Multi-scale and multi-site resampling of a study area in spatial genetics: implications for flying insect species

Julien M Haran [Corresp., 1, 2] , Jean-Pierre Rossi [3] , Juan Pajares [4] , Luis Bonifacio [5] , Pedro Naves [5] , Alain Roques [1] , Géraldine Roux [1]

1 UR633 Zoologie Forestière, INRA, Orléans, France

2 CBGP, CIRAD, Montpellier SupAgro, INRA, IRD, Univ. Montpellier, Montpellier, France CIRAD, CBGP, Montpellier, France

3 CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Univ. Montpellier, INRA, Montpellier, France

4 Sustainable Forest Management Res Inst, Universidad de Valladolid, Palencia, Spain

5 Instituto Nacional de Investigacao Agraria e Veterinaria, INIAV, Oeiras, Portugal

Corresponding Author: Julien M Haran
Email address: julien.haran@gmail.com

The use of multiple sampling areas in landscape genetic analysis has been recognized as a useful way of generalizing the patterns of environmental effects on organism gene flow. It reduces the variability in inference which can be substantially affected by the scale of the study area and its geographic location. However, empirical landscape genetic studies rarely consider multiple sampling areas due to the sampling effort required. In this study, we explored the effects of environmental features on the gene flow of a flying long-horned beetle (*Monochamus galloprovincialis*) using a landscape genetics approach. To account for the unknown scale of gene flow and the multiple local confounding effects of evolutionary history and landscape changes on inference, we developed a way of resampling study areas on multiple scales and in multiple locations (sliding windows) in a single large-scale sampling design. Landscape analyses were conducted in $3*10^4$ study areas ranging in scale from 220 to 1000 km and spread over 132 locations in the Iberian Peninsula. The resampling approach made it possible to identify the features affecting the gene flow of this species but also showed high variability in inference among the scales and the locations tested, independent of the variation in environmental features. This method provides an opportunity to explore the effects of environmental features on organism gene flow on the whole and reach conclusions about general landscape effects on their dispersal, while limiting the sampling effort to a reasonable level.

1 **Multi-scale and multi-site resampling of a study area in spatial**

2 **genetics: implications for flying insect species**

3 JULIEN HARAN[1,2,7], JEAN-PIERRE ROSSI[3], JUAN A. PAJARES[4], LUIS BONIFACIO[5],

4 PEDRO NAVES[5], ALAIN ROQUES[1] and GÉRALDINE ROUX[1,2]

5 [1] INRA, UR633 Zoologie Forestière, F-45075 Orléans, France.

6 [2] Université d'Orléans, rue de Chartres, Orléans cedex, France.

7 [3] CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Univ. Montpellier, Montpellier, France.

8 [4] Univ Valladolid INIA Palencia, Sustainable Forest Management Res Inst, Palencia, Spain.

9 [5] Instituto Nacional de Investigacao Agraria e Veterinaria – INIAV Oeiras, Portugal.

10 [7] Present address: CBGP, CIRAD, Montpellier SupAgro, INRA, IRD, Univ. Montpellier,
11 Montpellier, France CIRAD, CBGP, Montpellier, France

13 **Corresponding author**: Julien HARAN (CBGP, CIRAD, Montpellier SupAgro, INRA, IRD,

14 Univ. Montpellier, Montpellier, France CIRAD, CBGP, Montpellier, France. Tél.: +33 4 99 62 33

15 04, julien.haran@cirad.fr)

Abstract

The use of multiple sampling areas in landscape genetic analysis has been recognized as a useful way of generalizing the patterns of environmental effects on organism gene flows. It reduces the variability in inference which can be substantially affected by the scale of the study area and its geographic location. However, empirical landscape genetic studies rarely consider multiple sampling areas due to the sampling effort required. In this study, we explored the effects of environmental features on the gene flow of a flying long-horned beetle (*Monochamus galloprovincialis*) using a landscape genetics approach. To account for the unknown scale of gene flow and the multiple local confounding effects of evolutionary history and landscape changes on inference, we developed a way of resampling study areas on multiple scales and in multiple locations (sliding windows) in a single large-scale sampling design. Landscape analyses were conducted in $3*10^4$ study areas ranging in scale from 220 to 1000 km and spread over 132 locations in the Iberian Peninsula. The resampling approach made it possible to identify the features affecting the gene flow of this species but also showed high variability in inference among the scales and the locations tested, independent of the variation in environmental features. This method provides an opportunity to explore the effects of environmental features on organism gene flows on the whole and reach conclusions about general landscape effects on their dispersal, while limiting the sampling effort to a reasonable level.

51

52

53

54

55

56

57

58

Introduction

60   Landscape genetics examines the relationship between landscape and environmental
61   features and genetic structure (Manel et al., 2003; Manel & Holderegger 2013). It enables inference
62   about which environmental features facilitate or hinder gene flow (Zeller et al., 2012), which is a
63   key factor for understanding the persistence and evolution of species and populations and has
64   significant consequences for conservation planning (Castillo et al., 2014; Van Strien et al., 2014).
65   As an emerging and fast moving field, substantial effort to optimize the method is still required in
66   order to make relevant and optimum inferences (Anderson et al., 2010; Cushman et al., 2013;
67   Manel & Holderegger 2013). Landscape genetic analyses are usually conducted on a single scale
68   and in a single location (Zeller et al., 2012) and therefore provide results that are, strictly speaking,
69   only applicable to the particular area under study (Shortbull et al. 2011). In fact, genetic structure
70   is determined by multiple micro- and macro-evolutionary processes acting on different spatial and
71   temporal scales, and is rarely homogeneously distributed across a study species' distribution range
72   (Waters et al., 2013). For example, in addition to contemporary or historical environmental effects
73   on dispersal (Zellmer & Knowles 2009), the genetic structure of organisms is often influenced by
74   historic differentiation due to quaternary climate oscillations (Hewitt 2000), or by biased dispersal
75   due to local adaptation to specific environmental conditions (Sexton et al., 2014; Pflüger &
76   Balkenhol 2014). The diverse factors acting on different temporal and spatial scales may generate
77   genetic patterns that are inconsistent across locations or regions, which results in conflicting
78   signals of environmental factors acting on gene flow. In addition to these multiple limitations, the
79   landscape genetics approach is increasingly being applied to flying organisms with dispersal
80   abilities that are difficult to predict and for which barriers to dispersal may be hard to conceptualize
81   (Zeller et al. 2012). In these situations, an experimental design is difficult to set up a priori, which
82   can lead to limited inference (Dreier et al. 2014).

83   To integrate these factors and deal with the variability of the results in landscape genetic
84   analysis, several authors have highlighted the importance of two aspects: matching the study

85 design to the process being investigated (scale of sampling area, Cushman & Landguth 2010;
86 Manel & Holderegger 2013) and considering landscape level replications (multiple locations of
87 sampling areas, Short Bull et al., 2011). The study scale is fundamental in landscape genetics,
88 because species respond to environmental features on a continual range of scales, which may affect
89 correlation coefficients (Anderson et al., 2010; Manel & Holderegger 2013). This point has been
90 highlighted in several empirical studies and simulation exercises (Cushman & Landguth 2010;
91 Angelone et al., 2011; Galpern et al., 2012; Dudaniec et al., 2013; Keller et al., 2013), in particular
92 for organisms exhibiting wide home-ranges, such as large mammals (Galpern et al., 2012; Zeller
93 et al., 2014). The scale of the study design is also crucial for flying species for which the scale of
94 gene flow may be much larger than expected (Dreier et al. 2014). However, landscape genetic
95 studies still rarely consider this aspect (Zeller et al., 2012) and how it affects inference in the
96 detection of general effects of environmental features on dispersal and gene flow. Landscape-level
97 replication is another fundamental aspect in landscape genetics (Holderegger & Wagner 2008).
98 The term replication usually refers to the replication of sampling areas (sampling units; Short Bull
99 et al., 2011). Such an experimental design provides a "quantitative" dimension in landscape
100 genetic analysis, allowing conclusions to be drawn about the general effects of landscape features
101 on the dispersal of organisms. These replications are particularly useful to account for multiple
102 local confounding effects of the genetic structure of organisms detailed above. Few studies have
103 included replication in landscape genetic studies, and the number of replications considered is
104 often small (Driezen et al., 2007; Kindall & Van Manen 2007; Zalewski et al., 2009; Short Bull et
105 al., 2011).

106   In order to provide a more complete and comprehensive picture of the general effects of
107 the landscape on the dispersal of organisms, there is a need to integrate various scale and landscape
108 level replications of sampling areas in more empirical landscape genetic studies. However, such
109 exploration often remains limited due to the substantial sampling efforts required. Resampling
110 methods (Sjöstedt-de Lunas 2001) are an interesting perspective in landscape genetics, as they
111 offer the possibility of examining variation of inference in several sub-parts of a single sampling
112 design, potentially including variations of scale and location of study areas. In our study, we
113 developed a method combining multi-site and a multi-scale resampling of sliding windows (study
114 areas) to explore on which scale and in which locations environmental features fostered or hindered
115 the gene flow of a flying insect species: *Monochamus galloprovincialis* (Coleoptera,

116   Cerambycidae). We first characterized the broad-scale genetic structure of the beetle across the

117   study area, and we used the "isolation-by-resistance" (IBR) framework to model beetle dispersal

118   as a function of three relevant landscape features (temperature, elevation, and pine cover), called

119   IBR hypothesis (IBR-T, IBR-E and IBR-P respectively). For each of these IBR hypothesis, we

120   then carried out a landscape genetic analysis based in 30 576 resampled areas of extents ranging

121   from 220 to 1000 km and distributed in 132 sampling locations in the Iberian Peninsula. Lastly,

122   we searched for the main landscape features affecting gene flow in *M. galloprovincialis* and we

123   analyzed how the scale and location of the study area impacted inference.

124

Methods:

125

Sampling and genotyping

126

127   The study area covered the entire Iberian Peninsula (582 000 km^2) with altitudes ranging

128   from sea level up to 2444 m. *M. galloprovincialis* specimens were trapped between 2011 and 2013

129   at 137 sites spread over the Iberian Peninsula. We used multifunnel traps baited with a volatile

130   attractant (Galloprotect, SEDQ, Spain) installed during the summer to catch flying adults. The

131   traps used had a radius of attraction of 100 m (Jactel et al., 2015) and were placed in dense pine

132   stands (where beetle density is high; Jactel et al., 2015) to limit consanguinity among the

133   individuals caught. After collecting, adults were stored in 96.66% ethanol at 4°C. Despite intensive

134   trapping, *M. galloprovincialis* was not recorded in the central lowlands of Castilla y Leon, central

135   Galicia and Asturias districts (Fig. S1). We obtained 1050 individuals at 132 sites, with an average

136   sample per location of 7.68 individuals. Therefore, sampling consisted of a trade-off between the

137   number of localities and the size of demes, keeping the sampling effort at a realistic level. Details

138   of sampling localities and year of collection are given in table S1. Individuals collected at the same

139   locality were considered as one deme. The distribution of sites covered most of the pines forests

140   found in the Iberian Peninsula (Fig. S1).

