

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

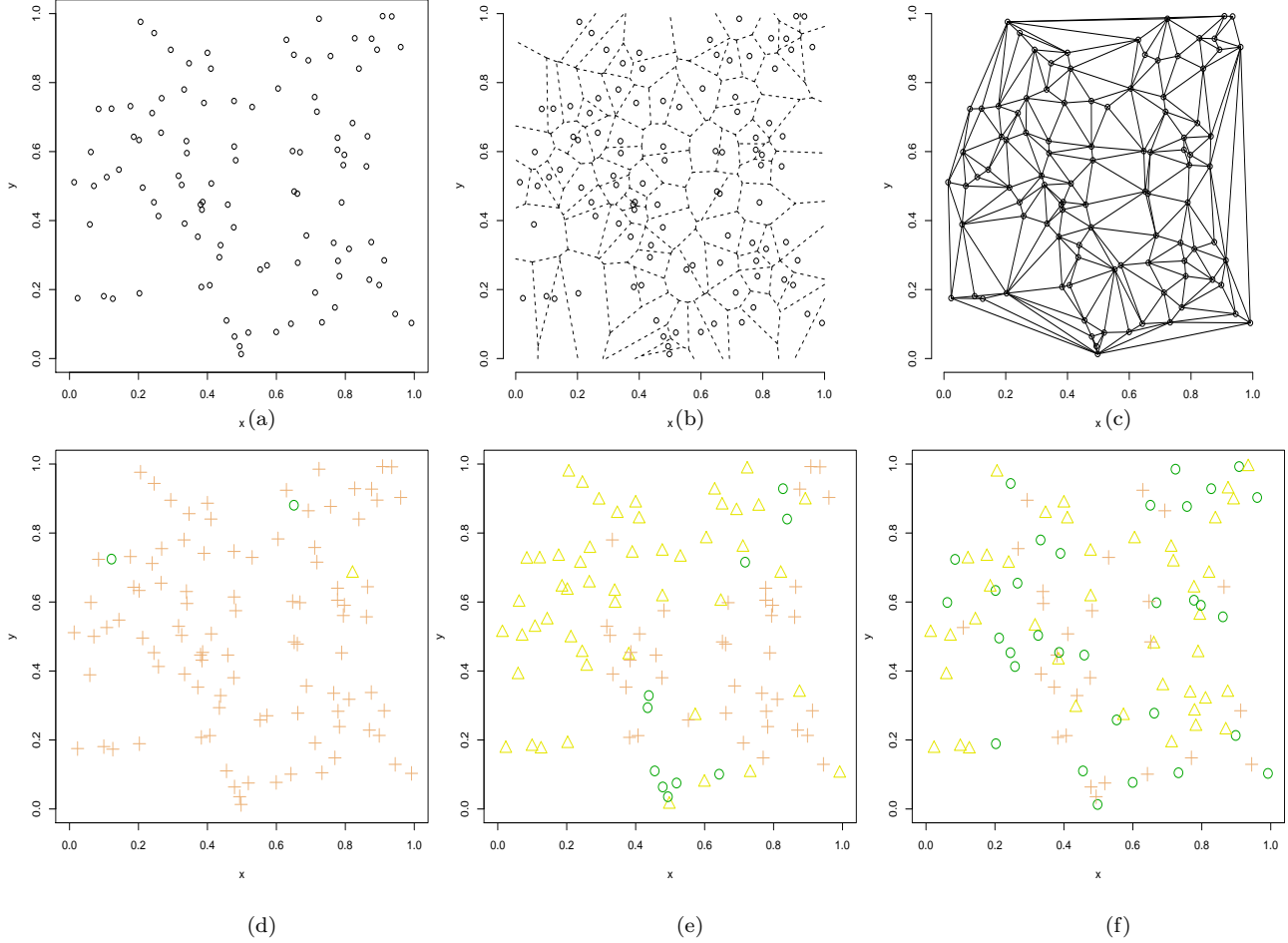


Figure I: Examples of simulated spatial patterns for population membership obtained by a constrained Voronoi tiling (induced by the sampling sites) and a Markov random field colouring. (a) locations of individuals here assumed to be regularly sampled over space; (b) Voronoi tessellation induced by the 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).

