Multi-scale and multi-site resampling of 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 UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Cirad, Montpellier, France
3 UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), INRA, Montpellier, France
4 Sustainable Forest Management Res Inst, Universidad de Valladolid, Palencia, Spain
5 Instituto Nacional de Investigigacao 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 to generalize the patterns of environmental effects on gene flow. It allows reducing the variability of inference, accounting for multiple scales and locations of study areas. Although several reviews have stressed the importance of this point, few studies have considered multiple sampling areas in analysis and formally tested their effects on inference. In this study, we present a method for resampling of study areas at multiple scales and multiple locations (sliding windows) to track the variation of inference in spatial genetics. We explored the effects of environmental features on gene flow of a flying long-horned beetle (*Monochamus galloprovincialis*) in $3 \times 10^4$ study areas ranging in scale from 220 to 1000 km and spread over 132 locations among the Iberian Peninsula. We show that there were no general or recurrent effects of environmental features detected among scales and locations, independent of variation in environmental features. Detection of environmental features on gene flow generally increased with an increasing scale of study, and was variable between locations. The resampling method presented here provides the opportunity to explore the effects of environmental features on gene flow of organisms in their whole extent and to conclude about general landscape effects on the dispersal of organisms, while keeping sampling effort to a reasonable level.
Multi-scale and multi-site resampling of study area in spatial genetics: implications for flying insect species

JULIEN HARAN\textsuperscript{1,2,7}, JEAN-PIERRE ROSSI\textsuperscript{3}, JUAN A. PAJARES\textsuperscript{4}, LUIS BONIFACIO\textsuperscript{5}, PEDRO NAVES\textsuperscript{5}, ALAIN ROQUES\textsuperscript{1} and GÉRALDINE ROUX\textsuperscript{1,2}

\textsuperscript{1} INRA, UR633 Zoologie Forestière, F-45075 Orléans, France.
\textsuperscript{2} Université d’Orléans, rue de Chartres, Orléans cedex, France.
\textsuperscript{3} INRA, UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Montpellier, France.
\textsuperscript{4} Univ Valladolid INIA Palencia, Sustainable Forest Management Res Inst, Palencia, Spain.
\textsuperscript{5} Instituto Nacional de Investigacao Agraria e Veterinaria – INIAV Oeiras, Portugal.
\textsuperscript{7} Present address: Cirad, UMR CBGP, Montpellier, France.

\textbf{Corresponding author}: Julien HARAN (Cirad, UMR CBGP, Montpellier, France. Tél. : +33 4 99 62 33 04   Fax. : +33 4 99 62 33 45, julien.haran@cirad.fr)

\textbf{Keywords}: gene flow, landscape genetics, insect dispersal, \textit{Monochamus galloprovincialis}, Iberian Peninsula.
Abstract

The use of multiple sampling areas in landscape genetic analysis has been recognized as a useful way to generalize the patterns of environmental effects on gene flow. It allows reducing the variability of inference, accounting for multiple scales and locations of study areas. Although several reviews have stressed the importance of this point, few studies have considered multiple sampling areas in analysis and formally tested their effects on inference. In this study, we present a method for resampling of study areas at multiple scales and multiple locations (sliding windows) to track the variation of inference in spatial genetics. We explored the effects of environmental features on gene flow of a flying long-horned beetle (*Monochamus galloprovincialis*) in $3 \times 10^4$ study areas ranging in scale from 220 to 1000 km and spread over 132 locations among the Iberian Peninsula. We show that there were no general or recurrent effects of environmental features detected among scales and locations, independent of variation in environmental features. Detection of environmental features on gene flow generally increased with an increasing scale of study, and was variable between locations. The resampling method presented here provides the opportunity to explore the effects of environmental features on gene flow of organisms in their whole extent and to conclude about general landscape effects on the dispersal of organisms, while keeping sampling effort to a reasonable level.

Introduction

Landscape genetics examines the relationship between landscape and environmental features and genetic structure (Manel et al., 2003; Manel & Holderegger 2013). It allows which
environmental features facilitate or hinder gene flow to be inferred (Zeller et al., 2012), which is a key factor for understanding the persistence and evolution of species and populations and has significant consequences for conservation planning (Castillo et al., 2014; Van Strien et al., 2014). As an emerging and fast moving field, the landscape genetic toolbox is far from being established, and an important effort toward method optimization is still required to make relevant and optimal inferences (Anderson et al., 2010; Cushman et al., 2013; Manel & Holderegger 2013). Landscape genetic analyses are usually conducted at a single scale and in a single location (Zeller et al., 2012) and therefore provide results that are strictly speaking only applicable to the particular area under study. Indeed, genetic structure is determined by multiple micro- and macro-evolutionary processes acting at different spatial and temporal scales, rarely homogeneously distributed across a study species’ distribution range (Waters et al., 2013). For example, in addition to contemporary or historical environmental effects on dispersal (Zellmer & Knowles 2009), the genetic structure of organisms is often influenced by historic differentiation due to quaternary climate oscillations (Hewitt 2000), or by biased dispersal due to local adaptation to specific environmental conditions (Sexton et al., 2014; Pflüger & Balkenhol 2014). The diverse factors acting at different temporal and spatial scales may generate genetic patterns that could be inconsistent across locations or regions, which results in conflicting signals of environmental factors acting on gene flow. This may drastically impede the ability to infer the general drivers of gene flow. To overcome this problem, several authors have pointed to the importance of matching study design to the process investigated (Anderson et al., 2010; Cushman & Landguth 2010; Galpern et al., 2012; Keller et al., 2013), or have stressed the need to consider landscape-level replications in landscape genetic analysis (Holderegger & Wagner 2008; Short Bull et al., 2011).

