

Important Dispersal Capacity of One Main Species Vector of Bluetongue and Schmallenberg Viruses *Culicoides Obsoletus* (Diptera: Ceratopogonidae) Revealed by Landscape Genetic Analyses

Antoine Mignotte (✉ antoine.mignotte@wanadoo.fr)

ASTRE, Univ Montpellier, Cirad, INRAE, Montpellier, France. 2 Cirad, UMR ASTRE, F-34398 Montpellier, France.

Claire Garros

CIRAD BIOS: CIRAD Departement Systemes biologiques

Simon Dellicour

Université Libre de Bruxelles: Universite Libre de Bruxelles

Maude Jacquot

INRAE Centre Clermont-Auvergne-Rhône-Alpes: Institut National de Recherche pour l'Agriculture l'Alimentation et l'Environnement Centre Clermont-Auvergne-Rhone-Alpes

Marius Gilbert

Université Libre de Bruxelles: Universite Libre de Bruxelles

Laetitia Gardès

CIRAD BIOS: CIRAD Departement Systemes biologiques

Thomas Balenghien

CIRAD Biological Systems Department: CIRAD Departement Systemes biologiques

Maxime Duhayon

CIRAD BIOS: CIRAD Departement Systemes biologiques

Ignace Rakotoarivony

CIRAD BIOS: CIRAD Departement Systemes biologiques

Maia de Wavrechin

CIRAD BIOS: CIRAD Departement Systemes biologiques

Karine Huber

CIRAD BIOS: CIRAD Departement Systemes biologiques

Research

Keywords: *Culicoides obsoletus*, landscape genetics, microsatellites, dispersal, Palearctic region

Posted Date: September 16th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-75603/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on February 3rd, 2021. See the published version at <https://doi.org/10.1186/s13071-020-04522-3>.

Abstract

Background: In the last two decades, recurrent epizootics of bluetongue virus and Schmallenberg virus have been reported in the western Palearctic region. These viruses affect domestic cattle, sheep, goats and wild ruminants and are transmitted by native hematophagous midges of the genus *Culicoides* (Diptera: Ceratopogonidae). *Culicoides* dispersal is known to be stratified, *i.e.* due to a combination of dispersal processes occurring actively at short distances and passively or semi-actively at long distances, allowing individuals to jump hundreds of kilometers.

Methods: Here, we aim to identify the environmental factors that promote or limit gene flows of *Culicoides obsoletus*, an abundant and widespread vector species in Europe using an innovative framework integrating spatial, population genetics and statistical approaches. A total of 348 individuals were sampled in 46 sites in France and have been genotyped using 11 newly-designed microsatellite markers.

Results: We found a low genetic differentiation and population structure for *C. obsoletus* across the country. Using three complementary inter-individual genetic distances, we did not detect any significant isolation-by-distance, but a significant anisotropic isolation-by-distance on a North-South axis. We have employed a multiple regression approach on distance matrices to investigate the correlation between genetic and environmental distances. Among all the environmental factors that have been tested, only cattle density seems to have an impact on *C. obsoletus* gene flow.

Conclusions: Such a high dispersal capacity of *C. obsoletus* over land territories calls for a re-evaluation of the impact of *Culicoides* on virus dispersal and highlights the urgent need to better integrate molecular, spatial and statistical information to guide vector-borne diseases control.

Background

During a vector-borne disease outbreak, a precise understanding of the dispersal capacity of the vector species is a key aspect to implement appropriate control strategies in order to limit the spread of the disease. Dipteran dispersal is known to be impacted by the composition and configuration of the landscape. When the dispersal of an organism is favored by a landscape characteristic (“environmental factor”) it can be categorized as a conductance factor [1]. On the contrary, when the dispersal is limited by a landscape characteristic, it can be categorized as a resistance factor [1, 2]. However, some types of dispersal allow overcoming resistance factor, as it is the case for “stratified dispersion”. *Culicoides* dispersal is described as “stratified”, due to a combination of dispersal processes occurring actively at short distances, and passively or semi actively at long distances [3].