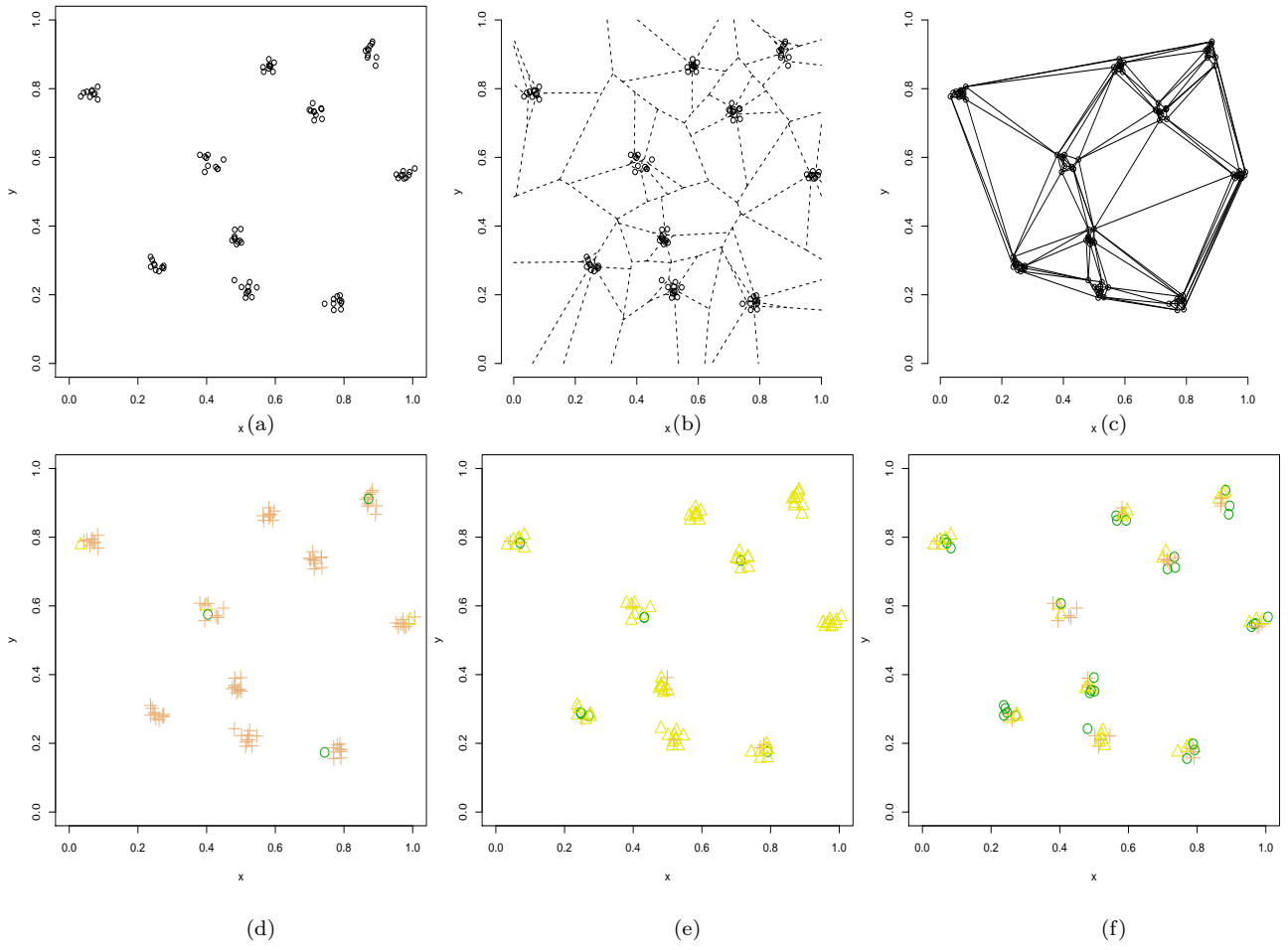


Figure II: Same as Fig. I with irregularly spaced sampling sites.

