

Multivariate analysis of genetic markers as a tool to explore the genetic diversity

Thibaut Jombart – Imperial College London

tjombart@imperial.ac.uk

Multivariate analysis of genetic markers



Biological processes (demography, dispersal, selection)

Spatial genetic structures





Biological processes (demography, dispersal, selection)

Spatial genetic structures



Identify genetic patterns to infer biological processes structuring the biodiversity

Multivariate analysis of genetic markers



Summarise the genetic diversity among individuals / populations

Multivariate analysis – rationale (1/2)





Multivariate analysis – rationale (1/2)



Multivariate analysis – rationale (1/2)



Summarise the genetic diversity among individuals / populations

Multivariate analysis – rationale (2/2)



Multivariate analysis of genetic markers



To name a few (used in genetics):

- Principal Component Analysis (PCA)
 - > centred / not centred / fancy centring
 - > scaled / not scaled / fancy scaling
 - > transformed for compositional data
- Principal Coordinates Analysis (PcoA), aka (Metric) Multidimensional Scaling (MDS)
 many genetic distances
- Correspondence Analysis (CA)
- Discriminant Analysis (DA)
- ... (review in Jombart et al. 2009, Heredity 102, 330-341)

What to do with principal components?(1/3)

Get a picture of genetic diversity



Multivariate analysis of genetic markers





(Mulley et al. 1979)

Multivariate analysis of genetic markers

What to do with principal components?(3/3)

Make geographical maps of the genetic differentiation



Why look for spatial genetic structures?

Most genetic models predict that genetic diversity should be spatially structured:









• Usual multivariate methods do not use spatial information.



• They can reveal `obvious' spatial patterns, but will overlook more subtle structures.

•To seek spatial genetic structures, we must find the part of the **genetic variability related to spatial proximity** between individuals/populations.



Spatial Principal Component Analysis (sPCA) (Jombart et al. 2009)



Population genetic software – the scary picture



(Excoffier & Heckel 2006)

`In a perfect world, research teams would be able to develop analysis tools to address their specific problem, but in practice they have to make their data fit the available tools, leading to obvious discrepancies between the initial goals and the results'

Taking genetic markers into the field of (multivariate) statistics

Population genetic software:

- very few multivariate methods
- no plasticity
- poor data interoperability

many multivariate methods

The **R** software:

- total plasticity
- tons of statistic methods (tests, modeling, Monte-Carlo)
- great graphics
- great interoperability (e.g., GIS)
- programming language
- free software





changed directly with the other programs



Purpose:

- take genetic markers into a suitable format
- adapt multivariate methods to genetic markers
- provide advanced data handling
- implement novel methods (*e.g.*, sPCA)





Multivariate analysis of genetic markers



Where to get information:

- reference: Jombart (2008) Bioinformatics 24: 1403-1405
- adegenet website: http://adegenet.r-forge.r-project.org/
- *adegenet* forum: adegenet_forum@lists.r_forge.r_project.org
- Here and now!