



Mapping Averaged Pairwise Information (MAPI): a new exploratory tool to uncover spatial structure.

Sylvain S. Piry, Marie Pierre Chapuis, Bertrand Gauffre, Julien Papaïx, Astrid Cruaud, Karine Berthier

► To cite this version:

Sylvain S. Piry, Marie Pierre Chapuis, Bertrand Gauffre, Julien Papaïx, Astrid Cruaud, et al.. Mapping Averaged Pairwise Information (MAPI): a new exploratory tool to uncover spatial structure.. Methods in Ecology and Evolution, 2016, 7 (12), pp.1463-1475. 10.1111/2041-210X.12616 . hal-01506806

HAL Id: hal-01506806

<https://hal.science/hal-01506806>

Submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Received Date : 24-Apr-2015
Revised Date : 22-Jun-2016
Accepted Date : 05-Jul-2016
Article type : Research Article
Editor : Stephane Dray

Mapping Averaged Pairwise Information (MAPI):

a new exploratory tool to uncover spatial structure.

Sylvain Piry¹, Marie-Pierre Chapuis², Bertrand Gauffre^{3,4}, Julien Papaix⁵, Astrid Cruaud¹, Karine Berthier⁶

¹ INRA, UMR 1062, Centre de Biologie pour la Gestion des Populations, 755 avenue du campus Agropolis, 34988 Montferrier sur Lez, France.

² CIRAD, UMR 1062, Centre de Biologie pour la Gestion des Populations, 755 avenue du campus Agropolis, 34988 Montferrier sur Lez, France.

³ CEBC-CNRS (UPR 1934), F-79360, Beauvoir sur Niort, France.

⁴ INRA, USC1339 (CEBC-CNRS), F-79360, Beauvoir sur Niort, France.

⁵ INRA, UR546 Biostatistique et Processus Spatiaux, F-84914 Avignon, France.

⁶ INRA, UR0407 Pathologie Végétale, F-84143 Montfavet, France.

Running title: MAPI: exploring spatial structure

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/2041-210X.12616

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

Keywords: pairwise information, geovisualisation, exploratory method, spatial genetics, landscape genetics, isolation-by-distance, barrier, gene flow, GIS, database

Corresponding author:

Sylvain Piry

INRA, UMR 1062, Centre de Biologie pour la Gestion des Populations

755 avenue du campus Agropolis

34988, Montferrier-sur-Lez

France

Tel: +33(0)499623337 - Email: piry@supagro.inra.fr

ABSTRACT (267 words)

1. Visualisation of spatial networks based on pairwise metrics such as (dis)similarity coefficients provides direct information on spatial organization of biological systems. However, for large networks, graphical representations are often unreadable as nodes (samples), and edges (links between samples) strongly overlap. We present a new method, MAPI (Mapping Averaged Pairwise Information), allowing translation from spatial networks to variation surfaces.

2. MAPI relies on i) a spatial network in which samples are linked by ellipses and, ii) a grid of hexagonal cells encompassing the study area. Pairwise metric values are attributed to ellipses and averaged within the cells they intersect. The resulting surface of variation can be displayed as a colour map in Geographical Information System (GIS), along with other relevant layers, such as land cover. The method also allows the identification of significant discontinuity in grid cell values through a nonparametric randomisation procedure.

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : [10.1111/2041-210X.12616](https://doi.org/10.1111/2041-210X.12616)

3. The interest of MAPI is here demonstrated in the field of spatial and landscape genetics. Using simulated test datasets, as well as observed data from three biological models, we show that MAPI is
- i) relatively insensitive to confounding effects resulting from isolation-by-distance (*i.e.* over-structuring),
 - ii) efficient in detecting barriers when they are not too permeable to gene flow and,
 - iii) useful to explore relationships between spatial genetic patterns and landscape features.
4. MAPI is freely provided as a PostgreSQL/PostGIS database extension allowing easy interaction with GIS or the R software and other programming languages. Although developed for spatial and landscape genetics, the method can also be useful to visualise spatial organisation from other kinds of data from which pairwise metrics can be computed.

INTRODUCTION

Computation of metrics to estimate similarity, difference or flow between samples is a common approach to study a wide range of social and biological systems (Borgatti et al. 2009; Miele et al. 2015). Computing pairwise metrics allows the construction of networks with nodes (samples) being connected by edges that are associated with the value of the metric of interest. When geographic coordinates of samples are available, such networks are inherently informative about the spatial organization of the system under study (Barthélemy 2011). This organization can be visualised by superimposing the network on a geographic map, with the thickness or colour of the edges displayed according to the pairwise values. However, such graphical representations are difficult to interpret for large networks as the numerous nodes and edges strongly overlap and obscure both the network and the underlying geographic space (Hennemann 2013). To better visualise spatial information, variation surfaces can be produced using interpolation procedures (e.g. kernel density estimation, kriging). To do so, pairwise metric values have first to be attributed to unique geographical points (*i.e.* transformed into punctual values) as, for example, to the middle of the edges connecting the samples in the network (e.g. Miller 2005).

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

In this paper we present a novel method, MAPI (Mapping Averaged Pairwise Information), to produce variation surfaces from pairwise metrics computed between georeferenced samples. In essence, MAPI is a non-parametric smoothing procedure applied to pairwise values rather than punctual values. To smooth out pairwise values, the straight lines connecting the samples in the network are replaced by ellipsoidal polygons for which the foci are located on the two samples being connected. Then, similarly to kernel density estimators for punctual values, ellipses constitute a geometric shape used to average information between overlapping connections. This smoothing procedure produces a two-dimensional geographical layer that can be easily visualised and customised in a GIS. Although this method may be of interest in other research fields, we will hereafter focus on spatial and landscape genetics, where pairwise metrics are widely used but flexible visualisation tools are still lacking for these measures (see Miller 2005; Vandergast et al. 2010, Etherington 2011, Petkova et al. 2016).

Although new methods, such as spatially-explicit clustering, are increasingly used in spatial genetics to describe genetic variation and identify barriers to gene flow (Guillot et al. 2009), pairwise genetic measures, such as distance, remain appealing as: i) they can be computed between individuals or populations, ii) they facilitate handling massive datasets, which are increasing with the advance in high-throughput sequencing (Duforet-Frebourg and Blum 2014) and, iii) their regression against geographic distances may be informative with respect to isolation-by-distance (IBD), i.e. the increase of genetic differentiation with spatial separation due to restricted dispersal (Wright 1943). The original IBD model has been further extended in landscape genetics to assess the impact of landscape heterogeneity on spatial genetic structure and gene flow. In this field, replacing straight line distances between samples by ecological distances has given rise to concepts of isolation-by-barrier, isolation-by-resistance and isolation-by-environment (Balkenhol et al. 2015). Nowadays, various approaches using least cost paths, circuit theory or population graphs have been developed to assess the impact of landscape features on genetic structure and gene flow (e.g. Coulon et al.

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

2004; Cushman et al. 2006; McRae 2006; McRae & Beier 2007; Garroway et al. 2008; Dyer & Nason 2004; Dyer 2015, Petkova et al. 2016).

Landscape genetics approaches often require *a priori* knowledge on species–landscape interactions to target relevant environmental variables, set cost/resistance values and account for demographic effects such as population size (e.g. Broquet et al. 2006; Weckworth et al. 2013). These methods can be difficult to apply when little is known about the species under study. In such a situation, we still rely on exploratory analyses to identify environmental variables of interest and draw up hypotheses that can be further tested using appropriate sampling schemes (Kelling et al. 2009; Richardson et al. 2016). Although the visualisation of pairwise genetic measures may be an obvious first step in such an approach (Etherington 2011), there are still very few tools allowing visualisation of such measures along with environmental layers without, first, attributing features such as cost or resistance values to habitat types. Among those tools, Allele In Space (Miller 2005) and the GIS toolbox of Vandergast et al. (2010) offer the rare possibility to produce variation surfaces of genetic distances that can be mapped over landscape layers (Wood et al. 2013; Adams & Burg 2015). Both tools rely on an inverse-distance-weighted interpolation of pairwise measures that are first spatially attributed to the middle of the segments linking the samples in the network. Confounding effects due to IBD can be somewhat limited by i) using residuals from the regression of genetic distances against geographic distances (Miller 2005) or, ii) limiting the network to the nearest neighbours for each sample (Vandergast et al. 2010).

Disentangling the relative contributions of IBD and environmental features in shaping spatial genetic patterns remains a central issue in landscape genetics (Bradburd et al. 2013; Wang et al. 2014). This question is often addressed using Mantel and partial Mantel tests between matrices of population- or individual-based genetic distances, Euclidean distances and ecological distances (e.g.