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

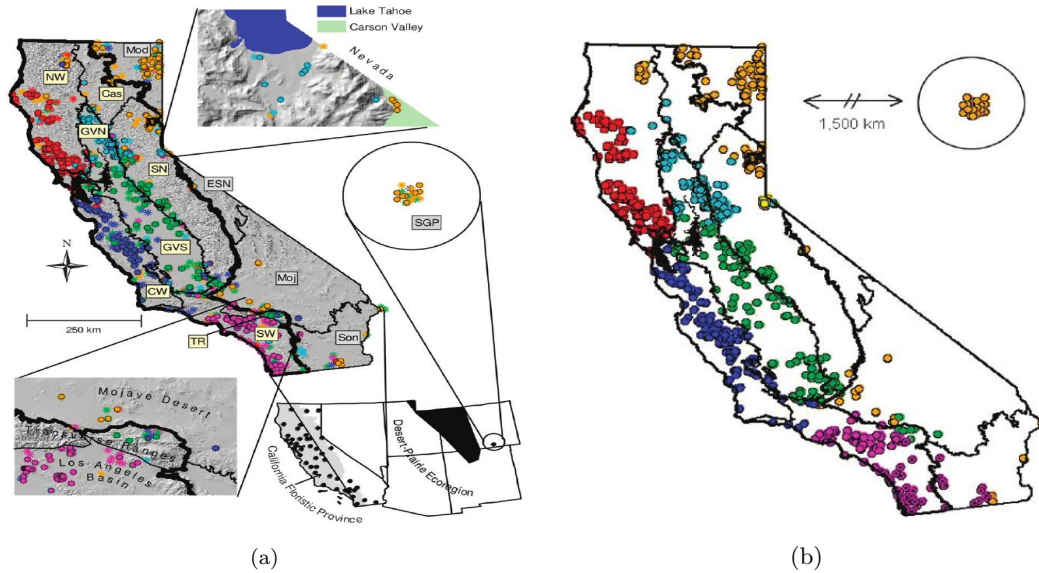


Figure III: Modal assignments of individual coyotes to different clusters using the (a) STRUCTURE algorithm ($K=6$) and the (b) spatially explicit GENELAND ($K=8$) algorithm. In the STRUCTURE-based assignments, coloured circles and asterisks indicate individuals with $> 80\%$ and $< 80\%$, respectively, 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 clusters 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 significant 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 increase. 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 white-tailed 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 1: List of main computer programs for spatial statistical analysis and spatially-explicit simulation of genetic data

Program	Spatial method(s)	Indiv. or pop.-level	Input	Output	OS	Web address	Main references
ADEGENET	boundary detection using Monmonier's algorithm, sPCA	both	indiv. genotypes & spatial coordinates	visual and text-based results	Win, Mac, Linux	adegenet.r-forge.r-project.org	Jombart (2008)
AIS	simple Mantel tests, spatial autocorrelation, Monmonier's maximum difference algorithm based on Delaunay triangulations	Indiv.	indiv. genotypes or genet. distances & spatial coordinates	visual and text-based results	Win	www.marksgeneticsoftware.net	Miller (2005)
BAPS	clustering	Indiv.	indiv. genotypes & spatial coordinates	cluster labels	Win, Mac, Linux	web.abo.fi/fak/mmf//mate/jc/software/baps.html	Corander et al. (2008)
EASYPOP	simulation of genotypic data under various migration models	both	simulation parameters	indiv. genotypes, genetic summary statistics, genealogy	Win, Mac	www.unil.ch/dec/page36926.fr.html	Balloux (2001)
GENECLUST	clustering	Indiv.	indiv. genotypes & spatial coordinates	cluster labels	Win, Mac, Linux	www.sophie-ancelet.com	François et al. (2006)
GENELAND	clustering	Indiv.	indiv. genotypes & spatial coordinates	cluster labels	Win, Mac, Linux	www2.imm.dtu.dk/~gigu/Geneland/	Guillot et al. (2005), Guillot et al. (2008), Guillot (2008)
GENEPOP	estimation of neighb. size by IBD analyses (a, e), Mantel tests	both	indiv. genotypes & spatial coordinates	matrices of geog. and genet. distances, IBD regression estimates and confidence intervals, Mantel P-value	Win, Linux	kimura.univ-montp2.fr/~rousset/Genepop.htm	Rousset (2007)
IBDSIM	simulation of genotypic data under general isolation by distance models	Indiv.	simulation parameters	indiv. genotypes, spatial coordinates, summary statistics (genotypes, effective dispersal distribution)	Win, Mac, Linux	kimura.univ-montp2.fr/~rousset/IBDSim.html	Leblois et al. (2009)