141   DNA was isolated from two legs per individual using a Nucleospin Kit (Macherey-Nagel,

142   Düren, Germany). Specimens were genotyped at 12 microsatellite loci (Mon01, Mon08, Mon17,

143   Mon23, Mon27, Mon30, Mon31, Mon35, Mon36, Mon41, Mon42 and Mon44; Haran & Roux-

144   Morabito 2014). Details of primer sequences and the protocol for genotyping are given in Table

145   S2. Results showing negative or ambiguous amplification of particular loci were repeated once

146   and considered null when still unsatisfactory. Individuals exceeding two missing loci were

147   removed for the analysis. Deviation from Hardy Weinberg Equilibrium (F_{is}) was estimated for

148   each deme, each inferred cluster and for the whole dataset using GENEPOP 4.2 (Raymon &

149   Rousset 1995). The frequency of null alleles at each locus was tested using FREENA (Chapuis &

150   Estoup 2007) among three large demes (n>19). Loci exceeding a rate of 7% of null alleles across

151   populations were discarded from further analysis. Allelic richness was computed for each deme

152   using rarefaction (HP-RARE, Kalinowski 2005). The absence of linkage disequilibrium between

153   pairs of loci was reported in a previous population-based study (Haran et al., 2015).

154

155   ***Genetic structure***

156   We used the Bayesian approach implemented in STRUCTURE 2.3.4 (Pritchard et al.,

157   2000) to identify the main genetic clusters among Iberian demes. STRUCTURE assigns

158   individuals to a predefined number of clusters based on allele composition and linkage

159   disequilibrium. We used the Delta K method (Evanno et al., 2005) to determine the number of

160   clusters (K) that best fitted the data. Genotypes were analyzed using default parameters (admixture

161   model, correlated allele frequency). We made ten repeats of a 200,000 burn-in period followed by

162   500,000 replicates of Markov Chain Monte Carlo (MCMC), for K values ranging from 1 to 20.

163   The results were uploaded in STRUCTURE HARVESTER (Earl et al., 2012) to determine the

164   optimum K. We also explored the existence of genetic clusters among demes using a principal

165   component analysis (PCA) performed on allele frequencies (Adegenet package, Jombart 2008).

166   To account for potential confounding effects of differentiated genetic clusters (possibly of

167   evolutionary history origin) on the inference of gene flow, the landscape genetic analyses of this

168   study (see below) were performed twice, once within the main cluster identified by STRUCTURE

169   and PCA, and once with the whole dataset including all clusters.

170   The genotypes of *M. galloprovincialis* were also analyzed taking a spatial approach in order

171   to identify nested levels of genetic structure linked to scales of study. The scores of the sampling

172   locations on axis 1 of the PCA are linear descriptors of the allele frequencies and, as such, can be

173   used as a univariate statistical measure of genetic composition. The scores may encapsulate

174   relevant spatial information, so we explored this point using a specific tool borrowed from

175   geostatistics: the variogram (Wagner et al., 2005, Goovaerts, 1997). The variogram is used in all

176   branches of life sciences in order to explore spatial patterns and determine the main spatial scales

177   on which structures occur. In our study, we analyzed the score of sample points on axis 1 using a

178   variogram to gain a better understanding of the spatial component of the variation encapsulated in

179   the first axis of the PCA. Let $z(u_\alpha)$, with $\alpha=1, 2, ... n$, be a set of n values of sample scores on a

180   PCA axis where u_α is the vector of spatial coordinates of the αth observation. In geostatistics,

181   spatial dependence is described in terms of dissimilarity between observations expressed as a

182   function of the separating distance (Goovaerts 1997). The average dissimilarity between data

183   separated by a vector h is measured by the empirical semi-variance $\hat{y}(h)$, which is computed as half

184   of the average squared difference between the data pairs:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{x=1}^{N(h)} \left[z(u_\alpha) - z(u_\alpha + h) \right]^2$$

185   (1)

186   where N(h) is the number of data pairs for a given lag vector h, $z(u_\alpha)$ and $z(u_\alpha+h)$ the score

187   values of all sample locations separated by a vector h. The more alike the observations at points

188   separated by h are, the smaller $\hat{y}(h)$ will be, and vice versa. The plot of $\hat{y}(h)$ against h is called a

189   variogram and represents the average rate of change of z with distance. Its shape describes the

190   pattern of spatial variation in terms of general form, scales and magnitude (Goovaerts 1997).

191   Variograms are good tools for depicting spatial structures and analyzing nested patterns (Burrough

192   1983); when structures occur on different spatial scales, the resulting variogram exhibits different

193   plateaus (horizontal flattening of the curve) in association with different scales (Robertson and

194   Gross 1994; Rossi 2003). The range of the variogram is the distance at which the plateau occurs.

195   Multi-plateau variograms exhibit different ranges which provide synthetic information about the

196   spatial scales in play. Readers are referred to Goovaerts (1997) for a thorough introduction to

197   variograms and geostatistics and to Wagner et al., (2005) and to Guillot et al. (2009) for an

198   introduction to this tool in the field of population genetics. Variograms were computed using the

199   R geoR package (Ribeiro and Diggle 2001).

200

201

202   *Landscape genetic analysis*

203   Conventional landscape genetic analyses were first carried out by computing pairwise
204   genetic distances and landscape resistance distances, and then by correlating them.

205   **Computing pairwise genetic distances.** Genetic distances were computed between pairs
206   of individuals using an individual-based metric (Shirk et al., 2010; Prunier et al., 2013). We first
207   constructed a matrix where each individual was a row and alleles were columns and where
208   genotypes were coded for each allele as 0 when absent, 1 when single at a locus (heterozygotes)
209   or 2 for homozygotes (Shirk et al., 2010). Thus, individuals were represented as a linear vector of
210   size n, where n was the total number of alleles encountered in all the individuals genotyped. We
211   then generated a semi matrix of distance between all pairs of individuals. We computed the Bray-
212   Curtis percentage of dissimilarity (Legendre & Legendre 1998) to estimate differentiation between
213   all pairs of individuals. Calculations were performed using the R vegan package (Oksanen et al.,
214   2016).

215   **Computing landscape resistance distances.** We then selected the environmental features
216   considered to be the most likely to influence the dispersal of *M. galloprovincialis* given the existing
217   knowledge of species requirements. Apart from Euclidian geographic distances (null model), we
218   considered three environmental features to be potential drivers of dispersal (pine density,
219   temperature, and elevation).

220   Environmental resistance to dispersal was modeled as a function of pine density as this
221   parameter determines the volume of resource available for *M. galloprovincialis* and is thought to
222   affect its foraging dispersal. As the dispersal behavior of this beetle in reaction to pine density is
223   not known, we modeled this parameter according to two alternative scenarios. (1) High pine
224   densities are positively correlated with beetle dispersal (scenario hereafter called Pc, "pines as
225   corridors"). In this scenario, a dense pine cover represents a corridor for dispersal due to the large
226   amount of resources available. Conversely, a low pine density would represent a barrier. (2) High
227   pine densities are negatively correlated with beetle dispersal (scenario hereafter called Pr, "pines
228   as resistance"). For this second scenario, it was assumed that a dense pine cover provides sufficient
229   resources for local populations, which would therefore not need to disperse. This scenario assumes

230   increased dispersal in low pine cover areas. To model resistance based on pine density, we

231   considered the sum of densities of all pine species encountered in a grid cell, because in the Iberian

232   Peninsula, *M. galloprovincialis* shows no preference for pine species. As shown in previous

233   studies, this beetle will live in the dead wood of any of the pine species considered in this study

234   (*Pinus pinaster, P. nigra, P. sylvestris, P. halepensis* and *P. radiata* (Naves et al., 2006; Haran et

235   al., 2015; 2017). Resistance to dispersal was also modeled as a function of the mean minimum

236   temperature (and its proxy: elevation, hereafter called scenarios T and E respectively), as low

237   summer temperatures tend to inhibit adult flying activity (Hernández et al., 2011), and because

238   low winter temperatures are likely to determine the survival or the development rate of *M.*

239   *galloprovincialis* larval instars (Naves & Sousa 2009). We consider that resistance increases when

240   the annual mean minimum temperatures decrease. We kept temperatures and elevation as distinct

241   environmental features for the analysis, as altitude and temperatures are not similarly correlated in

242   the North and South of Spain.

243

244   Resistance distances were computed using the gdistance package (van Etten 2012). Raster

245   layers of environmental features were imported at a resolution of 10 x 10 km. Such a resolution

246   was chosen because the mean flight distance of *M. galloprovincialis* reaches 16 km, based on flight

247   mill experiments (David et al., 2014). Temperature data (1950-2000) were downloaded from

248   Hijmans et al., (2005; http://www.worldclim.org; original resolution: 1 x 1 km), the pine density

249   from Tröltzsch et al., (2009; http://www.efi.int/; original resolution: 1 x 1 km) and elevation from

250   ARCGIS 9.3 (ESRI, Redlands, CA, USA; original resolution: 1 x 1 km). Null distances were not

251   encountered at a grain size of 10 x 10 km, as none of the sampling sites fell within neighboring

252   sites in the same grain. As temperature, elevation and pine density are continuous parameters, we

253   did not assign particular resistances to particular values, but directly used the values (except for

254   the scenario Pc for which values were set as negative). Pairwise resistance distances were

255   estimated based on random walk probabilities (Chandra et al., 1997, McRae 2006) and computed

256   using the commuteDistance command (gdistance package). Resistance distances were chosen

257   instead of least cost distances (LCD) because they are thought to be more reliable biologically and

258   produce fewer artifacts over long distances (McRae 2006). We constructed a semi matrix of

259   resistance distances between each pair of individuals. Values were normalized to a common scale

260   for further analysis. Collinearity was estimated using the variance inflation factor (VIF) based on

261   the formula VIF = $1/(1-R^2)$, where R^2 is the r-squared value of regression between variables. VIF

262   values > 10 are usually considered evidence for collinearity between environmental features

263   (O'Brien 2007). We did not detect collinearity between environmental features over the whole

264   area of study (VIF < 1 for all pairwise comparisons).

265   **Correlation analysis.** We tested correlations between the response (genetic distances

266   matrix, G) and resistance distances (resistance matrices; Isolation By Resistance: IBR) and the

267   geographic distances (Euclidian geographic distance; Isolation By Distance: IBD) using partial

268   Mantel tests (Cushman & Landguth 2010). Partial Mantel tests measure associations between two

269   distance matrices while partialling out a third distance matrix. We first used simple Mantel tests

270   to correlate IBD with G. We then tested the effect of IBR in partial Mantel tests. Support for the

271   IBR hypothesis was considered when: (1) IBR should be significantly correlated to G after

272   partialling out IBD ($p < 0.05$) and IBD should be non-significant with IBR partialled out ($p \geq 0.05$;

273   Cushman et al., 2006). Mantel and partial Mantel tests were performed using the vegan package

274   with 10^3 permutations. This approach is widely used in the field of landscape genetics (Cushman

275   et al., 2006; Cushman and Landguth 2010; Galpern et al., 2012; Castillo et al., 2014) and has been

276   shown to efficiently infer the drivers of gene flow (Cushman & Landguth 2010b). However, partial

277   Mantel tests have received criticism regarding their statistical performance (Guillot & Rousset

278   2013; Diniz-Filho et al., 2013), and are therefore preferably used together with complementary

279   approaches such as ordination methods (Kierepka et al., 2015). To overcome the potential

280   weakness of partial Mantel tests on our dataset, and to validate the statistical significance of

281   correlations, distance matrices were also regressed using commonality analysis (Prunier et al.,

282   2014). This method is based on variance-partitioning and therefore allows the relative importance

283   of the environmental features shaping genetic structure to be estimated, accounting for covariance

284   in the features tested. For the commonality analysis, the response G was regressed onto each

285   resistance matrix separately and each combination using the R yhat package (Nimon et al., 2013).