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

Cushman et al. 2006; Hagerty et al. 2011). There is an ongoing debate on the statistical performance of the Mantel test in spatial and landscape genetics (see Raufaste & Rousset 2001; Legendre & Fortin 2010; Cushman & Landguth 2010; Guillot & Rousset 2013; Cushman et al. 2013) and, recently, alternative methods using multivariate or geostatistical techniques have been proposed to disentangle the effects of geographic distance and environmental heterogeneity on spatial genetic structure (e.g. Duforet-Frebourg and Blum 2014; Bradburd et al. 2013; Galpern et al. 2014; Botta et al. 2015). Moreover, results from lot of methods do not include spatially-explicit variation surfaces that can be mapped over landscape layers to facilitate their interpretation.

In this general context, MAPI provides an approach to i) visualise pairwise genetic measures without confounding effects resulting from IBD, ii) test for spatial genetic discontinuity through a nonparametric randomisation procedure and, iii) explore relationships between observed genetic patterns and environmental heterogeneity, notably by using MAPI results for further statistical analyses. In this work, we assessed the efficiency of MAPI to accurately detect spatial discontinuity in pairwise genetic metrics by applying permutation tests on controlled simulated datasets of panmictic populations and populations under isolation-by-distance, including scenarios of separation by a linear barrier to gene flow. We also appraised the potential of MAPI in a landscape genetics framework by analysing genotypes simulated under landscape constraints with various spatial configurations of favourable and unfavourable habitats. We used a Bayesian conditional autoregressive model to analyse the relationships between MAPI results and landscape variables. To illustrate MAPI we also re-analysed published data from three biological models: 1) microsatellite genotypes from a rodent population under IBD, 2) DNA sequences from a plant virus exhibiting spatial genetic discontinuities and, 3) microsatellites and landscape data from both a forest specialist and a generalist species of ground beetles. Finally, we used controlled and observed datasets to explore MAPI sensitivity to parameter setting and sampling scheme.

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

METHODS

MAPI methodology

MAPI is implemented as an open-source SQL extension within open source database PostgreSQL 9.x and Postgis 2.x (1996-2013, The PostgreSQL Global Development Group: <http://www.postgresql.org/> - <http://postgis.net/>). Source code, user manual, and training datasets can be downloaded from: <https://www1.montpellier.inra.fr/CBGP/software/MAPI/>. Importation of data files and running of MAPI commands can be easily done from the software R (R core Team 2015). We hereafter detail the steps of MAPI, with Figures 1 and 2 illustrating the general framework and Table 1 providing hints for setting parameter values.

Data Input. The method requires sample geographical coordinates and pairwise metric values computed between samples (e.g. genetic distance).

Grid of cells. A grid of hexagonal cells is superimposed on the area defined by the convex hull of the sampling points. Based on the Nyquist Frequency concept, optimized cell size should be at most half the average distance between closest samples. The half-width (h_w) of the cells is computed as:

$$h_w = \sqrt{\frac{\beta^2}{\frac{3\sqrt{3}}{2}} \cdot \frac{A}{N}}$$
 with N the number of sampling points, A the surface's area defined by their convex

hull and β a parameter depending on their spatial dispersion: 0.5 for regular sampling and 0.25 for irregular sampling (see Hengl 2006). The grid is automatically generated by informing the β parameter. The final number of cells (n_c) in the grid is then computed as: $n_c \approx \frac{A}{2.598 \cdot h_w^2}$.

Alternatively, a user-defined grid of cells can be imported in PostgreSQL (see user manual for details on how to build or import a grid).

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

Network of ellipses. Network edges are formed by ellipsoidal polygons (hereafter referred to as ellipses) for which foci are the geographical locations of the two samples being connected. The shape of these polygons can be adjusted by two parameters: 1) the *eccentricity* of the ellipses which controls the smoothing intensity and must be greater than 0 (infinite circle) and smaller than 1 (straight line) and 2) the radius of the error circle that controls for uncertainty on sample coordinates (*error_circle_radius*). In addition, two optional parameters (*min_distance* and *max_distance*) limit the analysis to a given range of between-sample distances. Effects of these four parameters on the shape of the network are detailed in section 1 of the Appendix.

Surface. The ellipses receive the value of the metric computed between the samples they connect and the cells of the grid receive the weighted arithmetic mean (m_w) of the n_e ellipses intercepting

their geographical extent, computed as: $m_w = \frac{\sum_{i=1}^{n_e} v_i}{s_w}$, with v_i and a_i the metric value and area of the

ellipse i , respectively, and s_w , the sum-of-weights of the ellipses defined as: $s_w = \sum_{i=1}^{n_e} \frac{1}{a_i}$. This weighting procedure limits long distance effects as long and inflated ellipses participate far less than short ellipses to the computation of m_w . Cells not intersected by ellipse have no m_w value and are not included in the final result.

Test of significance for spatial structure. We implemented a nonparametric randomisation procedure, as schematised in the Figure 2, to test whether the pairwise metric values associated to the ellipses are independent of the sample locations (i.e. under the null hypothesis that pairwise metric values, and then m_w cell values, are randomly distributed in space). Sample locations from the observed dataset (Figure 2A1) are permuted n_p times (Figure 2B1) (parameter $n_permutations$). At each permutation, new m_w values are computed and stored to build a cumulative null distribution for each of the n_c cells of the grid (Figure 2B2). Each m_w cell value from the observed dataset is then ranked against its null distribution (Figure 2C1). For each cell, the proportion of permuted values that are smaller than the observed value provides a lower-tailed test (LT) p-value. An upper-tailed

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

test (UT) p-value is also computed for each cell as: $1 - (\text{LT p-value})$. As the probability to find significant cells only by chance (i.e. type I error) increases with the number of tests performed (i.e. n_c tests), the False Discovery Rate (FDR) procedure proposed by Benjamini & Yekutieli (2001) is applied to account for multiple testing under positive dependency conditions (i.e. spatial autocorrelation between cells) (Figure 2C2). The significance level at which FDR is controlled can be set by users through the parameter *my_alpha*. For example, when *my_alpha* is set to 0.05, this means that 5% of the cells detected as significant can be false positives. Finally, for each test, significant cells that are spatially connected are aggregated together. When using metrics estimating differences, such as distances, significant highest m_w values localise areas of higher dissimilarity than expected by chance (hereafter referred to as discontinuous areas) while significant lowest m_w values localise areas of higher similarity than expected by chance (hereafter referred to as continuous areas).

Data output. Cell-specific information, such as their geometry and associated m_w and p-values, are stored as a spatial PostgreSQL table which can be exported as a shapefile or text file and imported into GIS software or the software R using the rgdal package. These data can be used for further statistical analyses of relationships between genetic measures and landscape variables.

Visualisation. On the final graphical output (Figure 2D), the observed cell values (m_w) are displayed using a colour scale. Cells with significant extreme values after FDR correction are visualised as black contours mapped as an additional layer. When significant cells are spatially connected, only the perimeter of the polygon they form is represented. Vector or raster maps, satellite imagery or teledetection products can be inserted below the MAPI layers for landscape interpretation.

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

Simulated data in spatial genetics

Using backward-in-time (genealogy-based) simulation algorithms, we simulated datasets of 10 microsatellite genotypes for 200 diploid individuals distributed on the nodes of a 20x10 lattice (i.e. one individual per node). Three demographic models with different sets of parameter values were analysed (25 simulated datasets each): (1) one IBD population without barriers to gene flow (spatial test datasets 1; *IBD*); (2) two panmictic populations separated by a barrier (spatial test datasets 2; *Barrier*); (3) two IBD populations separated by a barrier (spatial test datasets 3; *Barrier & IBD*) (see Figure 3 to visualise the simulated lattice). A null model of one panmictic population without a barrier to gene flow was also considered (spatial test datasets 0; *Panmixia*). The permeability of the barrier is controlled by setting the $N_e m$ parameter, i.e. the product of the effective population size and migration rate per generation. Details on simulation algorithms, parameter settings and basic measures of genetic variation can be found in sections 2.1 and 2.3 of the Appendix. We investigated MAPI sensitivity to sampling effort by (1) randomly re-sampling 75 individuals out of the 200 simulated for each of the test datasets, which provided a test for how MAPI deals with both random irregular sampling and small sample size; (2) simulating additional test datasets with 100 loci for a representative subset of spatial datasets under both large regular and small irregular samplings (see Table 2 to identify test datasets).