IBDWS	simple and partial Mantel tests, Reduced Major Axis regression	both	indiv. genotypes or genetic distances & geog. distances & optional indicator variable (column or matrix) pairwise distances	visual and text-based results incl. IBD slope and intercept	all	ibdws.sdsu.edu/~ibdws	Jensen et al. (2005)
PASSAGE	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)	both	genet. and geog. distances	visual & text-based results incl. IBD slope and intercept	Win, Mac, Linux	www.passagesoftware.net	Rosenberg et al. (2001)
RMETASIM	spatially explicit simulation of genotypic data	Indiv.	simulation parameters	genotypes, summary statistics	Win, Mac, Linux	cran.r-project.org/web/packages/rmetasim	Strand (2002)
SCAT	assignment of samples across a continuous range of locations	Indiv.	indiv. genotypes & spatial coordinates	text-based results	Win, Linux	stephenslab.uchicago.edu/scat/scat.html	Wasser et al. (2004) Wasser et al. (2007)
SPAGEDI	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	both	indiv. genotypes & spatial coordinates	pairwise distances, text-based results	Win	www.ulb.ac.be/sciences/ecevol/spagedi.html	Hardy & Vekemans (2002)
SPLATCHE	simulation of genotypic data with possible incorporation of the influence of environmental parameters on migration	Pop.	simulation parameters incl. environmental data	indiv. genotypes, coalescence-related files, plots of simulated data	Win	cmpg.unibe.ch/software/splatche	Currat et al. (2004)
TESS	clustering	Indiv.	indiv. genotypes & spatial coordinates	cluster labels	Win, Mac, Linux	membres-timc.imag.fr/Olivier.Francois/tess.html	Chen et al. (2007) Guillot (2009)
TTPGA	simple Mantel tests	both	genet. and geog. distances	visual and text-based results	Win	www.marksgeneticssoftware.net	Miller (1997)
ZT	simple and partial Mantel tests	both	genet. and geog. distances & environmental variables	text-based results	Win, Mac, Linux	www.psb.ugent.be/~erbon/mantel/	Bonnet & de Peer (2002)