286

287   *Resampling on multiple scales and at multiple locations*

288   The above Partial Mantels test and commonality analysis were conducted in areas of

289   various spatial scales and in various locations in order to estimate the effects of scale and location

290   on the inference of the landscape genetic analysis for the four IBR hypothesis (IBR-T, IBR-E,

291   IBR-Pr and IBR-Pc). First, nested sampling areas (sliding windows) were generated over the full

292   extent of the Iberian Peninsula. These areas were constructed as circles of diameters ranging from

293   220 to 1000 km (steps of 20 km) and centered on each sampling location (scale therefore refers to

294   the extent, Mayer & Cameron 2003). Mantel tests were performed using all individuals found

295   within each defined area. Areas with a diameter below 220 km were not included, because it was

296   too small to gather neighboring demes for Mantel tests in the areas with scattered sampling. Then,

297   the statistical support of the landscape analysis performed in the sampling areas generated were

298   compared against their scale and location. The effect of scale was observed by summing the

299   number of areas with a supported IBR hypothesis and computing their mean Mantel r on each

300   scale. The effect of the geographic distribution (location) of sampling areas on the detection of the

301   landscape effect on gene flow was examined by mapping areas with a supported IBR hypothesis.

302   The map was obtained by summing the number of times that each individual was included in a

303   sampling area with IBR hypothesis support among all scales. The values obtained were corrected

304   accounting for intrinsic variation due to overlapping sampling areas. Values at each point were

305   interpolated using the Inverse Distance Weighted (IDW) method in ARCGIS 9.3 (ESRI, Redlands,

306   CA, USA). Given that landscape genetic analyses perform better in a contrasting landscape (i.e.

307   high amplitudes of values of resistant features; Jaquiéry et al., 2011; Cushman et al., 2013), we

308   also sought whether or not support for the IBR hypotheses was due to variations in the

309   environmental features tested. The resistance values of raster cells within each sampling area were

310   extracted and the standard deviation (SD) of those values was computed. We then calculated and

311   compared the mean standard deviation of areas with supported and non-supported IBR hypotheses

312   among the area scales tested. Commonality analyses (see above) were performed within each

313   sampling area generated. As for the Mantel tests, the variation in commonality coefficients

314   (percentage of variance explained by a unique and cumulative IBR hypothesis) was observed by

315   changing the scale and location of the sampling areas. The area maximizing commonality

316   coefficients was chosen for representation of the relative importance of environmental features in

317   shaping genetic structure. All computations were performed using R software version 3.0.2 (R

318   development Core Team 2013).

319

Results

Genotyping

Overall, 1050 individuals were successfully genotyped. Among the 3 populations of larger sizes tested (n>19), two loci exhibited substantial null allele frequencies (> 7%) and were therefore not considered for further analysis (Mon 01 and Mon 27). A significant heterozygote deficit was detected at four loci (Mon 30, 35, 42, 44). Corresponding null allele frequencies were low (<7%), so these loci were kept. After the removal of incomplete genotypes (n=58) and biased loci, we obtained a total of 992 individuals genotyped at ten loci. The average number of alleles per locus was 10.2 (range: 6-24). The number of alleles per deme (using rarefaction) ranged from 1.32 to 1.64 and F_{is} estimates from -0.27 to 0.38 (Table S1).

Genetic structure

Individuals formed an optimum number of two clusters under STRUCTURE analysis (Figure S3). The clusters showed a clear geographic structure, exhibiting a split between Portugal and western Galicia (West Iberian cluster) versus the rest of the Iberian Peninsula (Fig. 1A). The PCA gave similar results on the first axis (eigenvalue: 0.494 accounting for 14.3% of the total inertia), splitting demes into two distinct clusters (Fig. 1C). Estimates of population differentiation (F_{st}) between the three populations of large size (n>19) were moderate (Castro Daire /Catsellbell: 0.13; Castro Daire/Vale Feitoso: 0.13; Catsellbell/ Vale Feitoso: 0.05; p < 0.001).

The data points of the variogram were grouped into 26 distance classes ranging from 0 to 1252 km, with a distance interval of 50 km. The variogram revealed that the first axis of the PCA corresponded to a highly spatially structured pattern (Fig. 2). The semi-variance first progressively increased with increasing lag distance up to a distance of about 190 km and then reached a plateau. For distances of about 400 km, the semi-variance increased again and leveled off for distances further than 1000 km. The resulting plateau of the variogram showed the presence of a long-range spatial variation superimposed over a more local, i.e. short-scale genetic structure, occurring on scales of 200 to 400 km. For scales below 200 km, the variogram showed that genotypes were strongly spatially auto-correlated (i.e. non-independent).

348

Multiple scale and multiple location analysis

350   Analyses were conducted on both the whole dataset (992 individuals, 132 localities) and
351 within the Spanish cluster (790 individuals, 87 localities). Within each dataset, the number of
352 alleles observed across all individuals and loci was 116 and 102, respectively.

353   Over the whole study area (whole dataset), we generated a total of 30,576 sampling areas.
354 The mean number of individuals within sampling areas varied from 89.18 (SD: 42.42) on the
355 smallest scale (220 km) to 644.58 (SD: 158.07) on the largest scale (1000 km; Fig. S2). Significant
356 effects of environmental features were detected for all IBR hypotheses tested with partial Mantel
357 tests, but the frequency of areas exhibiting an IBR effect varied between scales and locations (Fig.
358 3A). Significant effects of environmental features were detected for the IBR-E, IBR-Pr and IBR-
359 T hypotheses in about 15-25% of the areas on the smallest scale (220 - 300 km). The frequency of
360 IBR-E and IBR-Pr gradually increased to reach 97% and 70% for areas of 1000 km. The frequency
361 of areas with a supported IBR-T hypothesis increased between scales to reach a peak around 600
362 km (82% of areas) and subsequently decreased. Significant IBR-Pc hypotheses were encountered
363 at a lower frequency. The number of areas with a positive effect of the Pc hypothesis was 4.07
364 areas, on average, for all the scales considered. Significant isolation by distance (IBD) was
365 observed for 60% of areas on the smallest scale. A first plateau of about 85% of areas was reached
366 for scales ranging between 400 and 700 km, and a second plateau of almost 100% of areas was
367 reached for scales above 700 km. Mean Mantel r values for areas with a supported IBR hypothesis
368 ranked between 0.05 and 0.25 and generally decreased when the scale increased (Fig. 3B).
369 Hypothesis IBR-T showed the highest mean Mantel r value out of all the IBR hypotheses for scales
370 above 360 Km.

371   Interpolation of supported IBR hypotheses and IBD was based on areas of scales ranging
372 from 220 to 600 km, because most of the variation in the detection of the effects of environmental
373 features was found on these scales (Fig. 3A). For most IBR hypotheses (IBR-E, IBR-Pr and IBR-
374 T) and IBD, the effects were mainly detected in the northern part of the study area (Fig. 4). In
375 contrast, these IBR hypotheses were the least frequently detected on the southern and eastern sides
376 of the Iberian Peninsula. For the IBR-Pc hypothesis, significant effects were detected mainly in

377 Andalucía, along the Betic system. Conversely, low or no effects for this hypothesis were detected
378 in the northern half of the Iberian Peninsula. The distribution of supported hypotheses was similar
379 for those performed on the whole dataset and on the Spanish cluster only (Fig. 4). For hypotheses
380 IBR-E, IBR-Pr and IBR-Pc, the variation of environmental features was lower on average in areas
381 exhibiting significant effects for scales up to 400 - 600 km (whole dataset; Fig. 5). Above this
382 scale, the mean standard deviation (SD) of areas with a supported IBR hypothesis was either equal
383 to or higher than the mean SD of areas with no support. For the IBR-T hypothesis, the mean SD
384 of areas with support was above the mean for non-supported areas, for most of the scales. The
385 results of the commonality analysis were in agreement with those of the Mantel tests. The IBR-T
386 and IBR-Pr hypotheses purely contributed to most of the total variance explained (20.77 to 32.65%
387 and 21.82 to 35.24%, respectively, Table 1), and the highest contribution to the total variance
388 explained was observed for the joint effects of IBR-E and IBR-T (54.31 to 56.43%). Conversely,
389 the IBR-Pc hypothesis made a limited contribution to the variance explained in pure effect (1.80
390 to 3.81%). As for the Mantel tests, the sampling areas showing the maximum explained variance
391 over all sampling areas (20.9 to 24%) were located in the western and northern parts of the Iberian
392 Peninsula and had a medium diameter (520 to 620 km).

393

Discussion:

395 The landscape genetics approach seeks to infer general drivers of gene flow for species in
396 a heterogeneous landscape context (Manel & Holderegger 2013). However, the ability of this
397 approach to make inferences about the effect of environmental features may vary due to multiple
398 evolutionary processes acting on the genetic structure of organisms on different spatial and
399 temporal scales. In this study, we explored potential barriers against and corridors for the dispersal
400 and gene flow of a flying insect in a large area with dramatic landscape changes. Based on multi-
401 scale and multi-site resampling of study areas, we successfully identified the effects of
402 environmental features on gene flow for independent study areas, and for different inference
403 methods. Resampling provided a quantitative value for the results, making it possible to control
404 the Type-1 errors associated with the inference methods employed (Shortbull et al. 2011). Apart
405 from identifying relevant environmental features, the resampling method showed patterns for the

406   frequency at which landscape effects were detected among locations, especially on the smallest
407   spatial scales. Interestingly, most of these patterns could be explained by the biology of the species
408   model, by identified artifacts of the sampling design and by known changes in the landscape
409   structure of the study area.

410

411   ***Importance of the scale of the study area for flying insects***

412   We observed a notable influence of scale on the detection of supported IBR hypotheses
413   with Mantel tests for most of the environmental features tested (IBR-E, IBR-T and IBR-Pr).
414   Support was scarcely detected on the smallest spatial scale (220-400 km) and was generally more
415   often detected with increasing scale. A larger study area had a larger sampling leading to better
416   detection of positive correlations via partial Mantel tests (Cushman & Landguth 2010; Landguth
417   et al. 2012). Beyond the simple effect of the sampling size, the spatial scale of gene flow and
418   resulting genetic structure also influenced the inference of the hypothesis tested. The smallest
419   spatial scales of 190-400 km corresponded to the distances at which the variogram showed an
420   initial plateau of genetic dissimilarity (Figure 2). This correspondence suggested that for this range
421   of scales, the dissimilarity between individuals of *M. galloprovincialis* was not enough to show a
422   significant effect of environmental features on gene flow in most study areas. In contrast, the peak
423   (for IBR-T) or inflection of the curves (for IBR-Pr, IBR-E) of the number of areas with supported
424   IBR hypotheses observed on scales ranging from 400 to 600 km corresponded to the increase in
425   dissimilarity in the variogram (Figure 2). Thus, scales above 400 km seemed more appropriate for
426   observing a genetic structure of *M. galloprovincialis* potentially structured by environmental
427   features. Interestingly, we found that the variation in the frequency of areas with support was
428   specific to each environmental feature tested. In line with previous studies on large mammals
429   (Zeller et al., 2014) and on insects (Rasic and Keyghobadi 2012), this study highlighted that each
430   environmental feature affected gene flow on a distinct scale, and that empirical studies including
431   several features should always consider a range of study scales.