Northern Europe experienced very sudden and rapid outbreaks of bluetongue virus (BTV) in 2006–2008 and of Schmallenberg virus (SBV) in 2011–2012 [4]. Both viruses have spread very quickly and massively across the whole western Palearctic region transmitted by native *Culicoides* species. *Culicoides obsoletus* has been identified as the main vector species responsible for the transmission of BTV and SBV to wild

and domestic ruminants in western Palearctic region [5, 6, 7]. *Culicoides obsoletus* is a very widespread species able to breed in a wide range of habitats [8, 9]. BTV spread is facilitated by favorable conditions for midge activity and viral replication and the environmental and climatic drivers of BTV transmission in Europe have been suggested [10, 11, 12, 13, 14]. For example, high ambient temperature reduce incubation period of the virus [15] whereas high precipitation and wind speeds can reduced the *Culicoides* flight activity [16, 17, 18].

Previous studies of mark-release-recapture on *Culicoides* species showed that the post-release dispersal distance travelled, during two nights, ranges from 1 to 2.5 km, and is linked to the gradual search for hosts or oviposition sites [19]. The maximum recapture travelled distance recorded was 6 km for *Culicoides mohave* in a particular desert landscape [20]. Yet, the potentially high dispersal capacity of *Culicoides* and high mortality of adults when manipulating represents a practical limitation in the context of a mark-release-recapture procedure. Dispersal phenomena driven by wind currents on a much larger scale have also been detected during outbreaks related to *Culicoides*-borne diseases [3, 18, 21, 22, 23, 24, 25, 26]. Introduction of BTV serotypes by wind-borne infected *Culicoides* has been demonstrated from Northern Africa to Southern Europe [27], from Kenya to South-Western Indian Ocean Islands [28], from Sardinia to the Balearic Islands [25], from France to Corsica [29], in Ireland [30], or in UK [31]. Most of these studies are supported by modelling analyses in which dispersal trajectories have been evaluated by atmospheric dispersion models over water bodies [30, 32, 33, 34, 35, 36]. *Culicoides* dispersal over land territories has been less investigated [37] and little attempt has been made to link it to environmental factors. Given this high dispersal capacity and the stratified dispersal pattern of *Culicoides*, it is crucial to determine inland connectivity among populations and to identify the potential environmental factors that promote or limit gene flows between them.

Classic approaches to study the impact of landscape on gene flow generally use population-based sampling and are often based on a limited number of sampled populations [28, 38]. However, *Culicoides* such as other dipterans species, are not spatially structured into separate populations and must be considered as a continuum of individuals heterogeneously distributed across landscape. In order to identify the environmental factors having an impact on the dispersal of *Culicoides*, an individual approach is more relevant as it avoids inter-population genetic distance misinterpretations [39]. Indeed, the individual approach in landscape genetics aims at maximizing the number of sampling sites, and thus brings a much greater statistical power to detect spatial patterns of genetic differentiation and the environmental factors that cause them [40]. In addition to the analysis of inter-individual genetic distances, population genetic structure can also be investigated, *e.g.* using gold-standard Bayesian clustering methods. However, if isolation-by-distance occurs, these clustering analyses could overestimate the actual number of genetic clusters [41]. It is therefore important to also incorporate a visualization approach into the analytical workflow [42]. MAPI allow to visually compare genetic dissimilarity with some environmental factors but also to develop working hypotheses. On the other side, it is also necessary to use as a complement statistical analyses of genetic and environmental distances.

The present study here aims at determining population inland connectivity at large geographical scales of a main vector species and to identify the environmental factors that promote or limit gene flows between *Culicoides* populations in France. For this purpose, we characterized 13 microsatellite markers dedicated to *C. obsoletus*. We are proposing a complementary framework integrating multiple approaches which can be applied more generally to the study of gene flow and links with environmental factors.

Methods

Culicoides obsoletus sampling and morphological identification

The data set analysed in the present study is composed of 368 individual females biting midges collected in 46 sites in France in April 2011, using the national surveillance network for *Culicoides* populations or local complementary collections (Table S1). Collections were made overnight using Onderstepoort Veterinary Institute (OVI) light traps set up at farms near cattle or sheep. All insects were stored in 70% ethanol. Morphological identification to species level has been carried out in the same way as described in a previous study [43] under a binocular microscope using the available identification keys [44].