Analyses of simulated test datasets with MAPI were performed using a grid of hexagonal cells defined by setting the parameter β to 0.5 and 0.25 for regular and irregular samplings, respectively (Table 1). The network was built using the genetic distance a_r (Rousset 2000), an error radius for sample location of 0.01 and ellipses with an eccentricity value of 0.975. Cells with extreme high m_w values were detected using 1,000 permutations of the sample locations and a significance level of 0.05 for computing significance for each cell. As the exact location of the barrier was known, we computed the proportion of the barrier covered by cells with significantly higher values

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

(hereafter referred to as Coverage). We also computed the proportion of the study area covered by cells with significantly higher values having no contact with the simulated barrier, which may result from stochasticity, edge effects or IBD (hereafter referred to as Unexpected). We also investigated the sensitivity of MAPI to the eccentricity value (0.900, 0.975 and 0.999) and number of cells (setting the parameter β to 0.15, 0.25, 0.5 and 0.75). To this aim, we used a subset of datasets simulated under both large regular sampling and small irregular sampling (see section 2.1 in the Appendix to identify test datasets).

Simulated data in landscape genetics

Using a forward-in-time (individual-based) simulation algorithm, we simulated datasets of 10 microsatellite genotypes for diploid individuals distributed in a landscape raster of about 50x50 cells.

Three landscape models, which consider two habitats with contrasted carrying capacity (20 and 2 for the favourable and unfavourable habitat, respectively) were analysed (20 simulated datasets each):

(1) a spatial transition from the favourable to unfavourable habitat with a high level of interpenetration (landscape test datasets 1; *Gradient*); (2) fragmentation of the favourable habitat in small areas isolated by a prevalent unfavourable habitat (landscape test datasets 2; *Fragmentation*);

(3) a random distribution of the favourable and unfavourable habitats over the study area (landscape test datasets 3; *Random*) (see Figure 4 to visualise the simulated landscapes). For the

latter, the detection of the habitat effect is unexpected since the spatial scale of autocorrelation in

habitat is very small. Details on simulation algorithms, parameter settings and basic measures of genetic variation can be found in sections 2.2 and 2.3 of the Appendix. For each of the simulated

datasets, we also investigated MAPI sensitivity to sampling effort by sampling 200 and 500

individuals under three sampling schemes : (1) random sampling, in which individuals were sampled

anywhere in the landscape regardless to the habitat; (2) balanced sampling, in which an equal

number of individuals was sampled from each habitat; (3) gridded sampling, in which the same

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

number of individuals was randomly sampled from each square of a 3x3 grid encompassing the landscape raster.

Analyses of simulated datasets with MAPI were performed using grids constituted of squared cells that matched the landscape raster pixels (side length=1), the genetic distance a_r (Rousset 2000), an error radius for sample location of 0.5 to consider that individuals can be anywhere within a cell, ellipses with an eccentricity of 0.975 and a minimal distance between samples of 0.1 to exclude intra-cell connections as individuals simulated within a same cell have the same geographical coordinates. For each simulation and sampling scheme, we extracted from each cell both, the m_w value and habitat type, which was expressed as a factor according to favourability (hab): 1 for the favourable habitat and 2 for the unfavourable habitat. We fitted a regression model as follows: $m_{wi} = \alpha_{hab(i)} + \varepsilon_i$ where $\alpha_{hab(i)}$ is an intercept depending on the habitat type in cell i . To account for spatial correlations it was assumed that the error term ε_i included two components: a structured component (i.e. spatially correlated) and an unstructured component (i.e. independently distributed). The unstructured heterogeneity term was assumed to be centred and normally distributed while the structured heterogeneity term was assumed to have a conditional intrinsic Gaussian autoregressive (CAR) distribution (Besag et al. 1991) with first- and second-order neighbours as neighbouring structure. The model was fitted in a Bayesian framework with an INLA approach (Blangiardo and Cameletti, 2015) using the R-INLA package (Martins et al. 2013) of the R software (R Core Team, 2015). The fit of the model to the data was assessed using posterior predictive checking (Gelman et al., 2004), and we systematically checked that residuals were not overly structured in space. Bayesian inference resulted in posterior densities for the parameters α_1 (favourable habitat) and α_2 (unfavourable habitat). As we used a genetic distance, we expect a high posterior probability for the intercept of the favourable habitat (α_1) to be less than the intercept of the unfavourable habitat (α_2). The significance was assessed with a threshold of 0.05.

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

Biological datasets

To illustrate MAPI behaviour in the presence of IBD, we analysed microsatellite genotypes from a rodent population that was previously characterised as a single genetic unit, only structured by IBD (Gauffre et al. 2008), using the clustering method GENELAND (Guillot et al. 2005). MAPI efficiency in detecting genetic discontinuities is further illustrated using DNA sequences from different strains of plant virus. These data were previously analysed (Desbiez et al. 2009; Joannon et al. 2010) using phylogenetic analyses, the clustering method SAMOVA (Dupanloup et al. 2002) and the maximum difference Monmonier's algorithm (Monmonier 1973). Finally, MAPI was applied to genetic distances computed from microsatellite genotypes of two species of forest ground beetles with contrasted level of habitat specialisation (i.e. specialist and generalist). Previous work using Mantel tests between genetic, geographic and landscape distances showed that open field areas were a stronger barrier to gene flow for the forest specialist (Brouat et al. 2003). See section 3 in the Appendix for further details on the datasets and related works.

RESULTS

All simulation setups and MAPI graphical outputs are available online at:

<https://www1.montpellier.inra.fr/CBGP/software/MAPI/>.

Sensitivity to isolation-by-distance

Simulation test datasets. MAPI did not detect unexpected areas of genetic discontinuity under strict IBD, regardless of its strength (see test datasets 1 in Table 2; Figures 3E, 3I). When models combine IBD and barrier effects, MAPI still did not detect unexpected areas of genetic discontinuity, i.e. which did not overlap with the simulated barrier (see test datasets 3 in Table 2; Figures 3F-H, 3J-L). Under

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

large regular sampling, the average percentage over replicates of unexpected discontinuous areas was less than 0.15% whatever the simulation setting considered (see 'Unexpected' in Table 2), with a maximal percentage across the 625 simulated datasets of 1.3%. Under small irregular sampling, the percentage of unexpected discontinuous areas was still minute, with an average lower than 0.53% whatever the simulation setting (Table 2) and a maximum across the 625 simulated datasets of 5.8%. Thus, gaps between samples did not drastically increase the proportion of unexpected significant discontinuous areas, even under strong IBD ($g=0.25$, slope = 0.055). The single exception was the combination of strong IBD and genotype datasets of 100 loci, for which the small and irregular sampling led to a more significant (but still relatively low) increase in the percentage of unexpected discontinuous areas, with an average over replicates of 2.75% and a maximum of 10.8% (Table 2). However, the spatial distribution of the unexpected significant cells hardly suggested the presence of a barrier to gene flow (see online illustrations).

Biological dataset. When applied on rodent microsatellite dataset, MAPI did not find any spurious significant area of genetic discontinuity that could be interpreted as a barrier to gene flow despite IBD, i.e. slope=0.005, p-value=0.005 (see section 3.1 in the Appendix).

Detection of a barrier to gene flow

Simulation test datasets. We found that MAPI was efficient in detecting strong to moderate barriers to gene flow, i.e. $N_e m \leq 1$; $F_{ST} \geq 0.1$ with a barrier coverage $\geq 75\%$ (see test datasets 2 in Table 2; Figures 3C-3D). When permeability increased ($N_e m > 1$; $F_{ST} < 0.1$), the method lost its accuracy to detect the barrier, i.e. the proportion of undetected barriers became large (see test datasets 2 in Table 2; Figure 3B). However, our simulations showed that IBD increased the power of the method to detect the linear barrier, especially for high levels of gene flow. For example, for $N_e m = 2.5$, even

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

weak IBD ($g=0.750$; slope= 0.005) increased the barrier coverage from 7% to 79%. As a result, MAPI performed well in identifying barriers to gene flow given that $N_e m < 10$ (see test datasets 3 in Table 2; Figure 3F-3H and 3J-3L). When individual samplings were small and irregular ($N=75$), the performance of MAPI to detect a weak barrier to gene flow ($N_e m > 1$) decreased by 30 to 70% (Table 2), though in these situations there was still often a graphical signal for a barrier (see Figure S3 in the Appendix). As expected, increasing genotyping effort improved further the detection of barriers for high levels of gene flow (Table 2). For example, for $N_e m = 2.5$, increasing the number of loci from 10 to 100 increased the barrier coverage from 7% to 98% in absence of IBD and for $N_e m = 10$, 30% of the barrier was still recovered.