The scale of study is fundamental in landscape genetics, because species respond to environmental features at a continuous range of scales (Anderson et al., 2010; Manel & Holderegger 2013). This point has been highlighted in several empirical studies and simulation exercises (Cushman & Landguth 2010; Angelone et al., 2011; Galpern et al., 2012; Dudaniec et al., 2013; Keller et al., 2013), in particular for organisms exhibiting wide home-ranges like large mammals (Galpern et al., 2012; Zeller et al., 2014). Despite an increasing number of studies explicitly accounting for scale effects, landscape genetics studies still rarely consider scale effects (Zeller et al., 2012) and how it affects inference on the detection of general effects of environmental features on dispersal and gene flow. Landscape-level replication is another
fundamental aspect in landscape genetics. The term replication usually refers to the replication of sampling areas (sampling units; Short Bull et al., 2011). Such experimental design provides a “quantitative” dimension in landscape genetics analysis, allowing conclusions to be drawn about the effects of landscape features on the dispersal of organisms. Few studies have included replication in landscape genetics studies (Driezen et al., 2007; Kindall & Van Manen 2007; Zalewski et al., 2009; Short Bull et al., 2011), and the number of replications considered is often low due to the sampling effort required.

There is currently an increasing demand to provide a more complete and comprehensive picture of the general landscape effects on the dispersal of organisms, including variation across scale and locations. However, such exploration often remains often limited due to the important sampling efforts required. In the present study, we assess a method to unify both dimensions in spatial genetic analysis. This method consists of a multi-site and a multi-scale resampling of sliding windows (study areas) and therefore has the potential to reduce the versatility of results while keeping sampling effort to a reasonable level.

As a study system, we explored which environmental features foster or hinder gene flow of a flying insect species, Monochamus galloprovincialis (Coleoptera, Cerambycidae). We performed an individual-based landscape genetic analysis among 3*10^4 resampled areas of extent ranging from 220 to 1000 km and distributed in 132 sampling locations in the Iberian Peninsula. M. galloprovincialis is the vector of the pinewood nematode (Bursaphelenchus xylophilus, PWN) in Europe. This species is native to Europe and is structured into several genetic clusters that are thought to correspond to postglacial recolonization patterns (Koutroumpa et al., 2013, Haran et al., 2015). The life cycle of this beetle occurs in the wood of declining pine trees (Pinus pinaster, P. sylvestris, P. nigra, P. halepensis, Naves et al., 2006; Hellrigl 1971). M. galloprovincialis is quite long-lived and shows rather high potential to dispersal in laboratory experiments and in the field (David et al., 2013; Mas et al., 2013). However, the role of major environmental features and parameters (elevations, low temperatures and the density of pine cover) as potential barriers to the dispersal of this species has been weakly explored and remains poorly understood (Haran et al., 2015; Torres-Vila et al., 2015).

Methods:
**Sampling and genotyping**

The study area covered the entire Iberian Peninsula (582,000 km²) with altitudes ranging from sea level up to 2444 m. *M. galloprovincialis* specimens were trapped between 2011 and 2013 at 137 sites spread over the Iberian Peninsula. We used multifunnel traps baited with a volatile attractant (Galloprotect, SEDQ, Spain) placed during the summer to catch flying adults. The traps used had a radius of attraction of 100m (Jactel et al., 2015) and were placed in dense pine stands (were beetle density is high; Jactel et al., 2015) to limit consanguinity among individuals caught. After collecting, adults were stored in 96.66% ethanol at 4°C. Despite intensive trapping, *M. galloprovincialis* was not recorded in five localities in the Central lowlands of Castilla y Leon, central Galicia and Asturias districts. We obtained a sampling of 1050 individuals at 132 sites. Seventy-seven sites had a size below 10 and 55 above or equal to this value, with an average sampling size per location of 7.68 individuals. Details of sampling localities and year of collection are given in table S1 (supporting information). Individuals collected at the same locality were considered as one deme. The distribution of sites covered most of the pines forests found in the Iberian Peninsula (Fig. S1; supporting information).

DNA was isolated from two legs per individuals using a Nucleospin Kit (Macherey-Nagel, Düren, Germany). Specimens were genotyped at 12 microsatellite loci (Mon01, Mon08, Mon17, Mon23, Mon27, Mon30, Mon31, Mon35, Mon36, Mon41, Mon42 and Mon44) following the method of Haran & Roux-Morabito (2014). Details of primer sequences and the protocol for genotyping are given in Table S2 (supporting information). Results showing negative or ambiguous amplification of particular loci were repeated once and considered null when still unsatisfactory. Individuals exceeding two missing loci were removed for analysis. Deviation from Hardy Weinberg Equilibrium ($F_{is}$) was estimated for each deme, each inferred cluster and for the whole dataset using GENEPOP 4.2 (Raymon & Rousset 1995). The frequency of null alleles at each locus was tested using FREENA (Chapuis & Estoup 2007) among three large size demes (n>19). Loci exceeding a rate of 7% of null alleles across populations were discarded from further analysis. The allelic richness was computed for each deme using rarefaction (HP-RARE, Kalinowski 2005). The absence of linkage disequilibrium between pairs of loci was already reported in a previous population-based study (Haran et al., 2015).
Genetic structure

We used the Bayesian approach implemented in STRUCTURE 2.3.4 (Pritchard et al., 2000) to identify the main genetic clusters among Iberian demes. STRUCTURE assigns individuals to a predefined number of clusters based on allelic composition and linkage disequilibrium. We used the Delta K method (Evanno et al., 2005) to determine the number of clusters (K) that best fit the data. Genotypes were analyzed using default parameters (admixture model, correlated alleles frequency). We made ten repeats of a 200,000 burn-in period followed by 500,000 replicates of Markov Chain Monte Carlo (MCMC), for K values ranging from 1 to 20. Results were uploaded in STRUCTURE HARVESTER (Earl et al., 2012) to determine the optimal K. We also explored the existence of genetic clusters among demes using a principal component analysis (PCA) performed on allele frequencies (package Adegenet, Jombart 2008). To account for potential confounding effects of differentiated genetic clusters (possibly of evolutionary history origin) on the inference of gene flow, landscape genetic analyses were performed twice, once within the main cluster identified by STRUCTURE and PCA, and once with the whole dataset including all clusters.