DNA extraction, amplification, sequencing and sequence analyses

A total of 368 individuals belonging to the *C. obsoletus* species complex were DNA-extracted, COI amplified and sequenced following the procedure detailed in Mignotte *et al.* (2020). To ensure that only *C. obsoletus* individuals were included in the final data set [45], sequences assignment was run according to Mignotte *et al.* (2020). The reference COI sequences used to make a specific assignment of all individuals are available in Table S2. All COI sequences were then aligned using the MUSCLE algorithm [46] implemented in the software package GENEIOUS v.6.0.5 (Biomatters, <http://www.geneious.com>). In order to estimate unbiased gene flows of *C. obsoletus*, only individuals identified as such were used for further analyses.

Development of microsatellite markers

Microsatellite markers were defined by next-generation sequencing following a similar protocol to the one described by Morillo *et al.* (2016). *Culicoides obsoletus* samples from France and United Kingdom (Table S3) were used as biological material for the DNA library. Among the 82 276 simple sequence repeats (SSRs) identified, primers were designed for 43 676 SSRs, of which 15 888 were dinucleotides, 23 656 were trinucleotides, and 4,132 contained complex SSR motifs. For the SSR screening, we selected 3 135 primer pairs flanking dinucleotide SSR motifs with a minimum of 12 repetitions and amplifying fragments between 150 and 300 base pairs (bp) in length and a half denaturation temperature (T_m) close to 55 °C. We screened a set of 30 primer pairs and optimized both polymerase chain reaction (PCR) and multiplex PCR conditions.

Microsatellite genotyping and loci filtering

Microsatellite makers were amplified by multiplex PCR with the Type-it-Microsatellites kit (Qiagen, Valencia, CA, USA) according to the protocol described in the manufacturer's manual and the annealing temperature given in Table S4. Standard conditions for PCR amplification included an initial denaturation step at 95 °C for 5 min, 35 cycles of denaturation for 30 s at 95 °C, annealing for 1 min at variable temperature (Table S4) and elongation for 1 min at 72 °C, followed by a final elongation of 5 min at 72 °C. Fragments were separated on an Applied Biosystems 3500xL Genetic Analyzer. Allelic size allocations for all individuals and microsatellite markers were performed using the program GeneMapper® version 5 (Life Technologies, Applied Biosystem) with double blind reading to limit the potential interpretation bias from the reader.

A first validation of the polymorphism of these 30 markers was carried out by amplifying and genotyping 48 individuals from various locations within the species geographic range. Then 13 loci showing good genetic profiles and clear allelic size variability were characterized as polymorphic markers and were selected for further analyses (Table S4).

The 368 *C. obsoletus* previously extracted were then amplified and genotyped at these 13 microsatellite loci. Genetic variability parameters, such as the number of alleles per locus, the allele size range, the observed and expected heterozygosity were estimated per locus, and per population over the entire dataset using the R packages *ggplot* [47], *poppr* [48] and *mmod* [49]. Short allele dominance could be a source of heterozygote deficiency in a microsatellites data set [50, 51]. In order to avoid such bias, Spearman's correlations rank between the size of the alleles and their frequencies were calculated in R [50]. Null allele frequencies were estimated using the R package *PopGenReport* [52], and a Bonferroni's correction was applied to all matched tests to take into account potential biases associated with multiple comparisons [53]. Correlation between F_{IS} and null allele was calculated in R, and locus with significant correlation were removed. The presence of an imbalance in genetic binding between each pair of loci has been tested by accurate Fisher tests performed with the R package *poppr* [48]. The same R package was used along with the *pegas* R package to identify a potential gap in the panmictic matching regime. For this purpose, we performed a chi-square tests to compare the observed heterozygosity deficits (H_o) to the expected heterozygosity (H_e) under the Hardy-Weinberg (HW) equilibrium with a significance threshold of 5%. We also calculated the fixation index (F_{IS}) [54].