432   Weak support of the IBR hypothesis on a small scale illustrated a general problem of the
433   landscape genetic analysis performed on flying species, which are naturally less affected by
434   environmental features than non-flying species. For such species, the combination of intensive
435   dispersal and associated gene flow and a limited number of environmental features affecting

436 dispersal makes inference difficult on small spatial scales (Dreier et al., 2014). In the *Monochamus*
437 species, substantial flying distances have been measured (up to 22 km in the field: Takasu et al.,
438 2000; Linit & Akbulut 2003; Hernandez et al., 2011; Gallego et al., 2012; Mas et al., 2013; David
439 et al., 2014), causing strong inbreeding of populations and leading to a weak local genetic structure
440 (Kawai et al., 2006; Shoda-Kagaya 2007; Haran et al., 2015). Considering a continuous range of
441 scales in the analysis prevented us from basing our conclusions on a scale for which the effect of
442 environmental features could not be detected. In this respect, an experimental design for flying
443 insect species should follow the approaches used for large mammals, for which multiple scales,
444 including very large scales, have been used to account for a scale of gene flow that is unknown a
445 priori (Galpern et al., 2012; Zeller et al., 2014).

446   ***Effect of geographic location on inference***

447   Based on the resampling of study areas in the East Iberian cluster, we observed the
448 existence of heterogeneous geographic distribution of supported resistance models (IBR-T, IBR-
449 E and IBR-Pr). Most variation in the distribution of support for IBR was observed on small and
450 intermediate scales (220-600 km), the smallest scale being larger than areas at which landscape
451 genetic analysis are usually conducted (Shortbull et al. 2011, Zeller et al. 2012). On these scales,
452 supported effects were mainly detected in the northern-central part of the Iberian Peninsula.
453 Conversely, effects were less supported in the rest of the Iberian Peninsula (center, south and
454 coasts). Variation of support are known to occur when multiple area of study are tested in
455 landscape genetics (Shortbull et al. 2011), however, in this case all areas with an effect were
456 located in the same region. These results highlighted that the genetic structure of *M.*
457 *galloprovincialis* was structured according to the environmental features tested in some areas but
458 not in others, independently of variations in the heterogeneity of the environmental features
459 (Figure 5). This observation is interesting, because one might expect a native species such as *M.*
460 *galloprovincialis* to have homogeneous dispersal in response to environmental features, at least
461 within a genetic lineage, and that this effect would be detected homogeneously across a study area.
462 Determining the exact origin of such heterogeneity is challenging. It is suggested that this variation
463 was a legacy of changes in the distribution of host trees in the Iberian Peninsula. Indeed, the
464 distribution and density of pine trees have been strongly affected by anthropogenic activities over
465 recent centuries (Ruiz-Benito et al., 2012; Lopez-Merino et al., 2014), resulting in local extinction,

466 as well as the connectivity and fragmentation of pine tree cover over time. For example, Abel-
467 Schaad et al., (2014) showed that pine trees locally disappeared from the Central Iberian System
468 during the middle ages. In contrast, these areas were 80% afforested with pines trees over the 1940-
469 1950 period. It is assumed that such recent modifications have dramatically affected the
470 distribution and abundance of *M. galloprovincialis*, and that the time since these modifications
471 occurred is too short to have affected the genetic structure of the beetle according to the
472 environmental features tested (Epps & Keyghobadi 2015).

473   This study also highlighted empirically how important it is to account for the genetic
474 differentiation of the species derived from evolutionary history when selecting the location of a
475 study area. We found that the effect of environmental features on gene flow was not detected in
476 study areas mainly located along the western Iberian coast, and these areas always overlapped the
477 West and East Iberian genetic clusters. In these cases, substantial genetic differentiation blurred
478 the genetic structure derived from the effect of environmental features on gene flow, which is more
479 recent and weaker in *M. galloprovincialis*. This effect of differentiated genetic clusters on
480 inference was also observed in the strength of the correlations: we observed a decrease in the mean
481 Mantel r for large-scale study areas (600-1000 km) which always overlapped the two genetic
482 clusters. It is difficult to predict exactly to what extent genetic divergence derived from
483 evolutionary history can contribute, or not, to the detection of an effect of environmental features
484 on gene flow, as this structure can also derive from old and stable barriers to dispersal. While the
485 vast majority of species exhibit differentiated genetic linages derived from evolutionary history
486 (Hewitt, 2000), the genetic structure of a species beyond the extent of the study area is rarely
487 explored in landscape genetic analysis (Zeller et al. 2012). Our results suggest that carrying out
488 analyses both within each genetic cluster and over the whole dataset is preferable, to avoid the
489 confounding effect of evolutionary history on landscape genetic analyses.

490

491   ***Relevance of inference for the model species***

492   This study enabled us to generalize patterns of dispersal for *M. galloprovincialis* and
493 confirmed previous observations made on the biology and ecology of this species. Mantel tests
494 and a commonality analysis supported the hypothesis that elevated areas and their associated
495 colder temperatures constitute barriers to gene flow for *M. galloprovincialis*. This result

496   corroborates the observations made for the congeneric species *M. alternatus* across the Ohu chain

497   of mountains in Japan (Shoda-Kagaya 2007) and on *M. galloprovincialis* across the Pyrenees

498   (Haran et al.*,* 2015; 2017). In addition, this hypothesis tends to be confirmed by several studies

499   showing that *M. galloprovincialis* larva development and survival (Naves and Sousa 2009), and

500   its ability to complete its development within one or two years (Tomminen 1993; Naves et al.,

501   2007b; Koutroumpa et al., 2008), are affected by low temperatures. Lastly, the adult flying activity

502   of this beetle was shown to be negatively affected by low daily temperatures (Hernández et al.,

503   2011), and to be directed mainly down-hill (Torrez-Vila et al., 2015). Our analysis also supports

504   the hypothesis that *M. galloprovincialis* exhibits limited dispersal when pine trees are abundant.

505   This species is known to develop on dead branches stemming from a self-pruning process

506   encountered in pines (Mäkinen 1999). Dead branches represent a resource that is quite well

507   distributed in space and time in pine stands. Such an abundance of resource is thought to cause

508   limited dispersal in adults. The philopatric behavior of *M. galloprovincialis* in relation to available

509   resources is consistent with flight observations of this species in the field (Torres-Vila et al., 2015),

510   or with the behavior of the pine processionary moth (*Thaumetopoea pityocampa*), another

511   oligophagous pine–associated insect (Demolin 1969). Conversely, the *Pr* hypothesis suggested

512   that low pine densities are not barriers to dispersal. This is in agreement with the suggestions of

513   Torres-Vila et al., (2015) that the dispersal of *M. galloprovincialis* tends to be enhanced across

514   open areas. Rossi et al., (2016) showed that areas without pine forests still show a homogeneous

515   distribution of scattered trees planted for ornamental use using observed and simulated data. We

516   suggest that pine trees outside forests provide a scattered but homogeneously distributed resource

517   that allows the dispersal of *M. galloprovincialis* across non-forested areas.

518

519   **Conclusions**

520   Our results show the importance of simultaneously considering a continuous range of

521   scales and multiple locations when exploring the effect of environmental features on dispersal in

522   highly mobile species. Multiple scales allow the effect of environmental features to be inferred at

523   the appropriate extent for each feature tested, while preventing the analysis from being focused on

524   an extent where intensive gene flow makes inference impossible due to the lack of genetic

525 structure. In addition, resampling of the study area across multiple locations helped to identify

526 variation in inference due to conflicting signals in genetic structure, and thereby allowed for

527 generalizing conclusions regarding the effects of environmental features on dispersal and gene

528 flow. As a result, the combination of a resampled study area on multiple spatial scales across

529 various locations in a landscape genetic analysis provides a more general picture of the effects of

530 environmental features on the gene flow of organisms and has the power to reduce the variability

531 of results while limiting the sampling effort.

532

Acknowledgments

534 We warmly thank Rolf Holderegger and Bertrand Gauffre for valuable comments on the

535 early versions of this manuscript. We also thank Jérôme Rousselet and Christelle Robinet for

536 interesting exchanges and help regarding the methodology and Peter Biggins for English editing.

537

Tables

539 **Table 1**: Commonality coefficients of both unique and common effects for the three sampling

540 areas with the highest variance explained. *Code pop*: population code for the center of the sampling

541 area. *Scale*: diameter of sampling area (km). *N*: number of individuals in sampling area. *Coef.*:

542 percentage of variance explained by environmental features (IBR hypotheses). *% Total*:

543 percentage of the contribution of environmental features to the total variance explained.

544 Environmental features and their associated Isolation By Resistance (IBR) hypothesis: *E*: elevation

545 model (high elevations = resistance to dispersal). *Pc*: Pine density model (high pine density =

546 corridors to dispersal. *Pr*: reversed pine density model (high pine density = resistance to dispersal).

547 *T*: temperature model (low minimum annual temperatures = resistance to dispersal).

548

549

550

551

Figure captions

553 **Figure 1**: Genetic clustering of 992 individuals of *Monochamus galloprovincialis* sampled at 132
554 locations. **A**: Assignment of individuals to clusters based on a STRUCTURE analysis for K=2. **B**:
555 Assignment of demes to clusters for k=2, displayed in geographic context (Iberian Peninsula, the
556 size of the pies refers to the size of the demes). **C**: PCA of individuals on first and second axes,
557 colors and ellipses refer to demes.

558 **Figure 2**: Empirical semi-variogram of genotypes of *Monochamus galloprovincialis*. The
559 variogram was fitted with an exponential model to highlight the first plateau. Data points are shown
560 with a spatial lag distance of 50 km.

561 **Figure 3**: Percentage of areas with supported IBR hypotheses for Mantel tests (**A**) and mean
562 partial Mantel r (**B**) of areas with support for the IBR hypotheses (p<0.05) with increasing scales
563 (whole dataset). *E*: Elevation, *T*: Mean minimum temperatures, *Pr* and *Pc*: pine densities as a
564 resistant feature and as a corridor respectively, IBD: Isolation by distance.

565 **Figure 4**: Distribution of supported IBR hypotheses through Mantel tests for the environmental
566 features tested (IBD: Euclidian distances; *T*: mean minimum temperatures; *E*: elevation; *Pr*: high
567 pine densities as barriers; *Pc*: high pine densities as corridors). Gray maps (A) refer to the
568 distribution of environmental features associated with resistance models (From white to black: low
569 to high resistance values). Colored maps refer to interpolations of supported IBR hypotheses on
570 the whole dataset (B) and within the western Iberian cluster only (C). From blue to red: low to
571 high frequency of study areas with supported resistance models.

572 **Figure 5**: Spatial heterogeneity (mean standard deviation, SD) of environmental features in areas
573 with supported and non-supported resistance hypotheses through Mantel tests with increasing
574 scales. Mean SD: mean standard deviation. *T*: mean minimum temperatures (A); *E*: elevation (B);
575 *Pr*: high pine densities as barriers (C); *Pc*: high pine densities as corridors (D). sig: significant,
576 non-sig: non-significant.

577

578

579 # References

580 Abel-Scaad D, Lopez-Saez JA, Pulido F. 2014. Heathlands, fire and grazing. A
581 paleoenvironmental view of Las Hurdes (Cáceres, Spain) history during the last 1200 years.
582 *Forest Systems* 23:247–258.
583 Anderson CD, Epperson BR, Fortin MJ, Holderegger R, James PMA, Rosenberg MS, Scribner
584 KT, Spear S. 2010. Considering spatial and temporal scale in landscape-genetic studies of
585 gene flow. *Molecular Ecology* 19:3565–3575.