The scores of sampling locations upon axis 1 of the PCA are linear descriptors of the allele frequencies and, as such, can be used as a univariate statistical measure of genetic composition. The scores may encapsulate relevant spatial information, so we explored this point using a specific tool borrowed from geostatistics: the variogram (Wagner et al., 2005, Goovaerts, 1997). The variogram is used in all branches of life sciences in order to explore spatial patterns and determine the main spatial scales at which structures occur. In the present study, we analyzed the score of sample points upon axis 1 using a variogram to better understand the spatial component of the variation encapsulated in the first axis of the PCA. Let \( z(u_\alpha) \), with \( \alpha = 1, 2, \ldots, n \), be a set of \( n \) values of sample scores upon a PCA axis where \( u_\alpha \) is the vector of spatial coordinates of the \( \alpha \)th observation. In geostatistics, spatial dependence is described in terms of dissimilarity between observations expressed as a function of the separating distance (Goovaerts 1997). The average dissimilarity between data separated by a vector \( h \) is measured by the empirical semi-variance \( \hat{\gamma}(h) \), which is computed as half of the average squared difference between the data pairs:

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

(1)
where \( N(h) \) is the number of data pairs for a given lag vector \( h \), \( z(u_0) \) and \( z(u_0 + h) \) the score values of all sample locations separated by a vector \( h \). The more alike the observations at points separated by \( h \) are, the smaller \( \hat{y}(h) \), and vice versa. The plot of \( \hat{y}(h) \) against \( h \) is called a variogram and represents the average rate of change of \( z \) with distance. Its shape describes the pattern of spatial variation in terms of general form, scales and magnitude (Goovaerts 1997).

Variograms are good tools to depict spatial structures and analyze nested patterns (Burrough 1983); when structures occurs at different spatial scales, the resulting variogram exhibits different plateaus in association with different scales (Robertson and Gross 1994; Rossi 2003). The range of the variogram is the distance at which the plateau occurs. Multi-plateau variograms exhibit different ranges which provide synthetic information about the spatial scales at play. Readers are referred to Goovaerts (1997) for a thorough introduction to variograms and geostatistics and to Wagner et al., (2005) for an introduction of this tool in the field of population genetics. Variograms were computed using the R package geoR (Ribeiro and Diggle 2001).

**Landscape genetics analysis**

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

We selected the environmental features considered to be the most likely to influence the dispersal of *M. galloprovincialis* given the existing knowledge of species requirements. Apart from Euclidian geographic distances (null model), we considered three environmental features to be potential drivers of dispersal (pine density, temperatures and elevation).
We modeled environmental resistance as a function of pine density as this parameter determines the volume of resource available for the *M. galloprovincialis* and is thought to affect its foraging dispersal. As the dispersal behavior of this beetle in reaction to pine density is not known, we modeled this parameter according to two alternative scenarios. (1) High pine densities are positively correlated with beetle dispersal. In this scenario, a dense pine cover represents a corridor for dispersal due to the high amount of resources available. Conversely, a low pine density would represent a barrier. (2) High pine densities are negatively correlated with beetle dispersal. For this second scenario, it was assumed that a dense pine cover provides sufficient resources for local populations, which would therefore not need to disperse. This scenario assumes increased dispersal in low pine cover areas. To model resistance based on pine density, we considered the sum of densities of all pine species encountered in a grid cell. Indeed, in the Iberian Peninsula, *M. galloprovincialis* is performing its life cycle in stressed or fresh dead wood of the most widespread pine species: *Pinus pinaster, P. nigra, P. sylvestris, P. halepensis* and *P. radiata* (Hellrigl 1971; Naves et al., 2006), and shows no specialization for any of these host species (Haran et al., 2015).

Resistance was modeled as a function of mean minimum temperatures, as low summer temperatures tend to inhibit adults flying activity (Hernández et al., 2011), and because low winter temperatures are likely to determine survival or the development rate of larval instars in *M. galloprovincialis* (Naves & Sousa 2009). This species performs its larval phase during winter, and instars may stop their development and eventually die after exposure to extended periods of cold temperatures (Naves and Sousa 2009). As no precise threshold is known for both flying activity and larvae survival, we consider that resistance increases when the annual mean minimum temperatures decrease. Elevation is often a proxy for temperature. We hypothesized that resistance to dispersal increases when elevation increases. We kept temperatures and elevation as distinct environmental features for the analysis, because temperature and altitude may not co-vary similarly at large scales (North to South of Spain; for collinearity see below). A summary of resistance scenarios of environmental features is given in Table 1.

Resistance distances were computed using the package gdistance (van Etten 2012). Raster layers of environmental features were imported at a resolution of 10 x 10 km. Such resolution was chosen because the mean flight distance of *M. galloprovincialis* reaches 16 km, based on flight mills experiments (David et al., 2014). Temperature data (1950-2000) were downloaded from
Hijmans et al., (2005; http://www.worldclim.org; original resolution: 1 x 1 km), the pine density from Tröltzsch et al., (2009; http://www.efi.int/; original resolution: 1 x 1 km) and elevation from ARCGIS 9.3 (ESRI, Redlands, CA, USA; original resolution: 1 x 1 km). For control purpose, resistance distances were also measured on layers with a resolution of 1 x 1 km. As temperature, elevation and pine density are continuous parameters, we did not assign particular resistances to particular values, but directly used the values (except for the $P_c$ hypothesis for which values were set as negative). Pairwise resistance distances were estimated based on random walk probabilities (Chandra et al., 1997, McRae 2006) and computed using the command commuteDistance (package gdistance). Resistance distances were chosen instead of least cost distances (LCD) because they are thought to be more reliable biologically and produce fewer artifacts over long distances (McRae 2006). We constructed a semi matrix of resistance distance between each pair of individuals. Values were normalized to a common scale for further analysis. Collinearity was estimated using the variance inflation factor (VIF) based on the formula $VIF = 1/(1-R^2)$, where $R^2$ is the r-squared value of regression between variables. VIF values > 10 are usually considered evidence for collinearity between environmental features (O’Brien 2007). We did not detect collinearity between environmental features over the whole area of study (VIF < 1 for all pairwise comparisons).