Biological dataset. In line with previous results published on the plant virus sequence dataset, MAPI identified a major area of genetic discontinuity bisecting the study area from North to South (see section 3.2 in the Appendix).

Sensitivity to parameter settings. The eccentricity and number of cells have little effect on the detection of unexpected significant areas (i.e. Unexpected in Figure 5). The worst situation (i.e. 3%) occurred when using very narrow ellipses ($e=0.999$; see Figure S1 of the Appendix) combined with a very high number of cells (i.e. $\theta = 0.15$). The combination of narrow ellipses ($e=0.999$), which result in a lower number of ellipses intercepting each cell, and a very low number of cells (i.e. $\theta = 0.75$) decreased MAPI efficiency to detect a barrier to gene flow (i.e. Coverage in Figure 5; Figure S3 in the Appendix). An eccentricity value of 0.975 associated with a θ value of 0.5 for regular sampling or 0.25 for irregular sampling ensures high barrier coverage, low rate of false positives and reasonable computational time. Lower values of eccentricity (0.8-0.95; inflated ellipses) provide same quality results as a value of 0.975 but at the cost of expended computational time and stronger smoothing effect (see broader aggregates of significant cells for a value of 0.9 in the Figure S3 in the Appendix).

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

The effect of the eccentricity parameter is further illustrated on virus data (Figure S8 in the Appendix). In all cases, the major area of genetic discontinuity was still uncovered by MAPI.

Assessment of landscape effects

Simulation test datasets. As expected, when the two habitats were distributed randomly, there was no obvious correspondence between the spatial variation in m_w and landscape cell values (Figure 4; see also Figure S4 in the Appendix to visualise examples of MAPI graphical outputs). This was confirmed by the results of the Bayesian conditional autoregressive model which showed that most of the posterior probabilities for a habitat effect were not significant (Table 3 and Figure S5 in the Appendix). A difference between habitats was, however, still detected from a few simulated test datasets, especially when using random sampling. For the gradient and fragmented landscapes, MAPI graphical outputs showed a relative convergence between variation in m_w cell values and spatial habitat configuration (Figure 4 and Figure S4 in the Appendix). Accordingly, the intercept was significantly less for the favourable habitat (α_1) than for the unfavourable habitat (α_2) for all datasets regardless of the sampling scheme and size for the fragmented landscape and for most of the datasets when large (500 genotypes) and random or gridded samplings were used for the gradient landscape (Table 3). When the sampling size was smaller (200 genotypes) the proportion of significant posterior probabilities for $\alpha_1 < \alpha_2$ decreased of 25 and 50% for the random and gridded samplings, respectively. Interestingly, under the gradient configuration, the balanced sampling led to very poor results and even, in a few cases, to false positives (i.e. $\alpha_1 < \alpha_2$) (Table 3 and Figure S5 in the Appendix).

Biological dataset. On ground beetle microsatellite dataset, MAPI analyses supported previously published conclusions by identifying a significant area of genetic discontinuity corresponding to a large open field for the forest specialist only (Figure 6). When using m_w cell values in the Bayesian conditional autoregressive model with the proportion of trees as an explanatory variable, we found a significant negative relationship only for the forest specialist (posterior probability=1). This means that, on average, between-individual genetic distances were lower within highly forested areas. No significant pattern was found for the generalist (see section 3.3 in the Appendix).

DISCUSSION

In this work, we presented a new method to translate networks of pairwise relationships into variation surfaces. MAPI is essentially a smoothing procedure using the overlap between ellipses as a way to share information between spatial connections. The surface produced is a grid of cells that provide information on the average intensity of the pairwise relationships crossing at the cell locations. The variation of the cell values over the surface allows to localise areas where the majority of the crossing connections correspond to very low or very high pairwise values. The significance of these areas can be assessed using a nonparametric randomisation procedure.

MAPI can be applied to genetic data to detect areas of high genetic continuity and discontinuity. When using neutral markers (e.g. microsatellites) and genetic differentiation measures, continuous and discontinuous areas reflect areas where gene flow is the highest and the lowest, respectively. Variation in gene flow intensity can result from different processes such as spatial heterogeneity in population density or migration success (Richardson et al. 2016). Here, using controlled simulations, we determined that MAPI performed well to detect genetic discontinuities resulting from a physical barrier as long as it is not too permeable to gene flow. When gene flow was

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

high, MAPI's performance was substantially improved with sampling effort in the number of loci and individuals. Under such spatial genetics models, the sensitivity analysis to eccentricity setting showed that a value of 0.975 ensured high barrier coverage, low rate of unexpected significant cells, high spatial resolution and reasonable computational time. This result might, however, differ in more complex scenarios than a linear barrier to gene flow (e.g. highly fragmented landscape).

A central feature of MAPI is its relative insensitivity to IBD, which is a critical issue for spatial and landscape genetics analyses (Guillot et al. 2009; Thomassen et al. 2010; Bradburd et al. 2013). In an ideal situation, with perfect regular sampling and no edge effects, all cells should theoretically display the same m_w value under strict IBD. When analysing highly irregular samplings, the method detected a few unexpected significant areas of genetic discontinuity as the cells located within the spatial gaps were only informed from long-distance connections. These highly discontinuous areas were, however, small and generally located on the border of the study area. As for many spatial exploratory analyses, regular individual-based samplings are more likely to provide reliable and interpretable results (Oyler-McCance et al. 2013; Balkenhol et al. 2015).

In landscape genetics, assessing the effects of environmental features on spatial patterns of genetic variation often requires going beyond the detection of barriers to gene flow. In this context, our controlled simulations and ground beetle datasets (Brouat et al. 2003) showed that overlaying MAPI graphical outputs on landscape layers may provide information on which environmental variables are potential candidates to explain observed genetic patterns (see Figure 4 and Figure S9 in the Appendix). These candidate variables can be further explored by using m_w cell values in post-MAPI analyses. To illustrate this possibility we used a regression model accounting for spatial autocorrelation in m_w cell values and fitted in a Bayesian framework. From our simulated datasets, the regression model successfully retrieved the expected relationships (i.e. smaller m_w values within the favourable habitat) for the fragmented and gradient landscapes but not when the two habitats

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

were randomly distributed. These results can be explained by the difference in the scale at which habitats are spatially autocorrelated in the three simulated landscapes. As long distance effects are limited in the computation of m_w by using an inverse-area-weighting procedure, the cells mainly reflect the average intensity of the shorter connections crossing at one location. Consequently, significant relationships between m_w cell values and landscape variables are expected to be detectable when the landscape is spatially structured in such a way that a cell is mostly influenced by short-distance connections occurring between samples located within a relatively homogeneous landscape aggregate. Contrarily, when the spatial scale of autocorrelation in landscape features is smaller than the resolution of the sampling, a cell reflects between- as much as within-habitat connections and habitat effect are likely to become undetectable (e.g. random landscape). In the datasets simulated under a fragmented landscape model, sampling size and strategy did not affect the detection of habitat effect. Under the gradient configuration, the habitat effect was well detected especially when a large number (i.e. 500) of genotypes were sampled following a random, or even better, a gridded sampling strategy. Contrarily, the balanced sampling strategy was inefficient to detect an effect. This result can be due to the multi-scale patterns of autocorrelation of the habitats that have different consequences. First, as the two habitats strongly interpenetrate, the central part of the study area looks like the random landscape model (i.e. small habitat patches). Within this zone, the balanced sampling produces numerous between-habitats connections at short distance that blur the relationship between the m_w cell values and habitat type. Second, over the whole study area, the scale at which the habitats are autocorrelated is quite large and, subsequently, autocorrelation in m_w cell values can be expected to be large as well. In such a situation, increasing the neighbouring structure from two to five cells in the conditional autoregressive model significantly improves the detection of the habitat effect from the balanced sampling (i.e. mean posterior probability = 0.909 and 0.927; percentage of significant probability = 60 and 75 for the datasets with 200 and 500 genotypes, respectively).

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

Analysis of MAPI cell values conjointly with landscape variables is a data-driven exploratory approach that can help identify candidate landscape features. Adequate sampling schemes and analyses aiming to test for the effects of those candidate variables should be conducted at an appropriate spatial scale before drawing conclusions (Richardson et al. 2016).