Genetic clustering analyses

In order to determine the most likely number of genetic clusters (K) and the probability of assignment of each individual to these clusters, we used the Bayesian clustering methods implemented in the program STRUCTURE (version 2.3.4) [55] and GENELAND (version 4.5.0) [56]. The algorithm implemented in STRUCTURE infers genetic clusters that minimize deviations from Hardy-Weinberg equilibrium. In STRUCTURE, we specified the following settings: correlated allele frequency, an admixture model, and a "locprior" model. These options respectively allow to: assume allele frequencies are similar among

populations, estimate for each individual the belonging proportion to each K clusters, and use sampling locations as prior information to assist the clustering, which is particularly adapted when there is a weak genetic structuring [57]. We performed 10 independent runs for each value of K varying from 1 to 10. Each run consisted in a burn-in of 100,000 iterations in the Markov Monte Carlo chain (MCMC) and a stationary phase of 1,000,000 iterations in the MCMC. The most probable number of clusters was inferred with the ΔK method (Evanno et al. 2005), *via* the STRUCTURE Harvester web platform [58]. This ΔK value is a way to determine the inflection in the $[\ln P(D)]$ curve. The clustering resulting from the Bayesian inference was transposed into percentages of assignment of each individual to the K inferred clusters and plotted on a map.

GENELAND also uses a Bayesian algorithm to infer population genetic clusters while taking into account the spatial position of individuals, making it a spatially-explicit clustering method. The most probable number of clusters was also determined by running the algorithm with K ranging from 1 to 10. The analysis was based on 1,000,000 MCMC iterations with a thinning of 1000, maximum rate of the Poisson process fixed to 100, maximum number of nuclei in the Poisson-Voronoi tessellation fixed to 300 and a burn-in of 100. We used the R package *graphics* to produce a distribution map of genetic structure resulting of STRUCTURE and GENELAND analyses.

Computation of genetic distances and mapping genetic dissimilarity

We computed three complementary inter-individual genetic distances. The first inter-individual genetic distance is based on a Factorial Correspondence Analysis (FCA) performed with the program GENETIX v. 4.05.2 [59]. The first ten FCA axes were used to calculate an Euclidean distance between all individuals using the R package *Ecodist* [60], hereafter called the “FCA” distance [61]. The second inter-individual genetic distance is the Rousset’s distance (a_R), [62], corresponding to the $F_{ST}/(1 - F_{ST})$ but for pairs of individuals. We used the SPAGeDi 1.5 program [63] to compute a_R genetic distance. The third inter-individual genetic metric is the Loiselle’s kinship coefficient (“LKC”, [64]), which, unlike FCA and a_R that are measures of genetic dissimilarity, is an index of similarity [65]. We used the R package *fields* [66] to investigate isolation-by-distance by performing linear regressions and Mantel tests between inter-individual genetic distances and log-transformed geographic distances with R package *fields* [66]. Mantel tests [67] were performed with the R package *vegan* [68]. Inter-individual genetic distances were mapped using the program MAPI [42]. The method implemented in MAPI allows to visualize spatial variation in genetic dissimilarity. It can also define areas where genetic dissimilarity is significantly lower or higher than expected by chance [42]. MAPI analyses were based on 1,000 permutations and an α value of 0.05.

Impact of environmental factors on genetic differentiation

We investigated the potential impact of several environmental factors on the genetic differentiation of *C. obsoletus*. In practice, we tested the association between matrices of pairwise genetic distances (see above) and matrices of environmental distances. These environmental distances were computed using the program CIRCUITSCAPE 4.0.5 [69, 70] and were based on distinct environmental rasters (*i.e.* geo-

referenced grids of environmental values) either treated as potential resistance (R) or conductance (C) factors:

- i. host densities: European index of distribution of roe deer (C) [71], European index of distribution of red deer (C) [71], global cattle, sheep and goat distributions in 2010 (C) [72].
- ii. terrestrial habitat: mean global elevation in 2010 (R) [73] and type of landscape cover (urban area, grassland, forest areas and croplands) (R). The four distinct land cover rasters (resolution: ~1000 m) were generated from the Corine Land Cover 2012 raster (CLC; www.eea.europa.eu; resolution: ~100 m) (Fig. S1).
- iii. meteorological and climate data: mean surface temperature (R), mean precipitation (R), and mean wind speed and direction (C). We used two alternative rasters for each variable: April 2011 mean and 2000–2010 mean. Surface temperatures were obtained from monthly day and night land surface temperature mean [74]. Precipitation and wind rasters were obtained from monthly means of precipitation, wind speed and the U and V components of wind (components of the horizontal wind towards the east and North) from the European Centre for Medium-range Weather Forecasts (Era5 Interim, <http://www.ecmwf.int>) over 10 years (2001–2010). Wind U and V components were used to compute prevailing meteorological wind direction (in degrees). Map and arrowheads indicating the mean wind direction over the 10 years at each raster pixel were generated with the R package *Rwind* (Fig. S2) [75].