We tested correlation between the response (genetic distances matrix, G) and resistance distances (resistance matrices; Isolation By Resistance: IBR) and geographic distances (Euclidian geographic distance; Isolation By Distance: IBD) using partial Mantel tests (Cushman & Landguth 2010). Partial Mantel tests measure association between two distances matrices while partialling out a third distance matrix. We first used simple Mantel tests to correlate IBD with G. We then tested the effect of IBR in partial Mantel tests. Support for IBR was considered when: (1) IBR should be significantly correlated to G after partialling out IBD ($p < 0.05$) and IBD should be non-significant with IBR partialled out ($p \geq 0.05$; Cushman et al., 2006). Mantel and partial Mantel tests were performed using the vegan package with $10^3$ permutations. This approach is widely used in the field of landscape genetics (Cushman et al., 2006; Cushman and Landguth 2010; Galpern et al., 2012; Castillo et al., 2014) and has been shown to efficiently infer the drivers of gene flow (Cushman & Landguth 2010b). However, partial Mantel tests have received criticism regarding their statistical performance (Guillot & Rousset 2013; Diniz-Filho et al., 2013), and are therefore preferably used together with complementary approaches such as ordination methods (Kierepka...)
et al., 2015). To overcome the potential weakness of partial Mantel tests on our dataset, and to validate the statistical significance of correlations, distance matrices were also regressed using commonality analysis (Prunier et al., 2014). This method is based on variance-partitioning and therefore allows the relative importance of the environmental features shaping genetic structure to be estimated, accounting for covariance in the features tested. For commonality analysis, the response G was regressed onto each resistance matrices separate and each combination using the R package yhat (Nimon et al., 2013).

Multiple scales and multiple locations analysis

We considered various spatial scales and various locations in the above landscape genetic analysis by generating nested sampling areas spread over the full extent of the Iberian Peninsula. Sampling areas were constructed as circles of diameters ranging from 220 to 1000 km (steps of 20 km) and centered at each sampling location. Scale dimension was therefore tested only in terms of the extent for this study (Mayer & Cameron 2003). Mantel tests were performed between all individuals found within each area defined. Areas of diameter below 220 km were not included, because it was too small to gather neighboring demes for Mantel tests in the less sampled areas, unbalancing the analysis. We then tracked the evolution of the number of areas with supported IBR hypothesis and the mean significant Mantel r with increasing scale. The geographic distribution of areas with a supported IBR hypothesis was obtained by summing the number of times that each individual was included in a sampling area with IBR hypothesis support among all scales. Obtained numbers (frequencies) were corrected accounting for intrinsic variation due to overlapping sampling areas. Frequencies at each point were interpolated using the Inverse Distance Weighted method (IDW) in ARCGIS 9.3 (ESRI, Redlands, CA, USA) to visualize variation in spatial distribution of areas which supported each IBR hypothesis. Landscape genetics analyses have been shown to perform better in a contrasted landscape (i.e. high amplitudes of values of resistant features; Jaquiéry et al., 2011; Cushman et al., 2013). We extracted resistance values of raster cells within each sampling area and computed the standard deviation (SD) of these values to determine whether support of the IBR hypotheses was due to variation in the environmental features tested. We then calculated mean standard deviation of areas with supported and non-supported IBR hypotheses among the scales of study. Commonality analyses (see above) were
performed within each sampling area generated. As for Mantel tests, we tracked the development of commonality coefficients (percentage of variance explained by unique and cumulated IBR hypothesis) among scales and locations. The sampling area maximizing commonality coefficients was chosen for representation of the relative importance of environmental features in shaping genetic structure. All computations were performed using the R software version 3.0.2 (R development Core Team 2013).

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). Significant heterozygote deficit was detected at four loci (Mon 30, 35, 42, 44). Corresponding null allele frequencies were low (<7%), so these loci were retained. 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). 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; supporting information).

Genetic structure

Individuals formed two clusters under STRUCTURE analysis (Delta K2= 1274.41; delta K3 = 73.41, see Figure S3 in supporting information). 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). 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).
Data points were grouped into 26 distance classes ranging from 0 to 1252 km, with a distance interval of 50 km. The variogram reveals that the first axis of PCA corresponds 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 shape of this variogram is typical of the presence of a long-range spatial variation superimposed over a more local, *i.e.* short-scale genetic structure occurring at scales of 200 to 400 km. For scales below 200 km, the variogram show that genotypes were strongly spatially auto-correlated (*i.e.* non-independent).

**Landscape genetics analysis**

Analyses were conducted both on the whole dataset (992 individuals, 132 localities) and within the Spanish cluster (790 individuals, 87 localities), for a total of 116 and 102 alleles analyzed respectively. Grain sizes of 1 x 1 km and 10 x10 km resulted in similar results. Null distances were not encountered at grain 10 x 10 km, as none of the sampling sites fell with neighbor sites in the same grain. Therefore, only the results obtained for grain 10 x 10 km will be reported below.

Over the whole area of study (whole dataset), we generated a total of 30 576 sampling areas. The mean number of individuals within sampling areas varied from 89.18 (SD: 42.42) at the smallest scale (220 km) to 644.58 (SD: 158.07) at the largest scale (1000 km; Fig. S2, supporting information). Significant effects of environmental features were detected for all IBR hypotheses tested with partial Mantel tests, but the frequency of areas exhibiting an IBR effect varied among scales and locations. The number of areas showing a significant effects of environmental features generally increased with increasing scale (Fig. 3A), but each of the four IBR hypotheses showed a different pattern. Significant effects of environmental features for *E, Pr* and *T* hypotheses were detected in about 15-25% of the areas at smallest scale (220 - 300 km). The frequency of *E* and *Pr* gradually increased to reach 90% and 60% for areas of 1000 km. The frequency of areas with a supported *T* hypothesis increased among scale to reach a peak around 600 km (≈ 80% of areas) and subsequently decreased again. Significant *Pc* hypotheses were encountered at a lower
frequency. The number of positive areas ranged from 0 to 1.11%, for an average number of 4.07
areas for each scale considered. No specific trend was observed when scale increased for the \(Pc\)
hypothesis. Significant isolation by distance (IBD) was observed for \(\approx 60\%\) of areas at smallest
scale. A first plateau of about 85\% of areas was reached for scales ranging between 400 and 700
km, and a second plateau of almost 100\% of areas was reached for scales above 700 km. Mean
Mantel \(r\) for areas with supported IBR hypothesis ranked between 0.05 and 0.25. Best values of
were observed at small scales and generally decreased when scale increased (Fig. 3B). IBR
hypothesis \(T\) showed the best Mantel \(r\) among all IBR hypotheses for scales above 360 Km.

