

A multivariate method to investigate the spatial patterns of the genetic variability

Thibaut Jombart¹ and Anne-Béatrice Dufour¹

¹ Université de Lyon; Université Lyon 1; CNRS; UMR 5558, Laboratoire de Biométrie et Biologie Evolutive, France.

A challenging issue in current molecular ecology is to take landscape information into account as well as genotyping data to infer biological processes [1]. Among landscape features, space is most likely to influence the genetic structuring of a set of individuals or populations, and must therefore be included in the statistical analysis of genotyping data. In order to reveal spatial patterns, a statistical method should be spatially explicit, *i.e.* it should directly take spatial information into account as a component of the adjusted model or of the optimized criterion. We propose a new spatially explicit multivariate method, *spatial principal component analysis* (sPCA), to investigate the spatial pattern of genetic variability using allelic frequency data of individuals or populations. The sPCA relies on a modification of the principal component analysis (PCA, [2]) so that not only the variance between the studied entities, but also their spatial autocorrelation is taken into account. This is achieved by finding a set of orthonormal axes in the multivariate space of biological entities which decompose both their variance and their spatial autocorrelation. The main results of the sPCA are maps of entities scores allowing a visual assessment of the spatial genetic structures. Global structures (patches or clines) are disentangled from local ones (strong genetic differences between neighbours) and from random noise. Two statistical tests are also proposed to detect the existence of both types of patterns. As an illustration, the results of PCA and sPCA are compared using simulated datasets. sPCA performed better than PCA to reveal spatial genetic patterns. The proposed methodology is implemented in the adegenet package [3] of the free software R [4].

References

- [1] S. Manel, M.K. Schwartz, G. Luikart and P. Taberlet (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, 18(4):189–197.
- [2] K. Pearson (1901) On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572.
- [3] T. Jombart (2008) The adegenet package for R: classes and methods for the multivariate analysis of genetic markers. URL: <http://pbil.univ-lyon1.fr/software/adegenet/>.
- [4] R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL: <http://www.R-project.org>.