A so-called “null raster” was also created and used as a negative control, *i.e.* a raster on which environmental distances computations correspond to a proxy of geographical distances. The null raster is a uniform raster with all cell values equal to 1. As in Dellicour *et al.* (2019), three values of k (10, 100 and 1000) were used to modify the potential impact of the environment on the resistance/conductance value. We used multiple regressions on distance matrices (MRDM) coupled with commonality analyses (CA; [76]) to identify unique and common contributions of predictors to the variance in the environmental distance (response variable). All the analyses were carried out with the R packages *ecodist* [60] and *yhat* [77]. In addition, we also performed univariate analyses by comparing (i) the determination coefficient R^2 obtained from the linear regression of genetic distances on distances computed on the environmental raster and (ii) the determination coefficient R^2 obtained from the linear regression of genetic distances on distances computed on the null raster. Only the environmental factors associated with a R^2 higher than that the one obtained from the linear regression based on environmental distances computed on the null raster were selected for the multivariate analyses (MRDM-CA) [78]. The same criteria to identify a suppressors were used for the multivariate analysis as in the commonality analyses descriptions [79, 80]. These suppressors allow to select environmental variables that are more explanatory than the geographical distance alone. An environmental variable is considered to be a suppressor if regression coefficient and correlation, or when the unique contribution and the common contribution are of opposite sign [79].

Anisotropic isolation-by-distance

In order to identify directional gene flows, we must calculate the projected distance between two sampling sites. The angle between sites was calculated using the R package *geosphere* [80] and transformed from degrees to radians. The set of angles between 0 and 360 were then tested as angles that maximize isolation-by-distance. The great-circle geographic distance between sampling sites was computed with the R package *fields* and then log-transformed [66]. The projected distance matrix was calculated for each angle between all sampling sites, using the formula below. In calculation of the projected geographical distance as a function of the angle between populations, d_{AB} is the geographical distance between population A and B and a_{AB} is the angle between populations A and B.

$$\log (| d_{AB} * \cos(a_{AB}) |)$$

A linear regression of the genetic distances on the projected geographical distances obtained for each angle was then performed. The angle that will maximize the R^2 of this regression with a positive regression coefficient was considered as the angle maximizing the isolation-by-distance signal [81]. In order to confirm this analysis with a similar but complementary approach, we evaluated the direction that maximizes the correlation between geographic and genetic distance using the PASSAGE program and a "Bearing" analysis. The workflow summarizing the genetic analyses carried-out is described in Figure S3.

Results

Characterization of microsatellite markers and genetic diversity

A total of 368 COI sequences of *C. obsoletus* were obtained after the DNA barcoding step (Table S3) and 368 *C. obsoletus* from 46 populations were genotyped with 13 microsatellite loci (Table S4). A filter has also been applied to exclude missing data. Individuals with more than 10% missing data were excluded from the analysis, *i.e.* 20 individuals. Loci OBSms25 and OBSms26 had the highest proportion of null alleles with 56% and 22% respectively, but also the highest F_{IS} values of 0.691 and 0.357 respectively (Table S5). The expected heterozygosity, which varies from 0.918 (OBSms5) to 0.404 (OBSms29) is always higher than that observed, varying from 0.872 (OBSms13) to 0.197 (OBSms25), reflecting a heterozygous deficit (Table S5). To ensure that all markers provide independent information, linkage disequilibrium between each pair of loci was tested. After Bonferroni correction, no significant linkage disequilibrium was observed between all loci. The OBSms25 locus was also the only locus with more than 5% missing data. Thus, only markers with a null percentage inferior to 10% were kept for the final analysis. In view of the high proportion of null alleles and the strong correlation between F_{IS} and the rate of null alleles ($r = 0.947$, $p\text{-value} > 0.001$), the OBSms25 and OBSms26 loci were excluded from the analysis. No significant correlation between allele size and allele frequency was found for any of the 11 remaining loci (Table S5). The genetic resolution of microsatellite markers depends on the number of markers and their polymorphism. All selected 11 loci were therefore polymorphic (Table S5) and the number of alleles present at a given locus varied from 40 for OBSms5 to 9 for OBSms29. No significant

isolation-by-distance was detected with the different inter-individual genetic distances, a_R ($R^2 < 0.001$, $p = 0.439$, Fig. S4), LKC ($R^2 < 0.001$, $p = 0.992$, Fig. S4), and FCA ($R^2 < 0.001$, $p = 0.361$, Fig. S4).