Interpolation of supported IBR hypotheses and IBD was based on areas of scales ranging
from 220 to 600 Km, because most of variation in the detection of effects of environmental features
was found at these scales (Fig. 3A). For most IBR hypotheses \((E, Pr\) and \(T\)) and IBD, effects were
mainly detected in the northern part of the area of study, corresponding to Cantabrian chain and
the western half of the Pyrenees (Fig. 4). In contrast, these IBR hypotheses were the least
frequently detected in a region comprising the eastern side of the Iberic and Betic mountain
systems. For the IBR hypothesis \(Pc\), significant effects were detected mainly in Andalucía, along
the Betic system. Conversely, low or no effects for this hypothesis were detected in the Northern
half of the Iberian Peninsula. The distribution of supported hypotheses was generally similar
between that performed on the whole dataset and on the Spanish cluster only (Fig. 4).

For hypotheses \(E, Pr\) and \(Pc\), the variation of environmental features was lower on average
in areas exhibiting significant effects for scales up to 400 - 600 km (whole dataset; Fig. 5). Above
this scale, the mean standard deviation (SD) of significant areas was either equal, or higher than
the mean SD of non-supported areas. For the \(T\) hypothesis, mean SD of significant areas was above
the mean for non-supported areas for most of the scales.

Regression models gave a maximum explained variance of 24% over all sampling areas
through commonality analysis (Table 2). Best values were obtained in various locations for
medium size scales (520-620 km) and for areas located in the Western and Northwestern part of
the area of study. Relative importance of unique and common effects of IBR hypotheses was
constant between the three areas exhibiting maximum explained variance. The features \(T\) and \(Pr\)
uniquely contributed to more than 20\% of the total variance explained (20.77 to 32.65\% and 21.82
to 35.24%, respectively). The best contribution to the total variance explained was observed for the common effects of $E$ and $T$ (54.31 to 56.43%).

Discussion:

The dispersion of a species to environmental features is generally expected to be consistent across its distribution range. However, our ability to make inferences about the effect of environmental features may vary due to multiple evolutionary processes acting on genetic structure at different spatial and temporal scales. In this study, we explored potential barriers and corridors to dispersal and gene flow of a flying insect in a large area with dramatic landscape changes, which occurred at various time scales. Based on multi-scale and multi-site resampling of study areas, we found evidence for consistent effects of environmental features on gene flow at both local and large scales, but observed a heterogeneous distribution of these effects among locations, especially at the lowest spatial scales.

Effect of scale and location on inference

We observed a notable influence of scale on the detection of supported IBR hypotheses with Mantel tests for most environmental features tested ($E$, $T$ and $Pr$). Support was scarcely detected at the lowest spatial scale (220-400 km) and generally more often detected with increasing scale. Indeed, 190-400 km corresponded to the distances at which the variogram given in Figure 2 showed an initial plateau of genetic dissimilarity. This correspondence suggested that at this range of scales, dissimilarity between individuals was often not appropriate to show a significant effect of environmental features on gene flow. In contrast, the peak (for $T$) or inflection of curves (for $Pr$, $E$) of number of areas with supported IBR hypotheses observed at scales ranging from 400 to 600 km corresponded to the increase in dissimilarity in the variogram. Thus, scales above 400 km seemed more appropriate to gather a genetic structure in $M. \text{galloprovincialis}$ that was determined by the environmental features tested. Interestingly, we observed that the development of the variation of frequency of areas with support was specific to each environmental feature tested.
Similar results were observed for a large mammal (Zeller et al., 2014) and for insects (Rasic and Keyghobadi 2012) when multiple scales were considered.

The shape of the variogram showed a drop of dissimilarity of genotypes below scales of 190 km. This drop indicate a lower genetic differentiation between demes distant of up to about 200 km. Weak genetic differentiation at such scale was shown, based on estimates of population differentiation ($F_{st}$), for *M. alternatus* and *M. galloprovincialis* in lowland valleys (Kawai et al., 2006; Shoda-Kagaya 2007; Haran et al., 2015). Direct measures of the dispersal ability of *Monochamus* species show that adults may fly over distances ranging from 2 to 22 km in the field (Takasu et al., 2000; Linit & Akbulut 2003; Hernandez et al., 2011; Gallego et al., 2012; Mas et al., 2013; David et al., 2014). These flight performances are thought to cause intensive gene flow and generate the weak genetic structure observed in this study at small spatial scales. This weak genetic structure was sufficient to detect IBD in a large proportion of the areas at small scales (<220 km), but IBR was rarely supported at such scales. Our results illustrate a general problem of landscape genetic analysis performed on species with an important potential for dispersal. This is particularly true for flying species, which are naturally less affected by environmental features than non-flying species. For such species, the combination of intensive dispersal and gene flow and a limited number of environmental features affecting dispersal make inference difficult at small spatial scales (Dreier et al., 2014). Considering a continuous range of scales in analysis prevented us from basing our conclusions on a scale at which the effect of environmental features could not be detected. Our observations are consistent with the cases of large mammals for which multiple scales, including very large scales, have been used to deal with uncertainties regarding the scale of gene flow (Galpern et al., 2012; Zeller et al., 2014).