ACKNOWLEDGEMENTS

The authors thank T. Jombart, G. Guillot, N. Verzelen, R Leblois and A. Estoup for helpful comments on previous versions of the manuscript as well as the associate editor and four referees for their very constructive remarks. The authors also thank C. Desbiez and C. Brouat for providing the virus and forest ground beetle data, respectively. K. Berthier acknowledges the INRA Metaprogram “Sustainable Management of Crop Health” for funding the COPACABANA project. This work is dedicated to late Serge Meusnier.

DATA ACCESSIBILITY

- Rodent microsatellite data: DRYAD entry doi.org/10.5061/dryad.jf7sn,
- Plant virus sequence data: Genebank accession numbers EU660581-EU660590, HM044202-HM044215,
- Ground beetle microsatellite data: Figshare entry https://figshare.com/articles/GroundBeetles_zip/3482804 and French National Institute for Agricultural Research, <https://www1.montpellier.inra.fr/CBGP/software/MAPI/> (file: MAPI_examples.zip).
- Simulated test datasets and R scripts: French National Institute for Agricultural Research, <https://www1.montpellier.inra.fr/CBGP/software/MAPI/> (file: MAPI_examples.zip).

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

REFERENCES

- Adams, R., & Burg, T. (2015). Influence of ecological and geological features on rangewide patterns of genetic structure in a widespread passerine. *Heredity*, **114**(2), 143–154.
- Balkenhol, N., Cushman, S., Storfer, A., & Waits, L. (2015). *Landscape Genetics: Concepts, Methods, Applications*. John Wiley & Sons.
- Barthélemy, M. (2011). Spatial Networks. *Physics Reports*, **499**, 1–101.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the False Discovery Rate in multiple testing under dependency. *Annals of Statistics*, **29**(4), 1165–1188.
- Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**(1), 1–20.
- Blangiardo, M., & Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons.
- Borgatti, S., Mehra, A., Brass, D., & Labianca, G. (2009). Network analysis in the social science. *Science*, **323**, 892–895.
- Botta, F., Eriksen, C., Fontaine, M. C., & Guillot, G. (2015). Enhanced computational methods for quantifying the effect of geographic and environmental isolation on genetic differentiation. *Methods in Ecology and Evolution*, **6**(11), 1270–1277.
- Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, **67**(11), 3258–3273.
- Broquet, T., Ray, N., Petit, E., Fryxell, J. M., & Burel, F. (2006). Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). *Landscape Ecology*, **21**(6), 877–889.
- Brouat, C., Sennedot, F., Audiot, P., Leblois, R., & Rasplus, J.-Y. (2003). Fine-scale genetic structure of two carabid species with contrasted levels of habitat specialization. *Molecular Ecology*, **12**(7),

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

1731–1745.

- Coulon, A., Cosson, J.-F., Angibault, J.-M., Cargnelutti, B., Galan, M., Morellet, N., Petit, Aulagnier, S. & E., Hewison, A.J. (2004). Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual-based approach. *Molecular Ecology*, **13**, 2841–2850.

- Cushman, S. A., McKelvey, K.S., Hayden, J. & Schwartz, M.K. (2006). Gene Flow in Complex Landscapes: Testing Multiple Hypotheses with Causal Modeling. *The American Naturalist*, **168**(4), 486–499.

- Cushman, S. A., & Landguth, E. L. (2010). Spurious correlations and inference in landscape genetics. *Molecular Ecology*, **19**(17), 3592–3602.

- Cushman, S. A., Wasserman, T. N., Landguth, E. L., & Shirk, A. J. (2013). Re-evaluating causal modeling with Mantel tests in landscape genetics. *Diversity*, **5**(1), 51–72.

-Desbiez, C., Joannon, B., Wipf-Scheibel, C., Chandeysson, C., & Lecoq, H. (2009). Emergence of new strains of Watermelon mosaic virus in South-eastern France: Evidence for limited spread but rapid local population shift. *Virus research*, **141**, 201–208

- Duforet-Frebourg, N., & Blum, M. G. (2014). Nonstationary patterns of isolation-by-distance: Inferring measures of local genetic differentiation with bayesian kriging. *Evolution*, **68**(4), 1110–1123.

Dupanloup, I., Schneider, S., & Excoffier, L. (2002). A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, **11**(12), 2571–2581.

- Dyer, R. J. (2015). Population Graphs and Landscape Genetics. *Annual Review of Ecology, Evolution, and Systematics*, **46**, 327–342.

- Dyer, R. J., & Nason, J. D. (2004). Population graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, **13**(7), 1713–1727.

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, **7** (12), 1463–1475. , DOI : 10.1111/2041-210X.12616

- Etherington, T. R. (2011). Python based GIS tools for landscape genetics: visualising genetic relatedness and measuring landscape connectivity. *Methods in Ecology and Evolution*, **2(1)**, 52–55.
- Galpern, P., Peres-Neto, P. R., Polfus, J., & Manseau, M. (2014). MEMGENE: Spatial pattern detection in genetic distance data. *Methods in Ecology and Evolution*, **5(10)**, 1116–1120.
- Garroway, C. J., Bowman, J., Carr, D., & Wilson, P. J. (2008). Applications of graph theory to landscape genetics. *Evolutionary Applications*, **1(4)**, 620–630.
- Gauffre, B., Estoup, A., Bretagnolle, V., and Cosson, J. F. (2008). Spatial genetic structure of a small rodent in a heterogeneous landscape. *Molecular Ecology*, 17:4619-4629.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Taylor & Francis.
- Guillot, G., Leblois, R., Coulon, A., & Frantz, A. C. (2009). Statistical methods in spatial genetics. *Molecular Ecology*, **18(23)**, 4734–4756.
- Guillot, G., Mortier, F., & Estoup, A. (2005). Geneland: a computer package for landscape genetics. *Molecular Ecology Notes*, **5(3)**, 712–715.
- Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution*, **4(4)**, 336–344.
- Hagerty, B. E., Nussear, K. E., Esque, T. C., & Tracy, C. R. (2011). Making molehills out of mountains: landscape genetics of the Mojave desert tortoise. *Landscape Ecology*, **26(2)**, 267–280.
- Hengl, T. (2006). Finding the right pixel size. *Computers & Geosciences*, **32(9)**, 1283–1298.
- Henneman, S. (2013). Information-rich visualisation of dense geographical networks. *Journal of Maps*, **9**, 68–75.
- Joannon, B., Lavigne, C., Lecoq, H., & Desbiez, C. (2010). Barriers to gene flow between emerging populations of Watermelon mosaic virus in Southeastern France. *Phytopathology*, **100(12)**:1373-1379.

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

- Kelling, S., Hochachka, W. M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., & Hooker, G. (2009). Data-intensive science: a new paradigm for biodiversity studies. *BioScience*, **59**(7), 613–620.
- Legendre, P., & Fortin, M.-J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**(5), 831–844.
- Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis*, **67**, 68–83.
- McRae, B. H. (2006). Isolation by resistance. *Evolution*, **60**(8), 1551–1561.
- McRae, B. H., & Beier, P. (2007). Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences*, **104**(50), 19885–19890.
- Miele, V., Picard, F., & Dray, S. (2014). Spatially constrained clustering of ecological networks. *Methods in Ecology and Evolution*, **5**(8), 771–779.
- Miller, M. (2005). Alleles In Space (AIS): Computer software for the joint analysis of interindividual spatial and genetic information. *Journal of Heredity*, **96**, 722–724.
- Monmonier, M. S. (1973). Maximum-Difference Barriers: An Alternative Numerical Regionalization Method. *Geographical Analysis*, **5**(3), 245–261.
- Oyler-McCance, S. J., Bradley C. F., & Landguth E. L. (2013) Sample design effects in landscape genetics. *Conservation Genetics*, **14**, 275–285.
- Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, **48**(1), 94–100.
- R Core Team. (2015). R: A language and environment for statistical computing. <http://www.R-project.org>.
- Raufaste, N., & Rousset, F. (2001). Are partial mantel tests adequate? *Evolution*, **55**(8), 1703–1705.
- Richardson, J. L., Brady, S. P., Wang, I. J., & Spear, S. F. (2016). Navigating the pitfalls and promise of

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, **7** (12), 1463–1475. , DOI : 10.1111/2041-210X.12616

landscape genetics. *Molecular Ecology*, **25(4)**, 849–863.

- Rousset, F. (2000). Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.

- Thomassen, H. A., Cheviron, Z. A., Freedman, A. H., Harrigan, R. J., Wayne, R. K., & Smith, T. B. (2010). Spatial modelling and landscape-level approaches for visualizing intra-specific variation. *Molecular Ecology*, **19(17)**, 3532–3548.