Clustering analyses and **mapping of genetic distances**

The Bayesian analysis performed with STRUCTURE revealed a low genetic structuring of *C. obsoletus*. Meanwhile, the exact number of genetic clusters was not clearly defined, the optimal value of K was 2 (Fig. S5); *i.e.* the co-existence of two genetic clusters was the most likely statistically (Fig. 1). The assignment of individuals to the different clusters revealed a weak genetic structuring of the populations. Although the assignment of a total of 284 out of the 348 *Culicoides* individuals to a genetic cluster is well supported ($Q > 0.90$), no clear association between spatial clustering and distribution can be observed, *i.e.* there is no association between *Culicoides* sampling location and the genetic clusters distributions. However, a West-East transect seems to emerge as samples collected on this transect were more frequently assigned to cluster 1, contrary to the North-South transect more represented by cluster 2. Genetic clusters inferred by GENELAND and STRUCTURE ($K = 2$) are consistent with these results, both clusters being found at all the sampling sites (Fig. 1 ab). Results for the second optimal K value ($K = 4$) of the STRUCTURE program gave a similar view, reinforcing the conclusion that populations are genetically weakly structured (Fig. S6). Additional STRUCTURE analysis has been carried out with only loci OBSms 4, 11 and 28 which are at Hardy Weinberg's equilibrium and no additional structure is detected using only this loci (data not shown). Genetic dissimilarity result from MAPI analyses with the a_R genetic distance showed a non-homogeneous pattern of dissimilarity. The maximum inter-individual genetic dissimilarity was of 0.182 and the lowest was of 0.002. The South of France was the most genetically dissimilar region. The center of France was also a zone of strong inter-individual dissimilarity, while eastern France showed low genetic dissimilarity. Another area spanning from the Paris region to Alsace (North east of France) showed low genetic dissimilarity (Fig. 1 cd). MAPI analyses with FCA inter-individual genetic distance, showed a homogeneous genetic dissimilarity. However, almost none of the areas was significantly genetically dissimilar in MAPI analysis with the three genetic distances used.

Investigating the impact of environmental factors

Univariate regressions revealed a significant positive association of FCA genetic distances with cattle distributions for $k = 100$ ($Q = 0.009$, $p\text{-value} = 0.001$) and $k = 1000$ ($Q = 0.029$, $p\text{-value} = 0.001$), with sheep distributions for $k = 100$ ($Q = 0.002$, $p\text{-value} = 0.002$) and for $k = 1000$ ($Q = 0.006$, $p\text{-value} = 0.001$), with elevation for $k = 100$ ($Q = 0.001$, $p\text{-value} = 0.005$) and for $k = 1000$ ($Q = 0.001$, $p\text{-value} = 0.006$), and with grassland for $k = 100$ ($Q = 0.009$, $p\text{-value} = 0.001$) and for $k = 1000$ ($Q = 0.011$, $p\text{-value} = 0.001$) distances (Table S6). Similarly, a_R inter-individual genetic distances were significantly associated with distances computed from elevation for $k = 100$ ($Q = 0.00018$, $p\text{-value} = 0.006$) and for $k = 1000$ ($Q = 0.00038$, $p\text{-value} = 0.007$), with grassland for $k = 100$ ($Q = 0.001$, $p\text{-value} = 0.001$) and for $k = 1000$ ($Q = 0.002$, $p\text{-value} = 0.001$) and with urban areas for $k = 100$ ($Q = 0.004$, $p\text{-value} = 0.001$) and for $k = 1000$ ($Q = 0.007$, $p\text{-value} = 0.001$) rasters (Table S6). However, no significant association was found between "LKC" genetic distance and the environmental distances (Q values always < 0).