Based on resampling of areas of study across the Iberian Peninsula, we have shown the existence of a heterogeneous distribution of supported resistance models. Most variation in the distribution of support for IBR was observed at small and intermediate scales (220-600 km). Supported effects were mainly detected in the north-central part of the Iberian Peninsula. Conversely, effects were less supported in the rest of Iberian Peninsula (center, south and coasts). Two hypotheses may explain this spatial heterogeneity in the supported resistance models. A first hypothesis is that differences in variation of environmental features exist across resampled areas. An area exhibiting contrasting environmental features is known to affect dispersal more strongly.
and thus increase the chance of detecting their effect (Short Bull et al., 2011; Cushman et al., 2013). However, our results showed that at the smallest spatial scales, variation of environmental features in areas with supported IBR hypothesis was no higher than for non-supported areas, for most resistance models. This indicated that the distribution of variation of environmental features was not the main factor determining heterogeneity in support of resistance models. A second hypothesis is the existence of a conflicting signal due to the inclusion of two differentiated genetic clusters, probably of evolutionary history origin, in a study area (West and East Iberian clusters). In that case, “historical” genetic differentiation can unbalance the analysis by blurring the genetic structure occurring in response to landscape features, which is expected to be more recent and weaker. Such an effect probably explained the lack of support along the western Iberian coast. Indeed, the western Iberian cluster formed a narrow band, and areas of study almost systematically overlapped with the eastern cluster there. Within the eastern Iberian cluster, however, we observed a lack of detection of supported IBR hypotheses in areas that covered only one cluster (eastern Iberian coast). A large part of the heterogeneity was therefore not due to conflicting signal due to differentiated genetic clusters.

The above results highlighted that at scales between 220 and 600 km, *M. galloprovincialis* was structured according to environmental features in some areas but not in others, independent of artifacts or variations in heterogeneity of the environmental features. This observation is interesting, because one could expect a native species such as *M. galloprovincialis* to have a homogeneous dispersal in response to environmental features, at least within a genetic lineage. Determining the exact origin of such heterogeneity is challenging. It is suggested that this variation was a legacy of changes in the distribution of host trees in the Iberian Peninsula. The distribution and density of pine trees have been strongly affected by anthropogenic activities during the last centuries (Ruiz-Benito et al., 2012; Lopez-Merino et al., 2014), resulting in local extinction, as well as the connectivity and fragmentation of pine tree cover across time. For example, Abel-Schaad et al., (2014) showed that pine trees locally disappeared from the Central Iberian System during the middle ages. In contrast, these areas have been afforested at 80% with pines trees during 1940-1950. It is assumed that such recent modifications have dramatically affected the distribution and abundance of *M. galloprovincialis*, and that the time since these modifications occurred is too short to have affected the genetic structure of the beetle according to the environmental features tested (Epps & Keyghobadi 2015).
Strength of the effects of environmental features

We observed a decrease in the mean Mantel r with increasing scale. Such observation suggests that in areas exhibiting support for IBR hypotheses, correlation is stronger at small spatial scales than at larger scales. Such a situation is expected because larger areas in this study (600-1000 km) often harbored two distinct genetic clusters derived from evolutionary history, which could unbalance analyses. Conversely, small areas with support for IBR hypotheses showed the highest mean Mantel r values. This result suggests that areas with significant IBR hypotheses exhibit a “pure” effect with a less conflicting signal (i.e. differentiated genetic clusters). Therefore, our results show a tradeoff between the sampling of small areas where effects of environmental features are strong but scarcely detected and the sampling of large surfaces, where this effect is weaker but often detected.

Elevated areas and pine cover are barriers to dispersal for M. galloprovincialis

One of our hypotheses was that elevated areas constitute barriers to gene flow for M. galloprovincialis. The two resistance models (T and E) support this hypothesis (Fig. 3) and corroborate observations made for M. alternatus across the Ohu chain mountain in Japan (Shoda-Kagaya 2007) and on M. galloprovincialis across the Pyrenees (Haran et al., 2015). Several factors may explain this result. Temperature affects larval development and survival in M. galloprovincialis (Naves and Sousa 2009) and its ability to complete its development within one or two years (Tomminen 1993; Naves et al., 2007b; Koutroumpa et al., 2008). In addition, adult flying activity is affected by low daily temperatures (Hernández et al., 2011). Therefore, low temperatures likely constitute a factor that prevents migration across elevated areas by impeding or slowing species dispersal and development. In addition to this effect of temperature, topography may also explain the effect of elevation on dispersal. Indeed, Torrez-Vila et al., (2015) have shown that adults tend to fly down-hill using mark-release-recapture experiments. Therefore, it is possible that slopes represent a break in the dispersal of this species.

The effect of pine on dispersal was modeled according to two mutually exclusive hypotheses: high densities of pines represent barriers (Pr) or corridors (Pc) to dispersal, Our results
show that *M. galloprovincialis* is mainly structured according to the first hypothesis. The second hypothesis (*Pc*) was not supported in Commonality analysis and scarcely detected through the Mantel test. This weak signal is thought to correspond to type I errors that have been reported for Mantel tests (Guillot & Rousset 2013) and the quantitative approach used in this study allowed such false positive to be rejected. The prevalence of the *Pr* hypothesis show that *M. galloprovincialis* exhibits a limited dispersal when its resource is abundant. This species is known to develop on dead branches stemming from a self-pruning process encountered in pines (Mäkinen 1999). Dead branches represent a resource that is quite well distributed in space and time. Such abundance of resource is thought to cause limited dispersal in adults. The philopatric behavior of *M. galloprovincialis* in relation to the available resources is consistent with the observation of flight of this species in the field (Torres-Vila et al., 2015), or with the behavior of the pine processionary moth (*Thaumetopoea pityocampa*), another oligophagous pine–associated insect (Demolin 1969). Conversely, the *Pr* hypothesis suggests that low pine densities are not barriers to dispersal, This is in agreement with the suggestions of Torres-Vila et al., (2015) that the dispersal of *M. galloprovincialis* tends to be enhanced across open areas. In fact, the Iberian Peninsula contains several wide areas where pine tree forests are absent (center of Castilla y Leon for example), and our results suggested that such areas do not represent barriers to dispersal. Rossi et al., (2016) have shown that areas without pine forests still show a homogeneous distribution of scattered trees planted for ornamental use using observed and simulated data. We suggest that pine trees out of forests provide a scattered but homogeneously distributed resource that allows the dispersal of *M. galloprovincialis* across non-forested areas.