Angelone S, Kienast F, Holderegger R. 2011. Where movement happens: scale-dependent landscape effects on genetic differentiation in the European tree frog. *Ecography* 34:714–722.

Burrough PA. 1983. Problems of superimposed effects in statistical study of the spatial variation in soil. *Agricultural Water Management* 6:123–143.

Castillo JA, Epps CW, Davis AR, Cushman SA. 2014. Landscape effects on gene flow for a climate-sensitive montane species, the American pika. *Molecular Ecology* 23:843–856.

Chandra AK, Raghavan P, Ruzzo WL, Smolensky R, Tiwari P. 1997. The electrical resistance of a graph captures its commute and cover times. *Computational Complexity* 6:312–340.

Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24:621–631.

Cushman SA, Landguth E. 2010. Scale dependent inference in landscape genetics. *Landscape Ecology* 25:967–979.

Cushman SA, McKelvey KS, Hayden J, Schwartz MK. 2006. Geneflow in complex landscapes: testing multiple models with causal modeling. *American Naturalist* 168:486–499.

Cushman SA, Shirk AJ, Landguth E. 2013. Landscape genetics and limiting factors. *Conservation Genetics* 14:263–274.

David G, Giffard B, Piou D, Jactel H. 2014. Dispersal capacity of *Monochamus galloprovincialis*, the European vector of the pine wood nematode, on flight mills. *Journal of Applied Entomology* 138:566–576.

Démolin G. 1969. Comportement des adultes de *Thaumetopoea pityocampa* Schiff. Dispersion spatiale, importance écologique. *Annales des Sciences Forestières* 26:89–102.

Diniz-Filho JAF, Soares TN, Lima JS, Dobrovoski R, Landeiro VL, Telles MPD, Rangel TF, Bini LM. Mantel test in population genetics. *Genetics and Molecular Biology* 36:475–485.

Dreier S, Redhead JW, Warren IA, Bourke AF, Heard MS, Jordan WC, Sumner S, Wang JL, Carvell C. 2014. Fine-scale spatial genetic structure of common and declining bumble bees across an agricultural landscape. *Molecular Ecology* 23:3384–3395.

Drizen K, Adriaensen F, Rondinini C, Doncaster CP, Matthysen E. 2007. Evaluating least-cost model predictions with empirical dispersal data: a case-study using radiotracking data of hedgehogs (*Erinaceus europaeus*). *Ecological Modelling* 209:314–322.

Dudaniec RY, Rhodes JR, Wilmer JW, Lyons M, Lee KE, McAlpine CA, Carrick FN. 2013. Using multilevel models to identify drivers of landscape-genetic structure among management areas. *Molecular Ecology* 22:3752–3765.

Earl DA, VonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4:359 361.

Epps CW, Keyghobadi N. 2015. Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. *Molecular Ecology* 24:6021–6040.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14:2611–2620.

626 Gallego D, Sanchez-Garcia FJ, Mas H, Campo MT, Lencina YJL. 2012. Estudio de la capacidad
627   de vuelo a larga distancia de *Monochamus galloprovincialis* (Olivier 1795). (Coleoptera:
628   Cerambycidae) en un mosaico agro-forestal., *Boletin de Sanidad Vegetal Plagas* 38:109–
629   123.

630 Galpern P, Manseau M, Wilson P. 2012. Grains of connectivity: analysis at multiple spatial scales
631   in landscape genetics. *Molecular Ecology* 21:3996–4009.

632 Guillot G, Leblois R, Coulon A, Frantz AC. 2009. Statistical methods in spatial genetics.
633   *Molecular Ecology*. 18: 4734–4756.

634 Goovaerts P. 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press,
635   Oxford.

636 Guillot G, Rousset F. 2013. Dismantling the Mantel tests. *Methods in Ecology and Evolution*
637   4:336–344.

638 Haran J, Rousselet J, Tellez D, Roques A, Roux G. 2017. Phylogeography of *Monochamus*
639   *galloprovincialis*, the European vector of the pinewood nematode. *Journal of Pest Science*.
640   doi: 10.1007/s10340-017-0878-4.

641 Haran J, Roques A, Barnard A, Robinet C, Roux G. 2015. Altitudinal barrier to the spread of an
642   invasive species: could the Pyrenean chain slow the natural spread of the pine wood
643   nematode? *PLoS ONE* 10(7):e0134126.

644 Haran J, Roux-Morabito G. 2014. Development of 12 microsatellites loci for the longhorn beetle
645   *Monochamus galloprovincialis* (Coleoptera Cerambycidae), vector of the pinewood
646   nematode in Europe. *Conservation Genetics Resources* 6:975–977.

647 Hellrigl KG. 1971. La bionomie des espèces de *Monochamus* (Coleoptera, Cerambycidae) et leur
648   importance pour la sylviculture et l'économie du bois. *Redia* 52:367–511.

649 Hernández R, Ortiz A, Pérez V, Gil JM, Sanchez G. 2011. *Monochamus galloprovincialis*
650   (Olivier, 1795) (Coleoptera: Cerambycidae), comportamiento y distancias de vuelo. *Boletin*
651   *de Sanidad Vegetal Plagas* 37:79–96.

652 Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405:907–913.

653 Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated
654   climate surfaces for global land areas. *International Journal of Climatology* 25:1965–1978.

655 Holderegger R, Wagner HH. 2008. Landscape genetics. *BioScience* 58:199–207.

656 Jaquiéry J, Broquet T, Hirzel AH, Yearsley J, Perrin N. 2011. Inferring landscape effects on
657   dispersal from genetic distances: how far can we go? *Molecular Ecology* 20:692–705.

658 Jactel H, Castagnone P, Mota M, Robinet C, Roux G, Sarniguet C. 2015. *Evaluation of emergency*
659   *measures to prevent the spread of the pine wood nematode within the European Union*.
660   ANSES opinion, Collective Expert Appraisal Report, 61 pp.

661 Jombart T. 2008. ADEGENET: a R package for the multivariate analysis of genetic markers.
662   *Bioinformatics* 24:1403–1405.

663 Kalinowski S. 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures
664   of allelic richness. *Molecular Ecology Notes* 5:187–189.

665 Kawai M, Shoda-Kagaya E, Maehara T, Zhou ZH, Lian CL, Iwata R, Yamane A, Hogetsu T.
666 2006. Genetic structure of pine sawyer *Monochamus alternatus* (Coleoptera :
667 Cerambycidae) populations in northeast Asia: Consequences of the spread of pine wilt
668 disease. *Environmental Entomology* 35:569–579.
669 Keller D, Holderegger R, van Strien MJ. 2013. Spatial scale affects landscape genetic analysis of
670 a wetland grasshopper. *Molecular Ecology* 22:2467–2482.
671 Kierepka EM, Latchi EK. 2015. Performance of partial statistics in individual-based landscape
672 genetics. *Molecular Ecology* 15:512–525.
673 Kindall JL, Van Manen FT. 2007. Identifying habitat linkages for American black bears in North
674 Carolina, USA. Journal of Wildlife Managment 71:487–495.
675 Koutroumpa F, VincentB, Roux-Morabito G, Martin C, Lieutier F. 2008. Fecundity and larval
676 development of *Monochamus galloprovincialis* (Coleoptera Cerambycidae) in experimental
677 breeding. *Annals of Forest Science* 65:707.
678 Legendre P, Legendre L. 1998. Numerical Ecology. Elsevier, Amsterdam.
679 Linit MJ, Akbulut S. 2003. Pine wood nematode phoresis: the impact on *Monochamus*
680 *carolinensis* life functions. *Nematology Monographs and Perspectives* 1:227–237.
681 Lopez-Merino L, Martinez Cortizas A, Reher GS, Lopez-Saez JA, Mighall TM, Bindler R. 2014.
682 Reconstructing the impact of human activities in a NW Iberian Roman mining landscape for
683 the last 2500 years. *Journal of Archaeological Science* 50:208–218.
684 Mäkinen H. 1999. Growth, suppression, and self-pruning of branches of Scots pine in southern
685 and central Finland. *Canadian Journal of Forest Research* 29:585–594.
686 Manel S, Holderegger R. 2013. Ten years of landscape genetics. *Trends in Ecology and Evolution*
687 28:614–621.
688 Manel S, Schwartz MK, Luikart G, Taberlet P. 2003. Landscape genetics: combining landscape
689 ecology and population genetics. *Trends in Ecology and Evolution* 18:189–197.
690 Mas H, Hernandez R, Villaroya G, Sanchez G, Pérez-Laorga E, González E, Ortiz A, Lencina JL,
691 Rovira J, Marco M, Pérez V, Gil M, Sánchez-García FJ, Bordón P, Pastor C, Biel MJ,
692 Montagud L, Gallego D. 2013. Dispersal behavior and long distance flight capacity of
693 *Monochamus galloprovincialis* (Olivier 1795), In: Schröder, T (ed.), Pine Wilt Disease
694 Conference 2013, pp. 22, Braunschweig, ISSN: 1866-590X.
695 Mayer AL, Cameron GN. 2003. Consideration of grain and extent in landscape studies of
696 terrestrial vertebrate ecology. *Landscape and Urban Planning* 65:201 – 217.
697 McRae B. 2006. Isolation by Resistance. *Evolution* 60:1551–1561.
698 Naves P, de Sousa E. 2009. Threshold temperatures and degree-day estimates for development of
699 post-dormancy larvae of *Monochamus galloprovincialis* (Coleoptera: Cerambycidae).
700 *Journal of Pest Science* 82:1–6.
701 Naves PM, Camacho S, de Sousa E, Quartau JA. 2007. Transmission of the pine wood nematode
702 *Bursaphelenchus xylophilus* through oviposition activity of *Monochamus galloprovincialis*
703 (Coleoptera : Cerambycidae). *Entomologica Fennica* 18:193–198.

704 Naves P, Sousa E, Quartau J. 2006. Feeding and oviposition preferences of *Monochamus*
705   *galloprovincialis* for some conifers under laboratory conditions. *Entomologica*
706   *Experimentalis et Applicata* 120:99–104.

707 Nimon K, Oswald F, Roberts JK. 2013. yhat: Interpreting Regression Effects. R package version
708   2.0-0.

709 O'Brien RM (2007. A caution regarding rules of thumb for variance inflation factors. *Quality and*
710   *Quantity* 41:673–690.

711 Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos
712   P, Stevens MHH, Szoecs E, Wagner H. 2016. vegan: Community Ecology Package. R
713   package version 2.3-3. http://CRAN.R-project.org/package=vegan

714 Pflüger FJ, Balkenhol N. 2014. A plea for simultaneously considering matrix quality and local
715   environmental conditions when analyzing landscape impacts on effective dispersal.
716   *Molecular Ecology* 23:2146–2156.

717 Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus
718   genotype data. *Genetics* 155:945–959.

719 Prunier JG, Colyn M, Legendre X, Nimon KF, Flamand MC. 2014. Multicollinearity in spatial
720   genetics: separating the wheat from the chaff using commonality analyses. *Molecular*
721   *Ecology* 24:263–283.

722 Prunier JG, Kaufmann B, Fenet S, Picard D, Pompanon F, Joly P. 2013. Optimizing the trade-off
723   between spatial and genetic sampling efforts in patchy populations: towards a better
724   assessment of functional connectivity using an individual-based sampling scheme.
725   *Molecular Ecology* 22:5516–5530.