Multivariate regressions revealed a significant positive association of FCA inter-individual genetic distance with cattle distributions for $k=1000$ ($r=0.1733$, $p\text{-value} < 0.001$) and with grassland for $k=1000$ ($r=0.1080$, $p\text{-value} < 0.001$) (Table S6). Also, multivariate regressions did not reveal any significant positive association of a_R inter-individual genetic distances with environmental distances. All environmental factors were considered as suppressors when using a_R to measure inter-individual genetic distances. Cattle distribution ($k=1000$) with FCA inter-individual genetic distance was the only environmental factor that showed a unique contribution to the variance in the dependent variable higher than 1% ($U=0.02$) (Table S6).

Anisotropic isolation-by-distance

Globally, correlation between the genetic distance and the distance along the bearing θ , d_θ , changed as a function of bearing θ . Anisotropic isolation-by-distance analyses revealed a significant positive correlation between a_R and FCA inter-individual genetic distances and orientational distances computed along North–South orientation and a significant negative correlation between genetic distances and orientational distances computed along a West–East orientation (Fig. 2a,b). For the analysis with LKC, which is a variable of genetic similarity, the North–South was the orientation that maximized the isolation-by-distance (Fig. 2c). Bearing analyses with PASSAGE software revealed a significant distance isolation on the North/South axis with the genetic distances a_R and FCA (Fig. 3a,b). A negative correlation between LKC genetic similarity and geographical distance was observed on the North/South axis (Fig. 3c).

Discussion

This study presented for the first time an extensive landscape genetic study carried out on a main vector species of interest for livestock diseases, *C. obsoletus*. Bayesian clustering analysis revealed a very weak genetic structure of *C. obsoletus* in France and a low level of inter-individual genetic differentiation was observed. We assessed the impact of a wide range of environmental factors on this pattern. Univariate and multivariate analyses highlighted an absence, or a weak impact of most of the tested environmental factors on inter-individual measures of pairwise genetic differentiation. No isolation-by-distance pattern was detected at the French scale with the set of inter-individual genetic distances used. However, an anisotropic isolation-by-distance analysis revealed a significant distance isolation on a North–South axis.

Important gene flows between C. obsoletus populations

The low overall inter-individual genetic dissimilarity highlighted by our study reflected the important level of gene flow for *C. obsoletus* and comes to reinforce what has been described previously for other *Culicoides* species. In particular, important level of gene flows have already been observed in France for *C. imicola* [29] and in Australia for *C. brevitarsis* [82]. Genetic studies on *C. imicola* in Europe revealed an important gene flow as reflected by the inference of two large genetic clusters: a “Central Mediterranean cluster” including Algeria, Sardinia, Corsica, Pyrénées-Orientales and Var French departments of France,

and a “Western Mediterranean cluster” including Morocco, Spain, Portugal and Majorca [26]. In North America, similar results were also found in *C. stellifer*. No barriers to gene flow could be identified in this species in the southeast United States [83]. In addition, no isolation-by-distance has been observed at the French scale. Identifying a genetic structure pattern is a real challenge when inter-individual genetic dissimilarity is low. Although STRUCTURE performs well at low levels of population differentiation ($0.02 < F_{ST} < 0.10$) by using prior population and correlated allele frequency models [84], when the differentiation is weaker, such as in the case of highly dispersed organisms, the latter may encounter difficulties [85]. This underlines the importance of using complementary tools of visualization approaches like MAPI. In addition, the use of an individual approach avoids to artificially consider individuals grouped into populations solely due to the effect of the trapping sites [2, 86]. Although our method of analysis seems complementary and coherent to detect a genetic structure, *C. obsoletus* is not genetically or geographically structured at the French scale.