- Vandergast, A., Perry, W., R.V., L., & Hathaway, S. (2010). Genetic landscapes GIS Toolbox: tools to map patterns of genetic divergence and diversity. *Molecular Ecology Resources*, **11**, 158–161.

- Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, **23(23)**, 5649–5662.

- Weckworth, B. V., Musiani, M., DeCesare, N. J., McDevitt, A. D., Hebblewhite, M., & Mariani, S. (2013). Preferred habitat and effective population size drive landscape genetic patterns in an endangered species. *Proceedings of the Royal Society of London B: Biological Sciences*, **280(1769)**, 20131756.

- Wood, DA, AG Vandergast, KR Barr, RD Inman, TC Esque, KE Nussear, and RN Fisher. 2012. Comparative phylogeography reveals deep lineages and regional evolutionary hotspots in the Mojave and Sonoran Deserts. *Diversity and Distributions*, **19**, 722-737.

- Wright, S. (1943). Isolation by distance. *Genetics*, **28**, 114–138.

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, **7** (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

Table 1 - Setting of MAPI parameter values.

Step	Parameter	Definition	Hints	Default	Simulated test data	
					Spatial framework	Landscape framework
<i>Grid of cells</i>	β	Defines the spatial resolution of the grid of cells	0.50 for regular sampling 0.25 for irregular sampling	NA	$\Rightarrow$ set to 0.5 for large regular samplings (200 samples); convert to 855 cells $\Rightarrow$ set to 0.25 for small irregular samplings (75 samples); convert to 843 cells on average	$\Rightarrow$ not used Importation of a grid with the resolution of the landscape raster pixels*
<i>Network of ellipses</i>	<i>Eccentricity</i>	Ellipse eccentricity	0.975 is a good starting point Run several independent analyses from 0.800 (inflated) to 0.999 (narrow) to assess pattern robustness	0.975	$\Rightarrow$ set to 0.975	$\Rightarrow$ set to 0.975
	<i>error_circle_radius</i>	Error circle on sample locations	Higher values for larger uncertainty on sample positions	10 map units	No error $\Rightarrow$ set to 0.01	Error $\Rightarrow$ set to 0.50
	<i>min_distance</i> <i>max_distance</i> (optional)	Minimum and maximum distances between	Start without filtering Filter on minimum distance to avoid local effects (see an illustration in section 3.1 in the	NA	$\Rightarrow$ not used	$\Rightarrow$ minimum distance set to 1
<i>Test for spatial structure</i>	<i>n_permutations</i>	Number of permutations (n_p)	$\geq 1,000$	1,000	$\Rightarrow$ set to 1,000	$\Rightarrow$ not used
	<i>my_alpha</i>	Significance level	$\alpha = 0.05$	0.05	$\Rightarrow$ set to $\alpha = 0.05$	$\Rightarrow$ not used

* When a grid is imported, the cells can be of any shape (e.g. hexagons, squares, etc.)

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

Table 2. Performance of MAPI under different spatial models. Parameters for each spatial model are indicated, including: the product of the effective population size and migration rate per generation ($N_e m$); the shape parameter of the geometric distribution (g); the number of simulated populations ($Ksim$); and the number of loci ($Nloc$). Performance is measured as the average percentage, over 25 replicates, of the barrier covered by significant discontinuous areas (Coverage), and the average percentage, over 25 replicates, of significant discontinuous areas having no contact with the barrier (Unexpected). Large sampling refers to the sampling of 200 genotypes distributed regularly on the lattice, and small sampling to the sampling of 75 genotypes drawn randomly from the lattice. The datasets with 100 loci are indicated in italics.

Spatial Model	$N_e m$	g	$Ksim$	$Nloc$	Coverage (%)		Unexpected (%)	
					Large sampling	Small sampling	Large sampling	Small sampling
Panmixia (test dataset)	NA	NA	1	10	NA	NA	0.10	0.05
<i>IBD</i> (test datasets 1)	NA	0.250	1	10	NA	NA	0.07	0.53
	NA	0.250	1	100	NA	NA	0.00	2.75
	NA	0.500	1	10	NA	NA	0.06	0.23
	NA	0.500	1	100	NA	NA	0.04	0.70
	NA	0.675	1	10	NA	NA	0.06	0.06
	NA	0.750	1	10	NA	NA	0.06	0.26
<i>Barrier</i> (test datasets 2)	0.1	NA	2	10	99	84	0.03	0.21
	1	NA	2	10	65	31	0.08	0.35
	2.5	NA	2	10	7	2	0.15	0.27
	2.5	NA	2	100	98	85	0.04	0.06
	10	NA	2	10	0	0	0.12	0.11
	10	NA	2	100	30	13	0.09	0.28
<i>Barrier & IBD</i> (test datasets 3)	0.1	0.250	2	10	100	90	0.04	0.29
	1	0.250	2	10	100	81	0.07	0.24
	2.5	0.250	2	10	90	66	0.04	0.61
	10	0.250	2	10	41	21	0.08	0.49
	0.1	0.500	2	10	100	92	0.02	0.21
	1	0.500	2	10	96	79	0.01	0.36
	2.5	0.500	2	10	92	62	0.07	0.21
	10	0.500	2	10	44	26	0.10	0.21
	10	0.500	2	100	99	82	0.05	0.31
	0.1	0.675	2	10	100	94	0.02	0.23
	1	0.675	2	10	96	74	0.02	0.35
	2.5	0.675	2	10	85	53	0.08	0.21
	2.5	0.675	2	100	100	96	0.03	0.36
	10	0.675	2	10	23	7	0.13	0.21
	0.1	0.750	2	10	99	94	0.03	0.15

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

	1	0.750	2	10	93	65	0.05	0.27
	2.5	0.750	2	10	79	47	0.07	0.24
	10	0.750	2	10	15	6	0.02	0.12

Table 3. Performance of MAPI under different landscape models. For each simulated dataset and sampling scheme, we extracted cell-specific m_w values and habitat type (1 or 2) to fit the following regression model: $m_{wi} = \alpha_{hab(i)} + \varepsilon_i$, where $\alpha_{hab(i)}$ is an intercept depending on the habitat type in cell i and error term ε_i includes a structured and an unstructured component. Here are presented the posterior probability that $\alpha_1 < \alpha_2$ ($P(\alpha_1 < \alpha_2)$) averaged over 20 replicates ($\bar{P}$) as well as the proportion of $P(\alpha_1 < \alpha_2) \geq 0.95$ ($P_{95\%}$) and ≤ 0.05 ($P_{5\%}$). Note that for all simulated test datasets, the regression model was well adjusted with a mean predictive p-value very close or equal to 0.5 and residuals not overly structured in space (see landscape genetics illustrations on the website).

Landscape model	Sampling scheme	$\bar{P}$		$P_{95\%}$		$P_{5\%}$	
		Large sampling	Small sampling	Large sampling	Small sampling	Large sampling	Small sampling
<i>Gradient</i> (test datasets 1)	Random	0.945	0.883	80	55	0	0
	Balanced	0.614	0.709	10	15	5	5
	Gridded	0.972	0.862	90	40	0	0
<i>Fragmentation</i> (test datasets 2)	Random	0.999	0.999	100	100	0	0
	Balanced	0.999	0.998	100	100	0	0
	Gridded	0.999	0.999	100	100	0	0
<i>Random</i> (test datasets 3)	Random	0.762	0.741	40	25	0	0
	Balanced	0.655	0.546	15	15	0	5
	Gridded	0.780	0.639	10	15	0	0

Figure legends.

Figure 1. Schematic view of MAPI components.

Figure 2. Steps in the test for spatial structure. Computation of m_w values for the observed (A1, A2) and n_p permuted datasets (B1, B2); Ranking of the observed m_w value against the cumulative null distribution for each cell (C1, example for one cell represented by a red hexagon on A2, B2 outputs); Lower-tail (LT) and upper-tail (UT) p-values of the cells (C2, significant LT and UT p-values after FDR corrections are delineated by thin and large black contours, respectively); Visualisation of the final

This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616

output with raw m_w values and significant areas (D).

Figure 3. Subset of MAPI graphical outputs from datasets simulated under panmixia (A), IBD (E, I), barrier to gene flow between panmictic (B-D) and IBD (F-H, J-L) populations. Significant continuous and discontinuous areas are denoted by thin and large black contours, respectively. The white line localise the simulated barrier. NB: colour scale varies.