726 Rasic G, Keyghobadi N. 2012. From broad scale patterns to fine-scale processes: habitat structure
727   influences genetic differentiation in the pitcher plant midge across multiple spatial scales.
728   *Molecular Ecology* 21:223–236.

729 Raymond M, Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact
730   tests and ecumenicism. *Journal of Heredity* 86:248–249.

731 R Development Core Team. 2013. R: a language and environment for statistical computing. R
732   Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

733 Ribeiro PJ, Diggle PJ. 2001. geoR: a package for geostatistical analysis. R-NEWS 1, 14–18.

734 Robertson GP, Gross KL. 1994. Assessing the heterogeneity of belowground resources:
735   quantifying pattern and scale. In: Caldwell M, Pearcy R (eds.), Exploitation of
736   Environmental Heterogeneity by Plants. Academic Press, pp. 237-253.

737 Rossi JP. 2003. Short-range structures in earthworm spatial distribution. *Pedobiologia* 47:582–
738   587.

739 Rossi JP, Garcia J, Roques A, Rousselet J. 2016. Trees outside forests in agricultural landscapes:
740   spatial distribution and impact on habitat connectivity for forest organisms. *Landscape*
741   *Ecology* 31:243-254.

742 Ruiz-Benito P, Gomez-Aparicio L, Zavala MA. 2012. Large-scale assessment of regeneration and
743   diversity in Mediterranean planted pine forests along ecological gradients. *Diversity and*
744   *Distributions* 18:1092–1106.
745 Serra P, Vera A, Francesc Tulla A, Salvati L. 2014. Beyond urbanerural dichotomy: exploring
746   socioeconomic and land-use processes of change in Spain (1991-2011). *Applied Geograpy*
747   55:71–81.
748 Sexton JP, Hangartner SB, Hoffmann AA. 2014. Genetic isolation by environment or distance:
749   which pattern of gene flow is most common? *Evolution* 68:1–15.
750 Shirk AJ, Wallin DO, Cushman SA, Rice CG, Warheit KI. 2010. Inferring landscape effects on
751   gene flow: a new model selection framework. *Molecular Ecology* 19:3603–3619.
752 Shoda-Kagaya E. 2007. Genetic differentiation of the pine wilt disease vector *Monochamus*
753   *alternatus* (Coleoptera: Cerambycidae) over a mountain range – revealed from microsatellite
754   DNA markers. *Bulletin of Entomological Research* 97:167–174.
755 Short Bull RA, Cushman SA, Mace R, Chilton T, Kendall KC, Landguth EL, Schwartz MK,
756   McKelvey K, Allendorf FW, Luikart G. 2011. Why replication is important in landscape
757   genetics: American black bear in the Rocky Mountains. *Molecular Ecology* 20:1092–1107.
758 Sjöstedt-de Luna, S., 2001. Resampling non-homogeneous spatial data with smoothly varying
759   mean values. Statistics & Probability Letters 53, 373–379. doi:10.1016/S0167-
760   7152(01)00040-2
761 Storey JD. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical*
762   *Society: Series B* 64:479–498.
763 Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF. 2007. Putting the
764   "landscape" in landscape genetics. *Heredity* 98:128–142.
765 Takasu F, Yamamoto N, Kawasaki K, Togashi K, Kishi Y, Shigesada N. 2000. Modeling the
766   expansion of an introduced tree disease. *Biological Invasions* 2:141–150.
767 Tomminen J. 1993. Development of *Monochamus galloprovincialis* Olivier (Coleoptera,
768   Cerambycidae) in cut trees of young pines (*Pinus sylvestris* L.) and log bolts in southern
769   Finland. *Entomologica Fennica* 4:137–142.
770 Torres-Vila L, Zugasti C, De-Juan JM, Olivia MJ, Montero C, Mendiola FJ, Conejo Y, Sanchez
771   A, Fernandez F, Ponce F, Esparrago G. 2015. Mark-recapture of *Monochamus*
772   *galloprovincialis* with semiochemical-baited traps: population density, attraction distance,
773   flight behavior and mass trapping efficiency. *Forestry* 88:224–236.
774 Tröltzsch K, Van Brusselen J, Schuck A. 2009. Spatial occurence of the major tree species group
775   in Europe derived from multiple data sources. *Forest Ecology and Management* 257:294–
776   302.
777 van Etten J. 2012. GDISTANCE: distances and routes on geographical grids. R package version
778   1.1-4.
779 Van Strien MJ, Keller D, Holderegger R, Ghazoul J, Kienast F, Bolliger J. 2014. Landscape
780   genetics as a tool for conservation planning: predicting the effects of landscape change on
781   gene flow. *Ecological Applications* 24:327–339.

782 Waters JM, Fraser CI, Hewitt GM. 2013. Founder takes all: density-dependent processes structure
783   biodiversity. *Trends in Ecology and Evolution* 28:78–85.
784 Wagner HH, Holderegger R, Werth S, Gugerli F, Hoebee SE, Scheidegger C. 2005. Variogram
785   analysis of the spatial genetic structure of continuous populations using multilocus
786   microsatellite data. *Genetics* 169:1739–1752.
787 Zalewski A, Piertney SB, Zalewska H, Lambin X. 2009. Landscape barriers reduce gene flow in
788   an invasive carnivore: geographical and local genetic structure of American mink in
789   Scotland. *Molecular Ecology* 18:1601–1615.
790 Zeller KA, McGarigal K, Beier P, Cushman SA, Winston Vickers T, Boyce WM. 2014.
791   Sensitivity of landscape resistance estimates based on point selection functions to scale and
792   behavioral state: pumas as a case study. *Landscape Ecology* 29:541–557.
793 Zeller KA, McGarigal K, Whiteley AR. 2012. Estimating landscape resistance to movement: a
794   review. *Landscape Ecology* 27:777–797.
795 Zellmer AJ, Knowles LL. 2009. Disentangling the effects of historic vs. contemporary landscape
796   structure on population genetic divergence. *Molecular Ecology* 18:3593–3602.

Appendices

797

Appendix 1: R script detailing the approach used in this study.

798

```
799   # Simplified version of the script used in this study. Provides an overview of the general method employed.
800   #------------------------------------------------------------------------
801   # create and plot background matrix with artificial barrier in middle
802   m <- matrix(1, nrow=10, ncol=10) ; m
803   m[,5] <- 4
804
805   library(raster)
806   r <- raster(m)
807   plot(r)
808
809   # create and plot transition matrix
810   library(gdistance)
811   t <- transition(r, transitionFunction=mean, 4, symm=TRUE, intervalBreaks=3)
812   plot(raster(t))
813
814   # create and plot sampling points and associated genetic data.
815   # (x coordinates, y coordinates, genetic data for 3 loci)
816   matG2 <- matrix(c(0.21, 0.22, 0.82, 0.23, 0.81, 0.83, 0.81, 0.21, 0.50, 0.51, 0.23, 0.83, 0, 0, 2, 0, 1, 1, 1, 2, 1, 1,
817   1, 0, 2, 1, 0, 1, 0, 0), ncol=5)
818   xcoord <- matG2[, 1] ; ycoord <- matG2[, 2]
819   P <- cbind(xcoord,ycoord)
820   points(P)
821
822
823   # construction of moving windows (sampling areas)
824   library("ade4") ; library("vegan")
825
826   # Define the extent of sampling areas and the interval wanted
827   Min <- 0.7 # Minimum radius of areas wanted
```

```
828    Max <- 0.9 # Maximum radius of areas wanted
829    Step <- 0.1 # interval wanted
830
831    # Loops to test correlations in sampling area on multiple scales and in multiple locations
832    resultsfinal <- cbind(1,1,1,1,1)
833    colnames(resultsfinal) <- c("xcoord","Ycoord", "Radius", "MantelR", "Pval")
834    for(Radius in seq(Min, Max, by = Step)){
835      results = NULL
836      for(i in 1:length(xcoord)){
837        Xcircle <- ( xcoord [i] + Radius*cos(seq(0,2*pi,length.out=100)))
838        Ycircle <- ( ycoord [i] + Radius*sin(seq(0,2*pi,length.out=100)))
839        polygon(Xcircle, Ycircle)
840
841        # extract individual data in each sampling are constructed
842        expr <- point.in.polygon(xcoord,ycoord,Xcircle,Ycircle)
843        xcoord[expr==1]
844        ycoord[expr==1]
845        coordPoly <- cbind (xcoord[expr==1],ycoord[expr==1])
846
847        # sort data and compute matrix of basic pairwise euclidian distances (not used further in this example)
848        CoordOrder<- coordPoly[order(coordPoly[,1],decreasing=FALSE),]
849        locOrder<-data.frame(CoordOrder)
850        DisGeoEucl<-dist(locOrder, method = "euclidean", diag = TRUE, upper = TRUE)
851
852        # compute corresponding matrix of genetic distances
853        listcoord = (1:6)[expr==1]
854        Genet = NULL ## fichier vide pour collage des données
855
856        for(h in listcoord){
857          tmp <- matG2[(matG2[, 1]==xcoord[h])and(matG2 [, 2]== ycoord[h]), ]
858          Genet = rbind(Genet,tmp)
859               }
860      GenetOrder<- Genet[order(Genet[,1],decreasing=FALSE),]
861      GenetOrderSanscoord <- GenetOrder[,-c(1,2)]
862      MatdistGenet<- vegdist(GenetOrderSanscoord, method="bray", binary=FALSE, diag=FALSE, upper=TRUE, na.rm = TRUE)
863      MatdistGenet <- as.dist(MatdistGenet)
864
865      # Compute matrix landscape "resistance" distances based on raster
866      spatiallocX <-  locOrder[,1] ## extraction des colonnes pour repasser en spatial
867      spatiallocY <-  locOrder[,2]
868      SpaLoc <- SpatialPoints(cbind(spatiallocX, spatiallocY))
869      Resdis<- commuteDistance(t, SpaLoc)
870      Resdis<-as.dist(Resdis, diag = TRUE, upper=TRUE)
871
872      # simple mantels test between genetic and landscape "resistance" distances
873      MantelpRes <- mantel.rtest(MatdistGenet, Resdis, nrepet = 99)
874      results <- rbind (results, cbind (xcoord [i], ycoord [i],Radius, MantelpRes[2], MantelpRes[4]))
875        }
876       resultsfinal <- rbind(resultsfinal,results)
877     }
878
879    # display result file with for each individual: x and y coordinates, radius of sampling area, mantel output and associated p-value
880    Resultsfinal
881
```

882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902

Supplementary Material

903

904 **Table S1**: Sampling details for the 132 demes. (Long. and Lat. refer to the geographic coordinates of the
905 sampling sites; N. is the number of individuals of demes; A. mean allelic richness; AR. corrected allelic
906 richness, accounting for variation in deme size; F_{is}. F_{is} estimate of deme, computed without Mon01 and
907 Mon 27)

908 **Table S2**: Details of primer sequence and genotyping.