References

- Alberto, F., Massa, S., Manent, P., Diaz-Almela, E., Arnaud-Haond, S., Duarte, C. M. & ao, E. A. S. (2008), 'Genetic differentiation and secondary contact zone in the seagrass *cymodocea nodosa* across the mediterranean atlantic transition region', *Journal of Biogeography* **35**, 1279–1294.
- Andrén, H. (1994), 'Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review', *Oikos* **71**, 355–366.
- Balloux, F. (2001), 'Easypop (version 1.7), a computer program for the simulation of population genetics', *Journal of Heredity* **92**, 301–302.
- Blanchong, J., Samuel, M., Scribner, K., Weckworth, B., Langenberg, J. & Filcek, K. (2008), 'Landscape genetics and the distribution of chronic wasting disease', *Biology Letters* **4**, 130–133.
- Bonnet, E. & de Peer, Y. V. (2002), 'zt: a software tool for simple and partial mantel tests', *Journal of Statistical software* **7**(10), 1–12.
- Chen, C., Durand, E., Forbes, F. & François, O. (2007), 'Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study', *Molecular Ecology Notes* **7**(5), 747–756.
- Corander, J., Sirén, J. & Arjas, E. (2008), 'Bayesian spatial modeling of genetic population structure', *Computational Statistics* **23**(1), 111–129.
- Coulon, A., Guillot, G., Cosson, J., Angibault, J., Aulagnier, S., Cargnelutti, B., Galan, M. & Hewison, A. (2006), 'Genetics structure is influenced by landscape features. Empirical evidence from a roe deer population', *Molecular Ecology* **15**, 1669–1679.
- Cullingham, C., Kyle, C., Pond, B., Rees, E. & White, B. N. (2009), 'Differential permeability of rivers to racoon gene flow corresponds to rabies incidence on Ontario, Canada', *Molecular Ecology* **18**, 43–53.
- Currat, M., Ray, N. & Excoffier, L. (2004), 'Splathe: a program to simulate genetic diversity taking into account environmental heterogeneity', *Molecular Ecology Notes* **4**, 139–142.
- Davis, J. & Stamps, J. (2004), 'The effect of natal experience on habitat preferences', *Trends in Ecology and Evolution* **19**, 411–416.

- Dionne, M., Caron, F., Dodson, J. & Bernatchez, L. (2008), ‘Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation’, *Molecular Ecology* **17**, 2382–2396.
- E. Leclerc, Y. Mailhot, M. M. & Bernatchez, L. (2008), ‘The landscape genetics of yellow perch (*perca flavescens*) in a large fluvial ecosystem’, *Molecular Ecology* **17**, 1702–1717.
- Florin, A. & Höglund, J. (2008), ‘Population structure of flounder (*platichthys flesus*) in the baltic sea: differences among demersal and pelagic spawners’, *Heredity* **101**, 27–38.
- François, O., Ancelet, S. & Guillot, G. (2006), ‘Bayesian clustering using hidden Markov random fields’, *Genetics* **174**, 805–816.
- Guillot, G. (2008), ‘Inference of structure in subdivided populations at low levels of genetic differentiation. The correlated allele frequencies model revisited’, *Bioinformatics* **24**, 2222–2228.
- Guillot, G. (2009), ‘On the inference of spatial structure from population genetics data using the Tess program.’, *Bioinformatics* . To appear.
- Guillot, G., Estoup, A., Mortier, F. & Cosson, J. (2005), ‘A spatial statistical model for landscape genetics’, *Genetics* **170**(3), 1261–1280.
- Guillot, G., Santos, F. & Estoup, A. (2008), ‘Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface.’, *Bioinformatics* **24**(11), 1406–1407.
- Hanson, T., Brunsfeld, S. J., Finegan, B. & Waits, L. (2008), ‘Pollen dispersal and genetic structure of the tropical tree *dipteryx panamensis* in a fragmented Costa Rican landscape’, *Molecular Ecology* **17**(8), 2060–2073.
- Hardy, O. & Vekemans, X. (2002), ‘Spagedi : a versatile computer program to analyse spatial structure at the individual or population levels’, *Molecular Ecology Notes* **2**, 618–620.
- Jensen, J., Bohonak, A. & Kelley, S. (2005), ‘Isolation by distance, web service’, *BMC Genetics* **6**(13).
- Jombart, T. (2008), ‘adeigenet: a r package for the multivariate analysis of genetic markers’, *Bioinformatics* **24**, 1403–1405.
- Latch, E. K., Scognamillo, D. G., Fike, J. A., Chamberlain, M. J. & Jr, O. E. R. (2008), ‘Deciphering ecological barriers to North American river otter (*lontra canadensis*) gene flow in the louisiana landscape’, *Journal of Heredity* **99**(3), 265–274.