**Conclusions**

In this study, we highlighted that elevated areas and dense pine cover constitute barriers to the dispersal of *M. galloprovincialis*. We also showed that this species exhibit substantial gene flow at a scale of less than about 200 km. Along with the results related to the species model, our results exemplify the importance of simultaneously considering a continuous range of scales and multiple locations when exploring the effect of environmental features on dispersal in highly mobile species. Multiple scales allow the effect of environmental features at the appropriate extent
for each features tested to be inferred, while preventing analysis from being focused at an extent where intensive gene flow makes inference impossible due to the lack of genetic structure. In addition, resampling of the study area across multiple locations can help to identify variation in inference due to conflicting signals in genetic structure, and therefore allow for generalizing conclusions regarding the effects of environmental features on dispersal and gene flow. As a result, the combination of a resampled study area at multiple spatial scales across various locations in landscape genetics analysis provides a more general picture of the effects of environmental features on gene flow and has the power to reduce the versatility of results while limiting the sampling effort.

Acknowledgments

This work was supported by the European project REPHRAME KBBE.2010.1.4-09 (FP7 Project, Analysis of the potential of the pine wood nematode (*Bursaphelenchus xylophilus*) to spread, survive and cause pine wilt in European coniferous forests in support of EU plant health policy). The first author was funded by the French Ministry of Research and Education. Field work was supported by COST Action FP1002 (COST-STSM-FP1002-14177). We warmly thank Rolf Holderegger and Bertrand Gauffre for valuable comments on the early versions of this manuscript. We also thank Jérôme Rousselet and Christelle Robinet for interesting exchanges and help regarding the methodology.

Data Archiving

Data of are available in the supplementary data of this paper (Table S3: Sampling locations and microsatellite genotypes)

Tables

Table 1: Summary of environmental features tested in isolation by resistance (IBR) models.
Table 2: Commonality coefficients of both unique and common effects for the three sampling areas with the highest variance explained. Code pop: code of population of the center of 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 contribution of environmental features to the total variance explained.

Figure captions

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, size of pies refer to the size of demes). C: PCA of individuals on first and second axis.

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: Development of the number of areas with supported IBR hypotheses for Mantel tests (A) and of mean partial Mantel r (B) of areas with support of IBR hypotheses (p<0.05) with increasing scale (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 all environmental features tested (Euclidian distances, IBD; mean minimum temperatures, T; elevation, E; high pine densities as barriers, Pr; high pine densities as corridors, Pc). Grey maps refer to the distribution of environmental features associated with resistance models. Colored maps refer to interpolations of supported IBR hypotheses on the whole dataset (central column) and within the western Iberian cluster only (right column). From blue to red: low to high frequency of supported resistance models.

Figure 5: Development of spatial heterogeneity (mean standard deviation, SD) of environmental features in areas with supported and non-supported resistance hypotheses through Mantel test with increasing scale. Mean SD: mean standard deviation, T, E, Pr and Pc refer to IBR hypotheses tested, sign: significant, non-sign: non-significant.

References


**Appendices**

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

```r
# Simplified version of the script used in this study. Provide an overview of the general method employed.
#---------------------------------------------------------------
# create and plot background matrix with artificial barrier in middle
m <- matrix(1, nrow=10, ncol=10) ; m
m[,5] <- 4

library(raster)

r <- raster(m)
plot(r)

# create and plot transition matrix
library(gdistance)
t <- transition(r, transitionFunction=mean, 4, symm=TRUE, intervalBreaks=3)
plot(raster(t))

# create and plot sampling points and genetic data associated.
# (x coordinates, y coordinates, genetic data for 3 loci)
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, 1, 0, 2, 1, 0, 1, 0, 0, ncol=5)
xcoord<- matG2[, 1] ; ycoord <- matG2[, 2]
P<-cbind(xcoord,ycoord)
points(P)

# construction of moving windows (sampling areas)
library("ade4") ; library("vegan")

# Define the extent of sampling areas and the interval wanted
Min <- 0.7 # Minimum radius of areas wanted
Max <- 0.9 # Maximum radius of areas wanted
Step <- 0.1 # interval wanted

# Loops to test correlations in sampling area at multiple scales and locations
resultsfinal <- cbind(1,1,1,1)
colnames(resultsfinal) <- c("xcoord","ycoord", "Radius", "MantelR", "Pval")
for(Radius in seq(Min, Max, by = Step)){
```

results = NULL
for(i in 1:length(xcoord)){
    Xcircle <- (xcoord[i] + Radius*cos(seq(0,2*pi,length.out=100)))
    Ycircle <- (ycoord[i] + Radius*sin(seq(0,2*pi,length.out=100)))
    polygon(Xcircle, Ycircle)

    # extract individuals data in each sampling are constructed
    expr <- point.in.polygon(xcoord,ycoord,Xcircle,Ycircle)
    xcoord[expr==1]
    ycoord[expr==1]
    coordPoly <- cbind(xcoord[expr==1],ycoord[expr==1])

    # sort data and compute matrix of basic pairwise euclidian distances (not used further in this example)
    CoordOrder<- coordPoly[order(coordPoly[,1],decreasing=FALSE),]
    locOrder <- data.frame(CoordOrder)
    DisGeoEucl<-dist(locOrder, method = "euclidean", diag = TRUE, upper = TRUE)

    # compute corresponding matrix of genetic distances
    listcoord = (1:6)[expr==1]
    Genet = NULL ## fichier vide pour collage des données
    for(h in listcoord){
        tmp <- matG2[(matG2[, 1]==xcoord[h])and(matG2[, 2]== ycoord[h]), ]
        Genet = rbind(Genet,tmp)
    }
    GenetOrder<- Genet[order(Genet[,1],decreasing=FALSE),]
    GenetOrderSanscoord <- GenetOrder[-c(1,2)]
    MatdistGenet <- vegdist(GenetOrderSanscoord, method="bray", binary=FALSE, diag=FALSE, upper=TRUE, na.rm = TRUE)
    MatdistGenet <- as.dist(MatdistGenet)

    # Compute matrix landscape "resistance" distances based on raster
    spatiallocX <- locOrder[,1] ## extraction des colonnes pour repasser en spatial
    spatiallocY <- locOrder[,2]
    SpaLoc <- SpatialPoints(cbind(spatiallocX, spatiallocY))
    Resdis<- commuteDistance(t, SpaLoc)
    Resdis<-as.dist(Resdis, diag = TRUE, upper=TRUE)

    # simple mantels test between genetic and landscape "resistance" distances
    MantelpRes <- mantel.rtest(MatdistGenet, Resdis, nrepet = 99)
    results <- rbind (results, cbind (xcoord[i], ycoord[i],Radius, MantelpRes[2], MantelpRes[4]))
}
resultsfinal <- rbind(resultsfinal,results)

# display result file with for each individual: x and y coordinates, radius of sampling area, mantel output and associated p-value
Resultsfinal
Supplementary Material

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

Table S2: Details of primer sequence and genotyping.