909 **Protocol**
910 Multiplexed PCR were performed in a 10 µL reaction volume using 25 ng of genomic DNA, 0.4 U of
911 DreamTaq DNA Polymerase (Thermo Scientific®), 0.75 µL Dream Taq Green Buffer (including 20 mM
912 MgCl2, Thermo Scientific®), 1 µM Betaine, 0.24 µL dNTP (10 µM) and deionized H2O. PCR
913 amplifications were run on a Veriti® 96 well fast thermal cycler (Applied Biosystems®) using the following
914 settings: a first denaturation step at 95°C for 10 min; 40 cycles of denaturation (30 s at 95°C), hybridization
915 (30 s at 55°C) and elongation (1 min at 72°C), and a final elongation step at 72°C for 10 min. One µL of
916 PCR products was denatured in a mix of 10 µL of formamide and 0.3 µL of 600 Liz marker before being
917 run on an ABI PRISM 3500 sequencer (Life Technologies®). Genotypes were read using GENEMAPPER
918 V 4.1 software (Applied Biosystems®).

919 **Table S3**: Sampling locations and microsatellite genotypes.

920 **Figure S1**: Distribution of sampling sites in the Iberian Peninsula. Black dots refer to populations of size
921 > 19 individuals. The green background refers to elevation (from pale to dark green: low to high
922 elevation).

923 **Figure S2**: Number of individuals in sampling areas across the spatial scale (Mean: black; +/- SD: gray).

924 **Figure S3**: Evolution of DeltaK for an increasing value of K (2 -20).

925

Figure 1

Genetic clustering of 992 individuals of *Monochamus galloprovincialis* sampled at 132 locations

A: Assignment of individuals to clusters based on a STRUCTURE analysis for K=2. **B**: Assignment of demes to clusters for k=2, displayed in geographic context (Iberian Peninsula, the size of the pies refers to the size of the demes). **C**: PCA of individuals on first and second axes, colors and ellipses refer to demes.

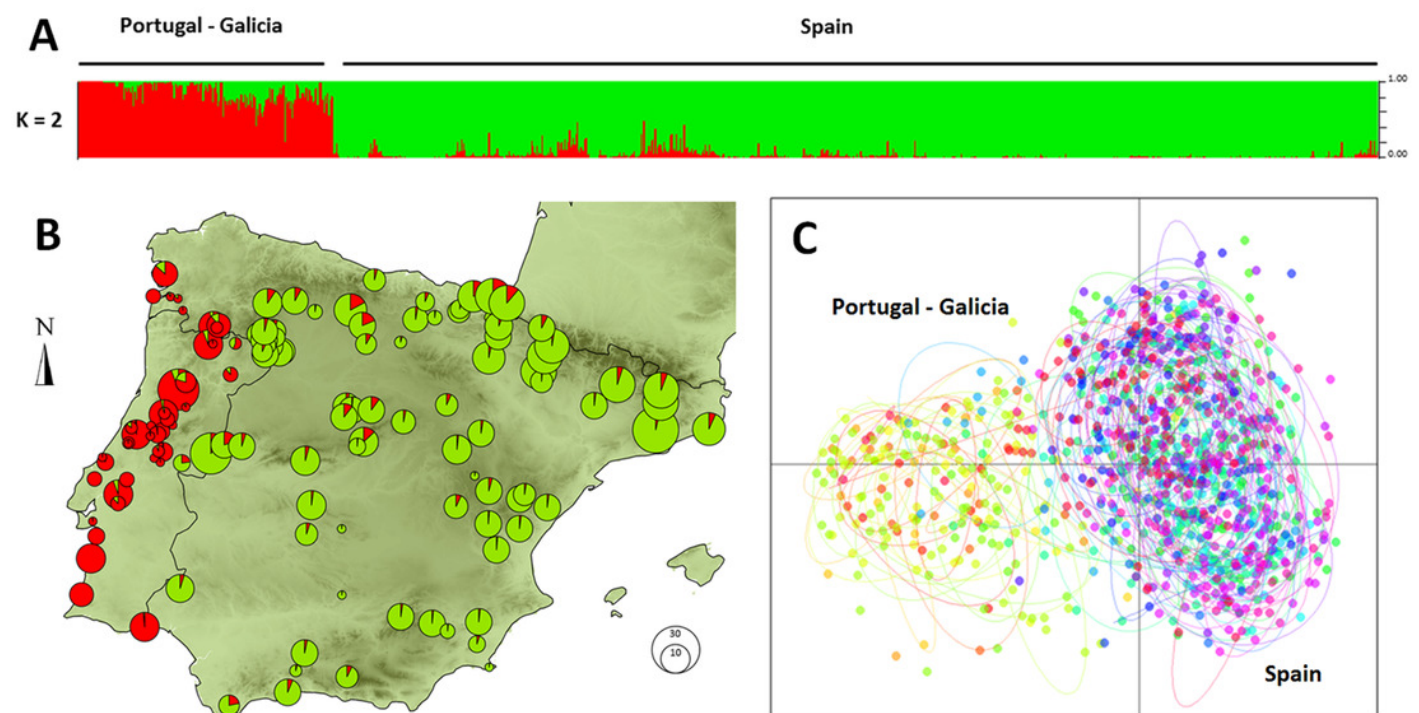



Figure 2

Empirical semi-variogram of genotypes of *Monochamus galloprovincialis*.

The variogram was fitted with an exponential model to highlight the first plateau. Data points are shown with a spatial lag distance of 50 km.

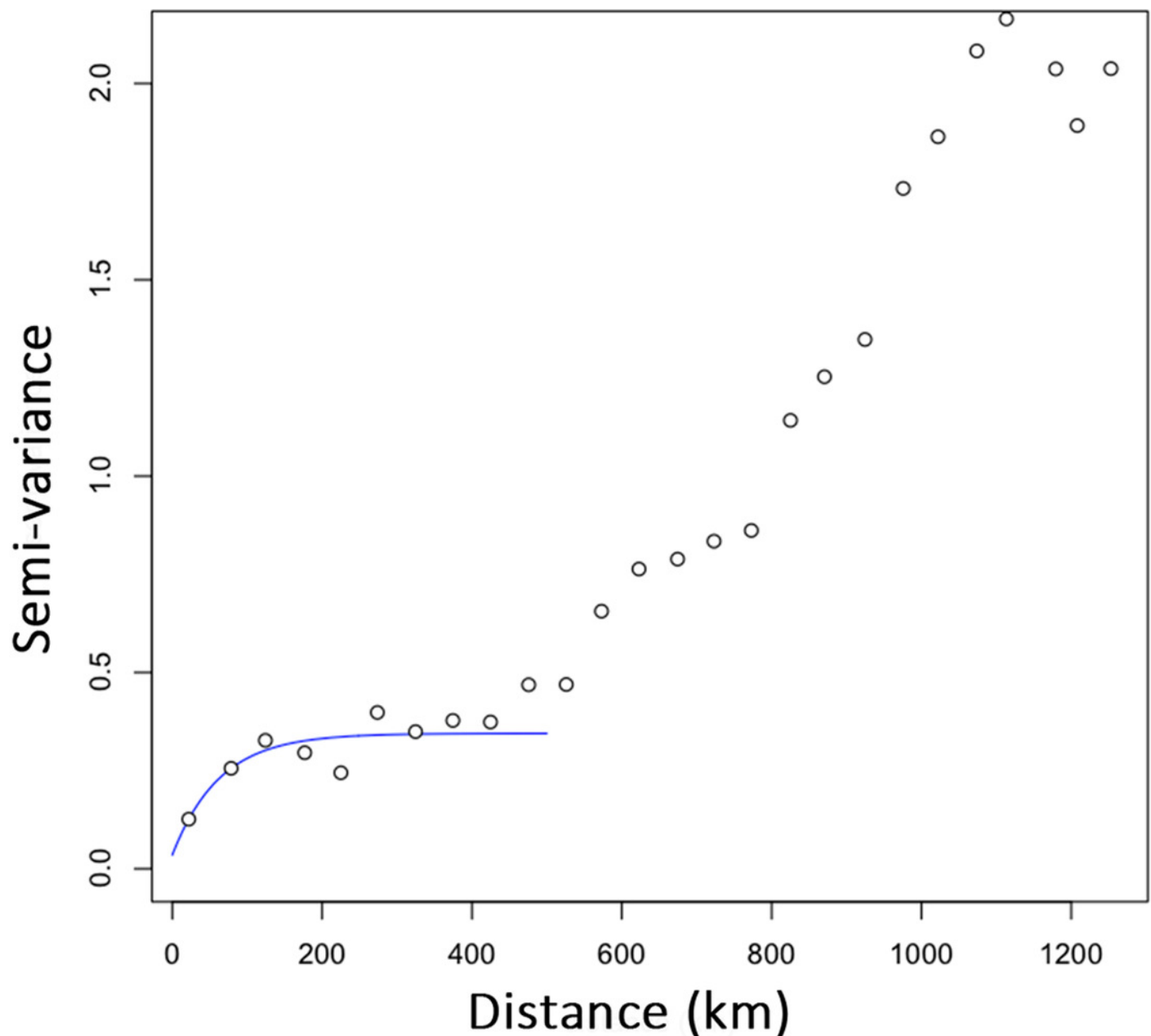



Figure 3

Percentage of areas with supported IBR hypotheses for Mantel tests (A) and mean partial Mantel r (B) of areas with support for the IBR hypotheses (p<0.05) with increasing scales (whole dataset).

E: Elevation, *T*: Mean minimum temperatures, *Pr* and *Pc*: pine densities as a resistant feature and as a corridor respectively, IBD: Isolation by distance.

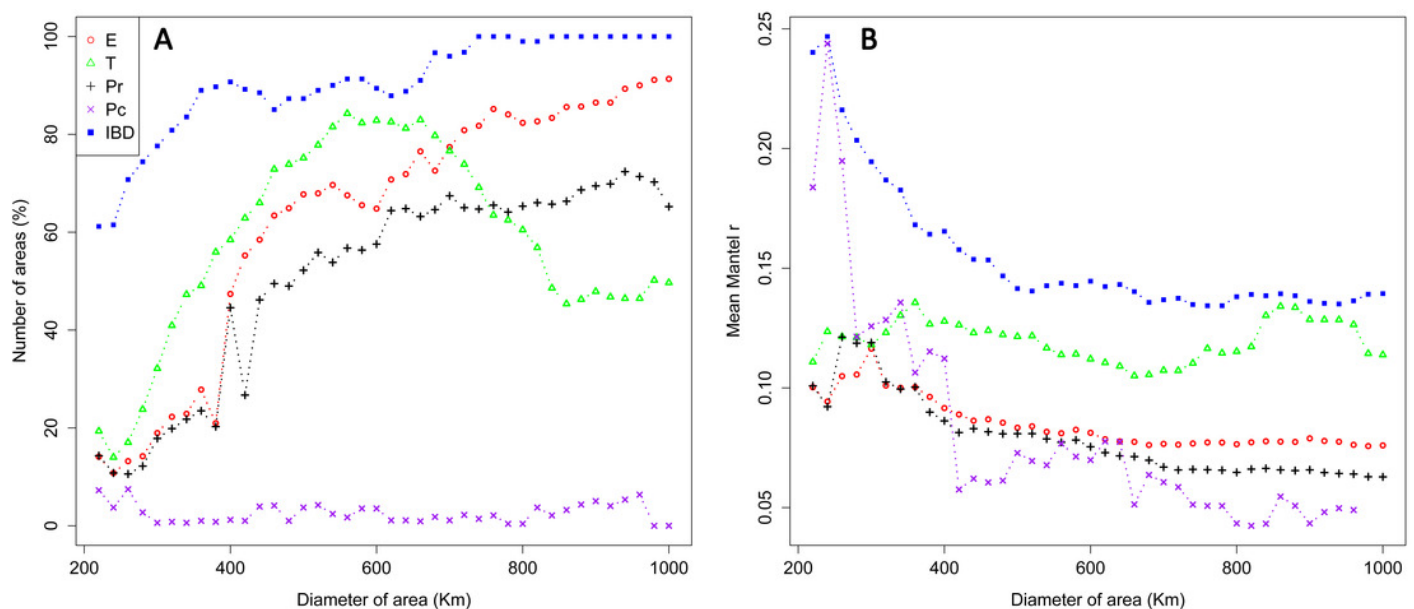



Figure 4

Distribution of supported IBR hypotheses through Mantel tests for the environmental features tested

(IBD: Euclidian distances; *T*: mean minimum temperatures; *E*: elevation; *Pr*: high pine densities as barriers; *Pc*: high pine densities as corridors). Gray maps (A) refer to the distribution of environmental features associated with resistance models (From white to black: low to high resistance values). Colored maps refer to interpolations of supported IBR hypotheses on the whole dataset (B) and within the western Iberian cluster only (C). From blue to red: low to high frequency of study areas with supported resistance models.