- Leblois, R., Estoup, A. & Rousset, F. (2009), 'IBDsim: a computer program to simulate genotypic data under isolation by distance', *Molecular Ecology Resources* **9**, 107–109.
- Miller, M. (1997), 'Tools for population genetic analyses (tfgpa) 1.3: A windows program for the analysis of allozyme and molecular population genetic data. computer software distributed by author.'
- Miller, M. (2005), 'Alleles in space (AIS): Computer software for the joint analysis of interindividual spatial and genetic information', *Heredity* **96**(6), 722–724.
- Pampoulie, C., íslason, D. G. & Daníelsdóttir, A. K. (2009), 'A seascape genetic snapshot of *sebastes marinus* calls for further investigation across the north atlantic', *ICES Journal of Marine Science* . doi:10.1093/icesjms/fsp199.
- Pope, L., Pope, B. R., Wilson, G., Woodroffe, R., Erven, K., Conyers, C., Franklin, T., Delahay, R., Cheeseman, C. & Burke, T. (2007), 'Genetic evidence that culling increases badger movement: implications for the spread of bovine tuberculosis', *Molecular Ecology* **16**, 4919–4929.
- Rees, E., Pond, B., Cullingham, C., Tinline, R., Ball, D. & nd White B.N. Kyle, C. (2008), 'Assessing a landscape barrier using genetic simulation modelling: implications for raccoon rabies management', *Preventive Veterinary Medicine* **86**, 107–123.
- Riley, S., Pollinger, J., Sauvajot, R., Bromley, E. Y., Fuller, T. & Wayne, R. (2006), 'A southern California freeway is a physical and social barrier to gene flow in carnivores', *Molecular Ecology* **15**, 1733–1741.
- Rosenberg, N., Burke, T., Elo, K., Feldman, M., Friedlin, P., Groenen, M., Hillel, J., Mäki-Tanila, A., Tixier-Boichard, M., Vignal, A., Wimmers, K. & Weigend, S. (2001), 'Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds', *Genetics* **159**, 699–713.
- Rousset, F. (2007), 'Genepop'007: a complete re-implementation of the Genepop software for windows and linux', *Molecular Ecology Notes* **8**(1), 103–106.
- Sacks, B., Bannasch, D. L., Chomel, B. B. & Ernst, H. (2008), 'Coyotes demonstrate how habitat specialization by individuals of a generalist species can diversify populations in a heterogeneous ecoregion', *Molecular Biology and Evolution* **25**(7), 1354–1395.
- Sacks, B. N., Brown, S. K. & Ernest, H. B. (2004), 'Population structure of california coyotes corresponds to habitat-specific breaks and illuminates species history', *Molecular Ecology* **13**, 1265–1275.

- Scott-McCairns, R. J. & Bernatchez, L. (2008), ‘Landscape genetic analyses reveal cryptic population structure and putative selection gradients in a large-scale estuarine environment’, *Molecular Ecology* **17**, 3901–3916.
- Sloan, C., Duell, E., Shi, X., Irwin, R., Andrew, A., Williams, S. & Moore, J. (2009), ‘Ecogeographic genetic epidemiology’, *Genetic epidemiology* .
- Strand, A. (2002), ‘Metasim 1.0: an individual-based environment for simulating population genetics of complex population dynamics’, *Molecular Ecology Notes* **2**, 373–376.
- Strasburg, J. (2006), ‘Roads and genetic connectivity’, *Nature* **440**, 875–876.
- Walsh, P., Biek, R. & Real, L. (2005), ‘Wave-like spread of Ebola Zaire’, *Public Library of Science, Biology* **3**, e371.
- Wasser, S., Mailand, C., Booth, R., Mutayoba, B., Kisamo, E. & Stephens, M. (2007), ‘Using DNA to track the origin of the largest ivory seizure since the 1989 trade ban’, *Proceedings of the National Academy of Sciences* **104**(10), 4228–4233.
- Wasser, S., Shedlock, A., Comstock, K., Ostrander, E., Mutayoba, B. & Stephens, M. (2004), ‘Assigning African elephants DNA to geographic region of origin: applications to the ivory trade’, *Proceedings of the National Academy of Sciences* **101**(41), 14847–14852.