Figure 4. Subset of MAPI graphical outputs for the gradient, fragmented and random landscape configurations and three sampling schemes (random: 500 genotypes sampled anywhere, balanced: 250 genotypes sampled in each habitat and gridded: 56 genotypes sampled in each cell of a 3x3 lattice (not shown) covering the study area). NB: colour scale varies.

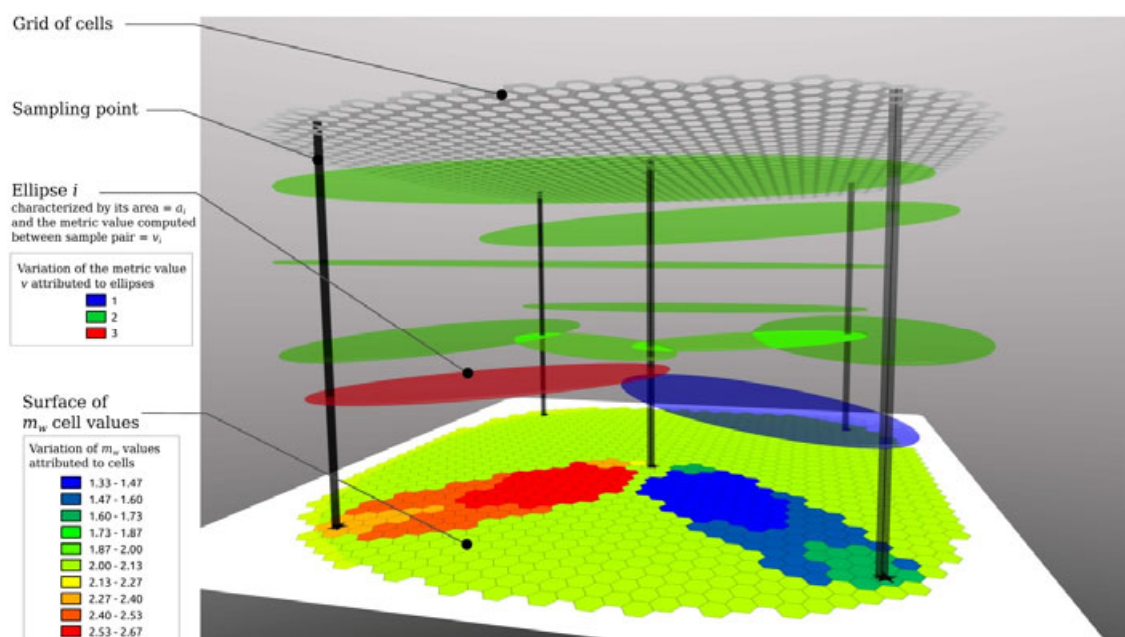
Figure 5. Interactions between *eccentricity* and θ parameter values on MAPI performance. For the different values of θ , the y-axis shows, when applicable, the percentage of the simulated barrier covered by significant discontinuous areas (i.e. Coverage, left panel) or the percentage of significant discontinuous areas having no contact with the barrier (Unexpected, right panel). Eccentricity is on the x-axis. We considered large regular sampling schemes (200 samples, top row) and small irregular sampling schemes (75 samples, bottom row). We used a subset of ten datasets simulated under spatial genetics models (see the section 2.1 in the Appendix to identify test datasets).

Figure 6. MAPI graphical output for the forest specialist species of ground beetle. Cell-specific m_w values appear as a colour scale. The hatched area corresponds to the location of the significant area of genetic discontinuity. Variation in tree coverage follows a greyscale from a minimum of 11% to a maximum of 75%. Sampling locations are illustrated by white circles proportional to sampling sizes.

This article is protected by copyright. All rights reserved.

Comment citer ce document :

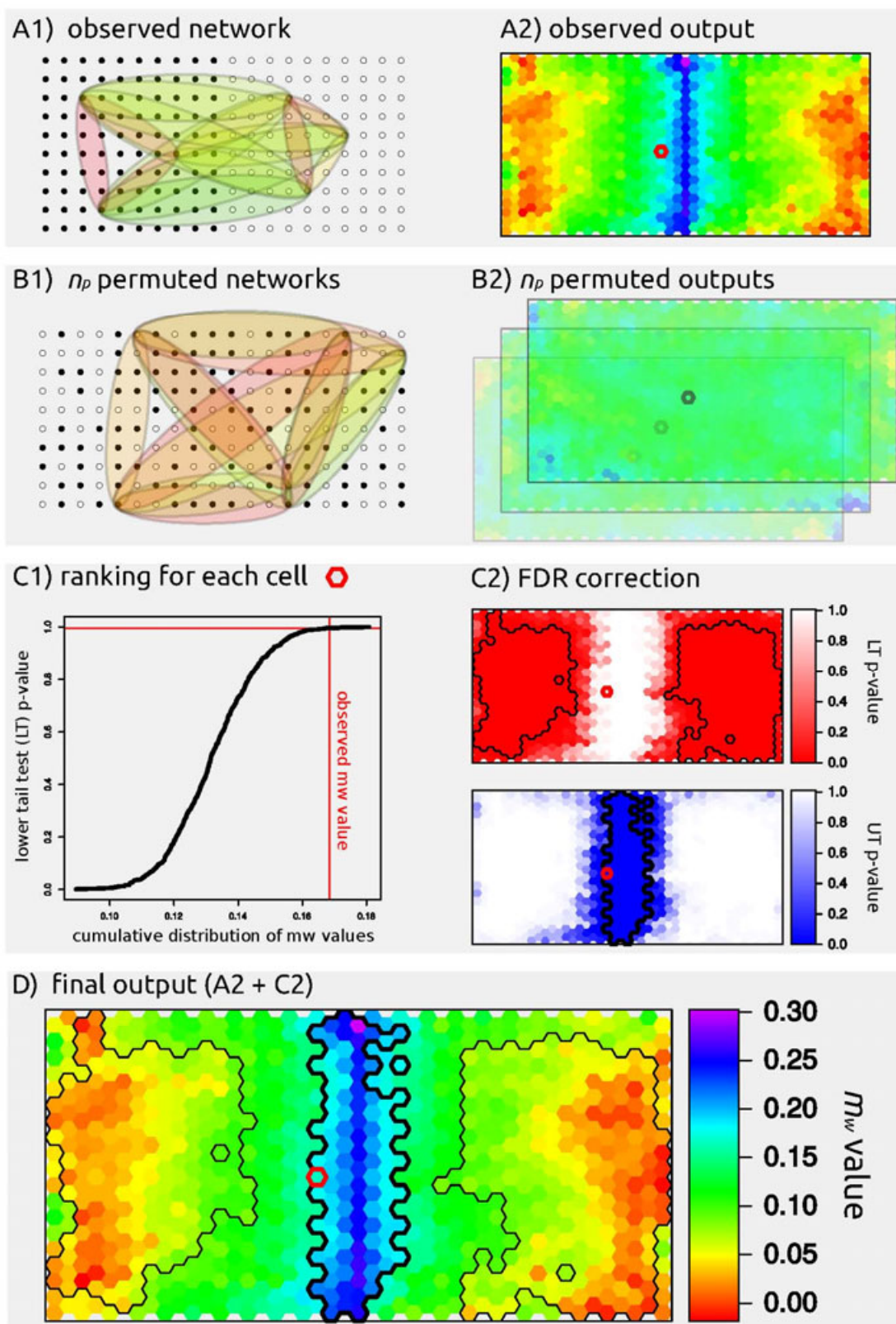
Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616



This article is protected by copyright. All rights reserved.