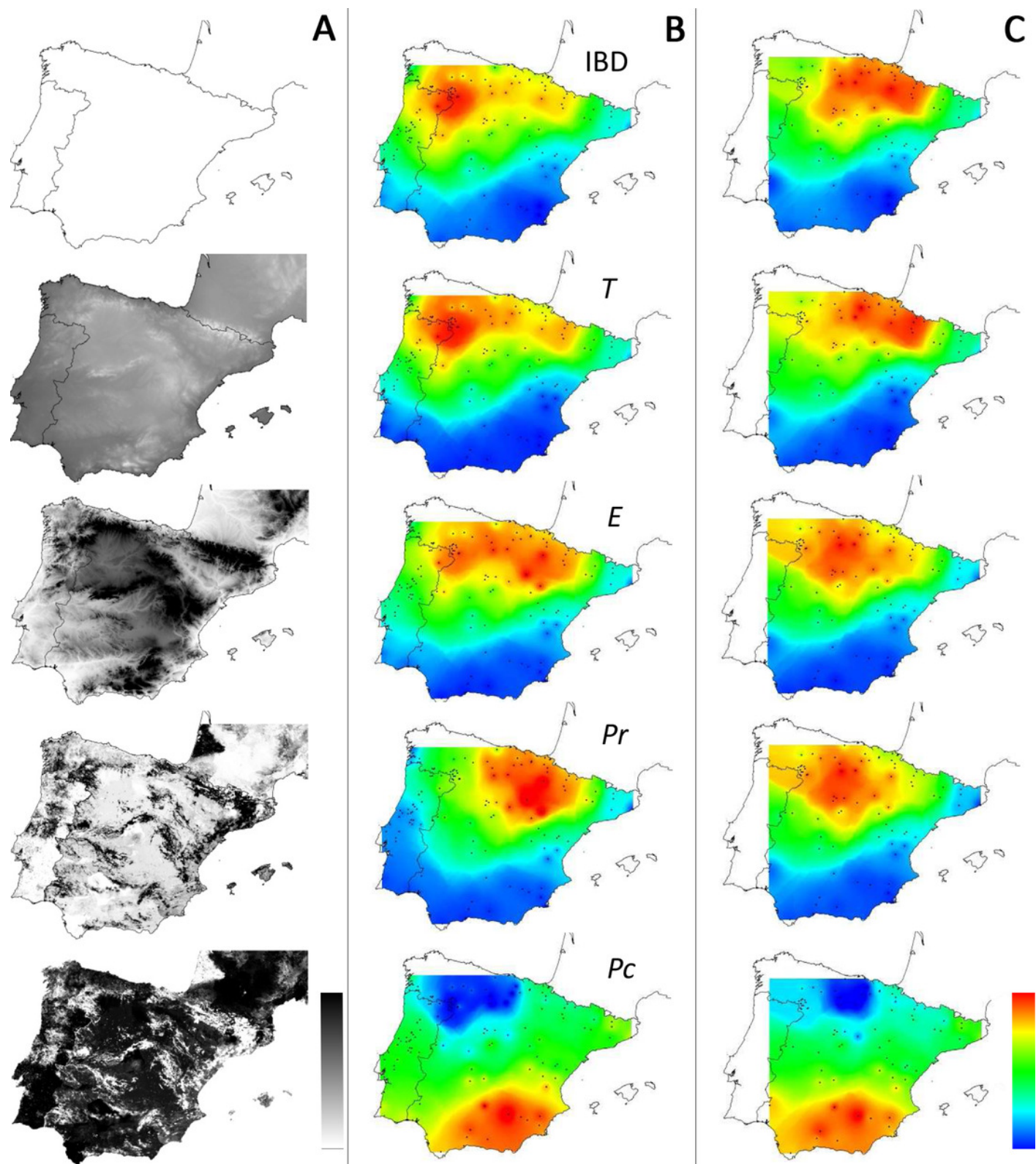
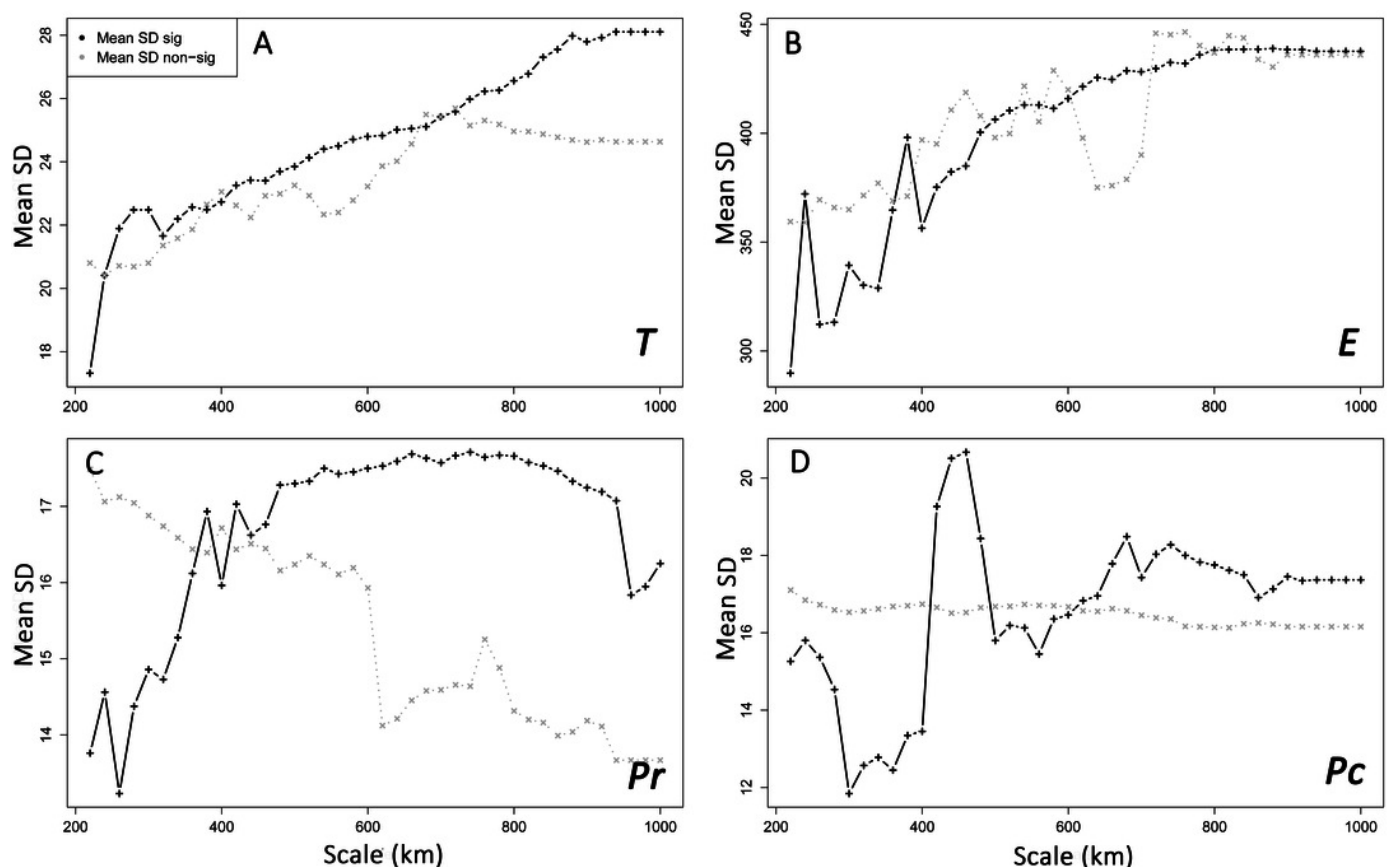



Figure 5

Spatial heterogeneity (mean standard deviation, SD) of environmental features in areas with supported and non-supported resistance hypotheses through Mantel tests with increasing scales.

Mean SD: mean standard deviation. *T*: mean minimum temperatures (A); *E*: elevation (B); *Pr*: high pine densities as barriers (C); *Pc*: high pine densities as corridors (D). sig: significant, non-sig: non-significant.

Commonality coefficients of both unique and common effects for the three sampling areas with the highest variance explained.

Code pop: population code for the center of the sampling area. *Scale*: diameter of sampling area (km). *N*: number of individuals in sampling area. *Coef.*: percentage of variance explained by environmental features (IBR hypotheses). *% Total*: percentage of the contribution of environmental features to the total variance explained. Environmental features and their associated Isolation By Resistance (IBR) hypothesis: *E*: elevation model (high elevations = resistance to dispersal). *Pc*: Pine density model (high pine density = corridors to dispersal. *Pr*: reversed pine density model (high pine density = resistance to dispersal). *T*: temperature model (low minimum annual temperatures = resistance to dispersal).

	Code pop	85		130		131	
	Scale	620		540		520	
	N	225		254		244	
1st-order	**IBR hypotheses**	**Coef.**	**% Total**	**Coef.**	**% Total**	**Coef.**	**% Total**
	E	0,008	3,408	0,001	0,351	0,002	0,807
	T	0,050	20,775	0,070	32,651	0,059	28,108
	Pc	0,004	1,806	0,008	3,678	0,008	3,811
	Pr	0,085	35,235	0,047	21,817	0,046	22,214
2nd-order	**E,T**	0,136	56,426	0,117	54,314	0,115	54,953
	E,Pc	-0,003	-1,255	-0,001	-0,243	-0,001	-0,430
	T,Pc	0,001	0,426	0,007	3,236	0,010	4,893
	E,Pr	-0,002	-0,897	0,031	14,375	0,020	9,721
	T,Pr	-0,013	-5,398	0,018	8,406	0,014	6,540
	Pc,Pr	0,024	10,049	0,008	3,579	0,011	5,333
3rd-order	**E,T,Pc**	0,014	5,724	-0,005	-2,311	-0,001	-0,474
	E,T,Pr	-0,023	-9,712	-0,069	-32,007	-0,047	-22,356
	E,Pc,Pr	0,003	1,313	0,026	12,118	0,030	14,509
	T,Pc,Pr	-0,009	-3,795	0,006	2,961	0,010	4,950
	E,T,Pc,Pr	-0,034	-14,106	-0,049	-22,926	-0,068	-32,581
	Sum	0,240	100	0,216	100	0,209	100