Protocol

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

Table S3: Sampling locations and microsatellite genotypes.

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

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

Figure S3: Evolution of DeltaK among an increasing number of K (2-20).
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, size of pies refer to the size of demes). **C**: PCA of individuals on first and second axis.
Figure 2 (on next page)

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.
Development of the number of areas with supported IBR hypotheses for Mantel tests (A) and of mean partial Mantel r (B) of areas with support of IBR hypotheses (p<0.05) with increasing scale (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 (on next page)

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

Euclidian distances: IBD; mean minimum temperatures: $T$; elevation: $E$; high pine densities as barriers: $Pr$; high pine densities as corridors: $Pc$. Grey maps refer to the distribution of environmental features associated with resistance models. Colored maps refer to interpolations of supported IBR hypotheses on the whole dataset (central column) and within the western Iberian cluster only (right column). From blue to red: low to high frequency of supported resistance models.
**Figure 5** (on next page)

Development of spatial heterogeneity (mean standard deviation, SD) of environmental features in areas with supported and non-supported resistance hypotheses through Mantel test with increasing scale.

Mean SD: mean standard deviation, $T$, $E$, $Pr$ and $Pc$ refer to IBR hypotheses tested, sign: significant, non-sign: non-significant.
Table 1 (on next page)

Summary of environmental features tested in isolation by resistance (IBR) models.
<table>
<thead>
<tr>
<th>Environmental features</th>
<th>Code</th>
<th>Associated IBR hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elevation</td>
<td>$E$</td>
<td>High elevations = resistance to dispersal</td>
</tr>
<tr>
<td>Pine density</td>
<td>$P_c$</td>
<td>High pine density = corridors to dispersal</td>
</tr>
<tr>
<td></td>
<td>$P_r$</td>
<td>High pine density = resistance to dispersal</td>
</tr>
<tr>
<td>Mean min. temperatures</td>
<td>$T$</td>
<td>Low min. temperatures = resistance to dispersal</td>
</tr>
</tbody>
</table>
**Table 2** (on next page)

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

*Code pop*: code of population of the center of 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 contribution of environmental features to the total variance explained.
<table>
<thead>
<tr>
<th>IBR hypotheses</th>
<th>Coef.</th>
<th>% Total</th>
<th>Coef.</th>
<th>% Total</th>
<th>Coef.</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st-order</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E$</td>
<td>0,008</td>
<td>3,408</td>
<td>0,001</td>
<td>0,351</td>
<td>0,002</td>
<td>0,807</td>
</tr>
<tr>
<td>$T$</td>
<td>0,050</td>
<td>20,775</td>
<td>0,070</td>
<td>32,651</td>
<td>0,059</td>
<td>28,108</td>
</tr>
<tr>
<td>$Pc$</td>
<td>0,004</td>
<td>1,806</td>
<td>0,008</td>
<td>3,678</td>
<td>0,008</td>
<td>3,811</td>
</tr>
<tr>
<td>$Pr$</td>
<td>0,085</td>
<td>35,235</td>
<td>0,047</td>
<td>21,817</td>
<td>0,046</td>
<td>22,214</td>
</tr>
<tr>
<td>2nd-order</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E,T$</td>
<td>0,136</td>
<td>56,426</td>
<td>0,117</td>
<td>54,314</td>
<td>0,115</td>
<td>54,953</td>
</tr>
<tr>
<td>$E,Pc$</td>
<td>-0,003</td>
<td>-1,255</td>
<td>-0,001</td>
<td>-0,243</td>
<td>-0,001</td>
<td>-0,430</td>
</tr>
<tr>
<td>$T,Pc$</td>
<td>0,001</td>
<td>0,426</td>
<td>0,007</td>
<td>3,236</td>
<td>0,010</td>
<td>4,893</td>
</tr>
<tr>
<td>$E,Pr$</td>
<td>-0,002</td>
<td>-0,897</td>
<td>0,031</td>
<td>14,375</td>
<td>0,020</td>
<td>9,721</td>
</tr>
<tr>
<td>$T,Pr$</td>
<td>-0,013</td>
<td>-5,398</td>
<td>0,018</td>
<td>8,406</td>
<td>0,014</td>
<td>6,540</td>
</tr>
<tr>
<td>$Pc,Pr$</td>
<td>0,024</td>
<td>10,049</td>
<td>0,008</td>
<td>3,579</td>
<td>0,011</td>
<td>5,333</td>
</tr>
<tr>
<td>3rd-order</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E,T,Pc$</td>
<td>0,014</td>
<td>5,724</td>
<td>-0,005</td>
<td>-2,311</td>
<td>-0,001</td>
<td>-0,474</td>
</tr>
<tr>
<td>$E,T,Pr$</td>
<td>-0,023</td>
<td>-9,712</td>
<td>-0,069</td>
<td>-32,007</td>
<td>-0,047</td>
<td>-22,356</td>
</tr>
<tr>
<td>$E,Pc,Pr$</td>
<td>0,003</td>
<td>1,313</td>
<td>0,026</td>
<td>12,118</td>
<td>0,030</td>
<td>14,509</td>
</tr>
<tr>
<td>$T,Pc,Pr$</td>
<td>-0,009</td>
<td>-3,795</td>
<td>0,006</td>
<td>2,961</td>
<td>0,010</td>
<td>4,950</td>
</tr>
<tr>
<td>$E,T,Pc,Pr$</td>
<td>-0,034</td>
<td>-14,106</td>
<td>-0,049</td>
<td>-22,926</td>
<td>-0,068</td>
<td>-32,581</td>
</tr>
<tr>
<td>Sum</td>
<td>0,240</td>
<td>100</td>
<td>0,216</td>
<td>100</td>
<td>0,209</td>
<td>100</td>
</tr>
</tbody>
</table>