>Input</th>
<th>Output</th>
<th>OS</th>
<th>Web address</th>
<th>Main references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adegenet</td>
<td>boundary detection using Monmonier’s algorithm, sPCA</td>
<td>both</td>
<td>indiv. genotypes &amp; spatial coordinates</td>
<td>visual and text-based results</td>
<td>Win, Mac, Linux</td>
<td>adegenet.r-forge.r-project.org</td>
<td>Jombart (2008)</td>
</tr>
<tr>
<td>Ais</td>
<td>simple Mantel tests, spatial autocorrelation, Monmonier’s maximum difference</td>
<td>Indiv.</td>
<td>indiv. genotypes or genet. distances &amp; spatial coordinates</td>
<td>visual and text-based results</td>
<td>Win</td>
<td><a href="http://www.marksgeneticsoftware.net">www.marksgeneticsoftware.net</a></td>
<td>Miller (2005)</td>
</tr>
<tr>
<td>BAPS</td>
<td>clustering</td>
<td>Indiv.</td>
<td>indiv. genotypes &amp; spatial coordinates</td>
<td>cluster labels</td>
<td>Win, Mac, Linux</td>
<td>web.abo.fi/fak/mnf//mate/jc/software/baps.html</td>
<td>Corander et al. (2008)</td>
</tr>
<tr>
<td>Geneclust</td>
<td>clustering</td>
<td>Indiv.</td>
<td>indiv. genotypes &amp; spatial coordinates</td>
<td>cluster labels</td>
<td>Win, Mac, Linux</td>
<td><a href="http://www.sophie-ancelet.com">www.sophie-ancelet.com</a></td>
<td>François et al. (2006)</td>
</tr>
<tr>
<td>Geneland</td>
<td>clustering</td>
<td>Indiv.</td>
<td>indiv. genotypes &amp; spatial coordinates</td>
<td>cluster labels</td>
<td>Win, Mac, Linux</td>
<td>www2.imm.dtu.dk/~gigu/Geneland/</td>
<td>Guillot et al. (2005), Guillot et al. (2008), Guillot (2008)</td>
</tr>
<tr>
<td>Genepop</td>
<td>estimation of neighb. size by IBD analyses (a, e), Mantel tests</td>
<td>both</td>
<td>indiv. genotypes &amp; spatial coordinates</td>
<td>matrices of geog. and genet. distances, IBD regression estimates and confidence intervals, Mantel P-value</td>
<td>Win, Linux</td>
<td>kimura.univ-montp2.fr/~rousset/Genepop.htm</td>
<td>Rousset (2007)</td>
</tr>
<tr>
<td>Ibdsim</td>
<td>simulation of genotypic data under general isolation by distance models</td>
<td>Indiv.</td>
<td>simulation parameters</td>
<td>indiv. genotypes, spatial coordinates, summary statistics (genotypes, effective dispersal distribution)</td>
<td>Win, Mac, Linux</td>
<td>kimura.univ-montp2.fr/~rousset/IBDSim.html</td>
<td>Leblois et al. (2009)</td>
</tr>
<tr>
<td>Software</td>
<td>Description</td>
<td>Parameters</td>
<td>Results</td>
<td>Compatibility</td>
<td>Website</td>
<td>References</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>------------</td>
<td>---------</td>
<td>---------------</td>
<td>---------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>IBDWS</td>
<td>simple and partial Mantel tests, Reduced Major Axis regression</td>
<td>indiv. genotypes or genetic distances &amp; geog. distances &amp; optional indicator variable (column or matrix) pairwise distances</td>
<td>visual and text-based results incl. IBD slope and intercept</td>
<td>all</td>
<td>ibdws.sdsu.edu/~ibdws</td>
<td>Jensen et al. (2005)</td>
<td></td>
</tr>
<tr>
<td>PASSAGE</td>
<td>networks (e.g. Delaunay triangulation, tessellations), correlograms, semivariograms, kriging and spatial interpolation, simple and partial Mantel tests, local indicators of spatial association, boundary detection (wombling)</td>
<td>both genet. and geog. distances</td>
<td>visual &amp; text-based results incl. IBD slope and intercept</td>
<td>Win, Mac, Linux</td>
<td><a href="http://www.passagesoftware.net">www.passagesoftware.net</a></td>
<td>Rosenberg et al. (2001)</td>
<td></td>
</tr>
<tr>
<td>RMETASIM</td>
<td>spatially explicit simulation of genotypic data</td>
<td>Indiv. simulation parameters</td>
<td>genotypes, summary statistics</td>
<td>Win, Mac, Linux</td>
<td>cran.r-project.org/web/packages/rmetasim</td>
<td>Strand (2002)</td>
<td></td>
</tr>
<tr>
<td>SPAGEDI</td>
<td>IBD analyses (various estimators of genetic distance or relatedness) incl. Mantel tests and estimation of neighbourhood size by IBD analyses (a, Fst/(1-Fst), kinship coefficient), spatial autocorrelation</td>
<td>both individ. genotypes &amp; spatial coordinates</td>
<td>pairwise distances, text-based results</td>
<td>Win</td>
<td><a href="http://www.ulb.ac.be/sciences/ecoevol/spagedi.html">www.ulb.ac.be/sciences/ecoevol/spagedi.html</a></td>
<td>Hardy &amp; Velemans (2002)</td>
<td></td>
</tr>
<tr>
<td>SPLATCHE</td>
<td>simulation of genotypic data with possible incorporation of the influence of environmental parameters on migration</td>
<td>Pop. simulation parameters incl. environmental data</td>
<td>individ. genotypes, coalescence-related files, plots of simulated data</td>
<td>Win</td>
<td>cmpg.unibe.ch/software/splatche</td>
<td>Currat et al. (2004)</td>
<td></td>
</tr>
<tr>
<td>TFPGA</td>
<td>simple Mantel tests</td>
<td>both genet. and geog. distances</td>
<td>visual and text-based results</td>
<td>Win</td>
<td><a href="http://www.marksgeneticsoftware.net">www.marksgeneticsoftware.net</a></td>
<td>Miller (1997)</td>
<td></td>
</tr>
<tr>
<td>ZT</td>
<td>simple and partial Mantel tests</td>
<td>both genet. and geog. distances &amp; environmental variables</td>
<td>text-based results</td>
<td>Win, Mac, Linux</td>
<td><a href="http://www.psb.ugent.be/~erbon/mantel/">www.psb.ugent.be/~erbon/mantel/</a></td>
<td>Bonnet &amp; de Peer (2002)</td>
<td></td>
</tr>
</tbody>
</table>
References


Miller, M. (1997), ‘Tools for population genetic analyses (tfpga) 1.3: A windows program for the analysis of allozyme and molecular population genetic data. computer software distributed by author.’.