Comment citer ce document :

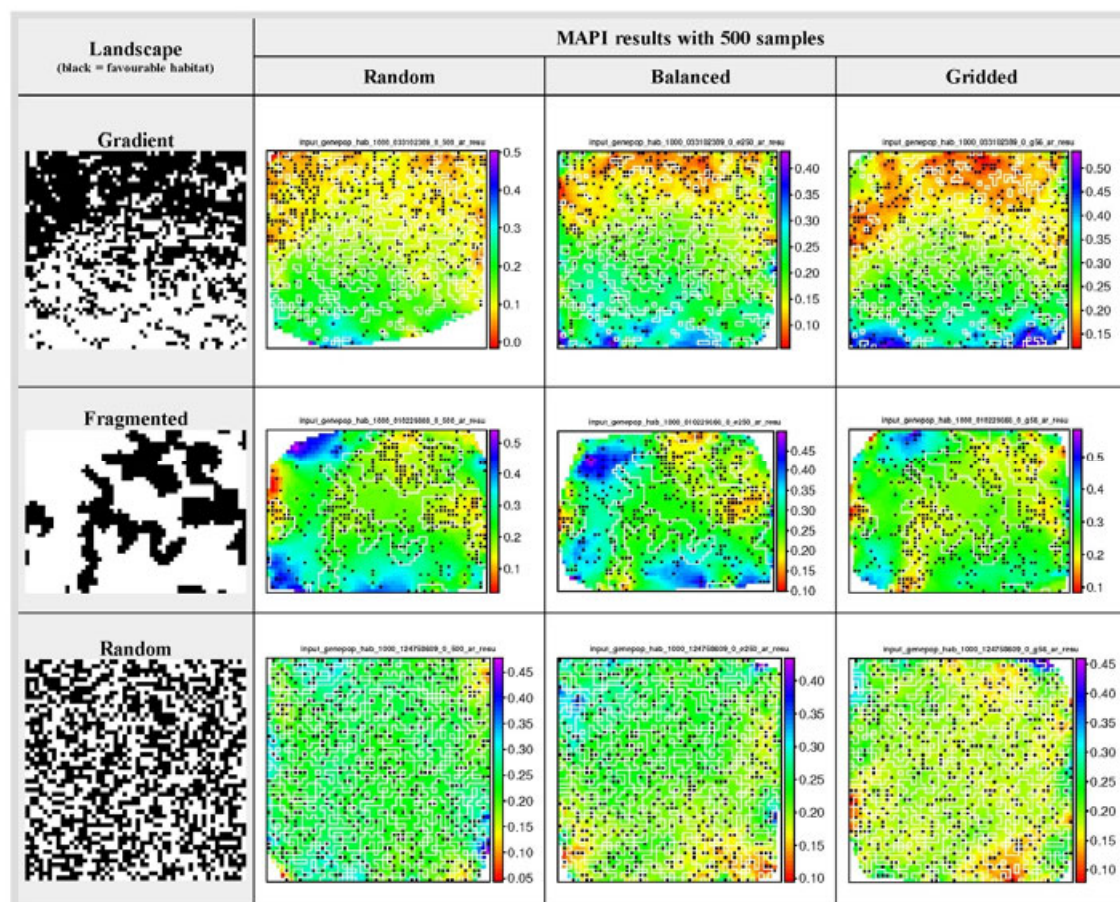
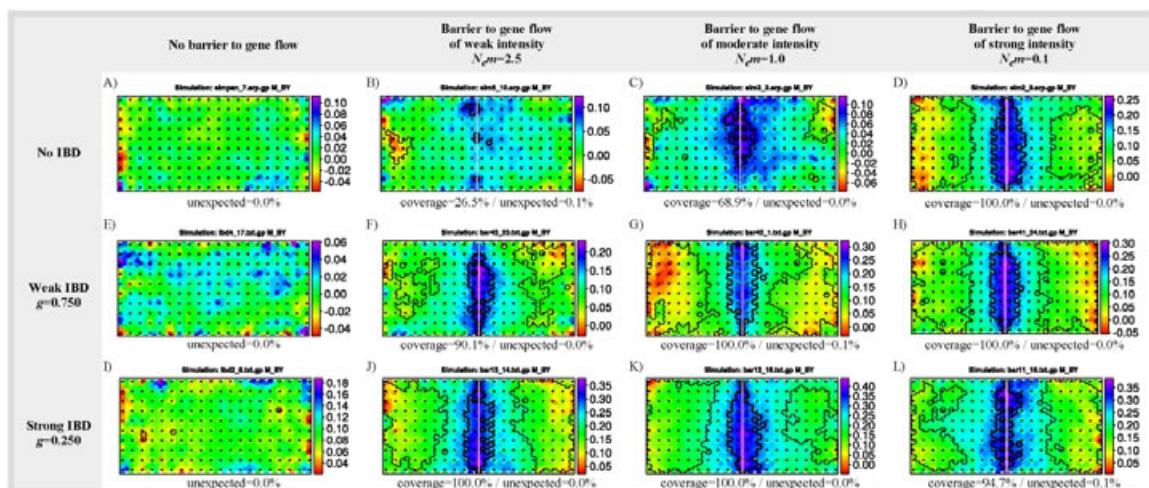
Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616



This article is protected by copyright. All rights reserved.

Comment citer ce document :

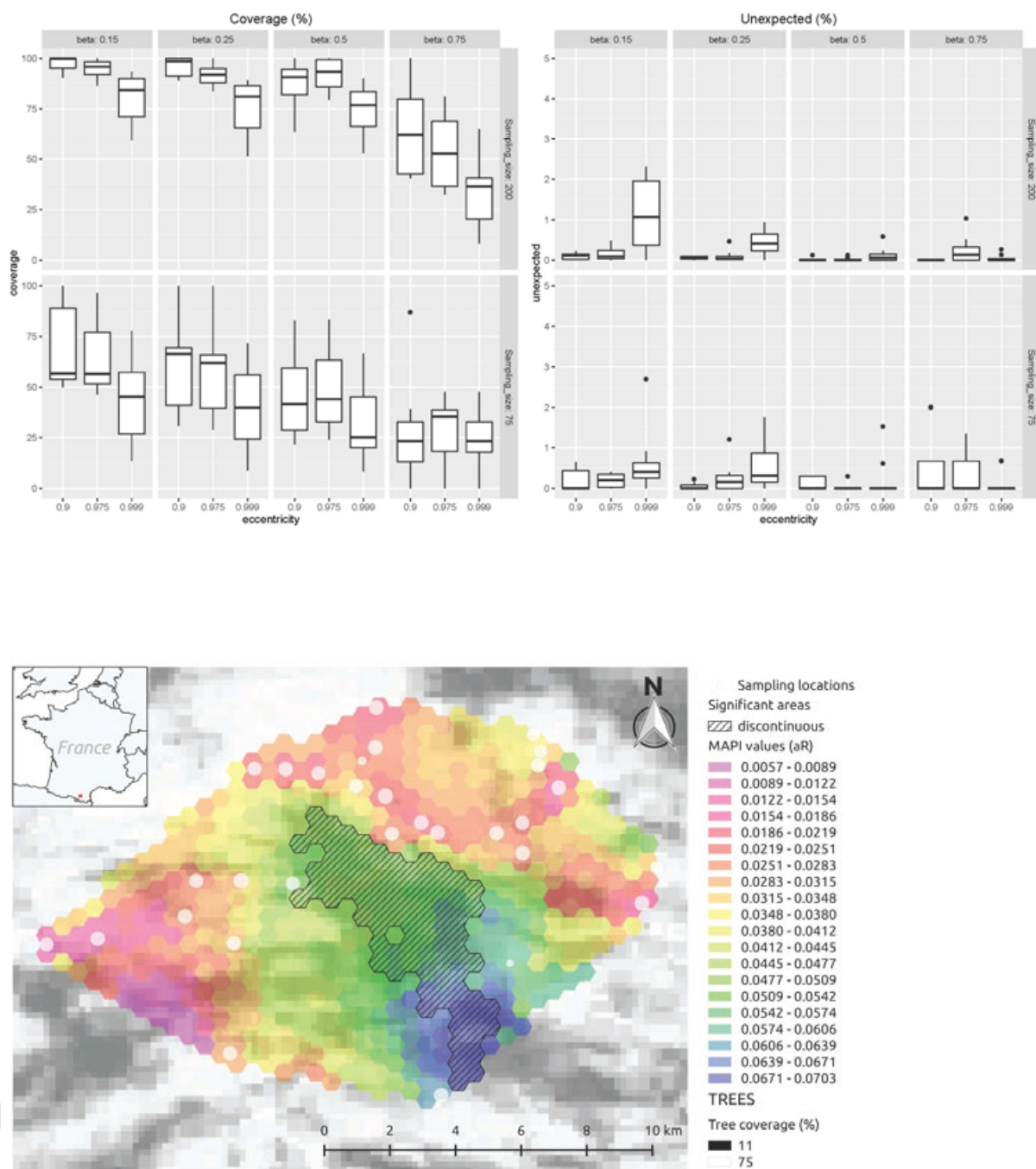
Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616



This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616



This article is protected by copyright. All rights reserved.

Comment citer ce document :

Piry, S. (Auteur de correspondance), Chapuis, M. P., Gauffre, B., Papaix, J., Cruaud, A., Berthier, K. (2016). Mapping averaged pairwise information (MAPI): a new exploratory tool to uncover spatial structure. *Methods in Ecology and Evolution*, 7 (12), 1463-1475. , DOI : 10.1111/2041-210X.12616