Livestock densities as potential drivers of C. obsoletus gene flow

It is undeniable that large gene flows homogenize the genetic diversity of *Culicoides*. This can be explained by active dispersal and host-seeking movements, as suggested by our results. While the unique contribution of all environmental factors tested was very small, the one of cattle density is the strongest we have detected. Host density as a conductance factor for *Culicoides* is obviously a factor consistent with the biology of the species. These results are in line with those obtained in landscape genetics studies of the BTV, which had identified distributions of cattle and sheep as a key factors in BTV dispersal [13]. In addition, previous studies showed that dairy cattle density is negatively correlated with BTV spread. Although paradoxical with the previous conclusions, this could be explained by the fact that dairy cattle are clustered around the milking parlour and move little, forming a fixed feed source, limiting the spread of *Culicoides*. On the contrary, beef cattle disperse much more in the pastures and by this phenomenon could encourage the diffusion of *Culicoides* in search of a source of blood [87]. In view of these converging results, of the marked preference of certain species of *Culicoides* vector for cattle [88], and of BTV emergence or reemergence events in cattle in the Netherlands in 2006 and in France in 2015 [89, 90], it seems crucial to closely monitor the surveillance of *Culicoides* in the vicinity of beef cattle farms [88].

In addition, it has been shown that BTV spread is facilitated at low elevation, up to 300 m [38, 87]. Altitude, an environmental factor tested in this study, failed to explain the inter-individual genetic differentiation of *C. obsoletus*. The altitude therefore does not act as a barrier between sampling sites located at relatively low altitude. It would thus be interesting to integrate in future studies sampling sites with a high altitude above 1,000 m, which is possible for *C. obsoletus* highlighting a large altitudinal range [91]. However, it is possible that the altitude does not directly impact the dispersal activity of *Culicoides*, but only the replication or viral infection of BTV due to low temperature. *Culicoides obsoletus* is extremely generalist and can take a blood meal on a wide range of hosts [92]. The study of phylogenetically close but ecologically very different species of *Culicoides*, such as *C. chiopterus* [93],

might then allow the identification of very different dispersal patterns and an importance of bovine density, the latter being exclusively dependent on it for egg-laying [94].

Long-dispersal of C. obsoletus

The most genetically dissimilar individuals are mainly located in the southernmost populations of the sampling area. Multiple non-exclusive lines of argument might explain the significant anisotropic isolation-by-distance observed on the North/South axis in France.

First, the significant anisotropic isolation-by-distance could be explained by the wind dispersal. Dispersion phenomena caused by wind currents are already established for *Culicoides*, mainly over the seas [3, 18, 21, 22, 23, 24, 25, 29]. The reflecting dispersal events can be described as passive and active because recapture beam has been carried out downwind and upwind of the prevailing wind direction [19]. The map of the average wind direction in France over the last 10 years shows potential answers. It can be seen that the southernmost sampling sites (with the most dissimilar individuals) are in an area where the wind direction is different from the rest of France. It should be noted that the diffusion of the BTV, and thus the dispersion of the *Culicoides*, has already been associated with the wind direction. For example, 2% of BTV infections occurred at distances greater than 31 km [95, 96] during the 2006 epizootic. The South of France was initially sampled, however a poor conservation of *Culicoides* DNA did not allow us to perform barcoding and microsatellite genotyping of these populations. The study of this geographical area could thus complete this study and potentially identify a stronger genetic structure. This will make it possible to decide whether this premise of differentiation observed in this study is due to an older phylogeographic structure, a different wind dispersion in this geographical area, or a random pattern. It is also questionable whether the very low indications of genetic structure could be due to differences in genetic diversity between the south and the north. A lower genetic diversity in the north could then be a sign of an expansion process that would contrast with the interpretation of a large-scale dispersion [97, 98]. Indeed the spatial expansion of populations is generally accompanied by a gradients of reduction in the genetic diversity of the population during the expansion process, caused by serial founder events, creating a genetic bottleneck through a founding effect [99, 100]. Population sampling at the same scale could allow comparison and estimation of genetic diversity.

Second, anisotropic isolation-by-distance may be due to an artifact due to sampling methods. Indeed, if the extent of the sampling varies depending on the directions, the distances projected from the angles may represent different distance distributions and lead to the over-representation of strong genetic metrics value in the direction of the scatterplot and the hit of positive correlation signals. Moreover, the absence of correlation does not necessarily mean more gene flow but an absence of isolation-by-distance which can also result from a strong drift and thus less gene flow, a drift which depends on both dispersion and population sizes. It is therefore essential to use as a complement, as we have done here and come up with a similar result, an approach that weights the geographical distances between populations according to their orientation with respect to a given angle axis passing through the barycenter, as implemented in PASSAGE.

Consideration for future work

Our results underline the importance of methodological development of wind dispersal models of *Culicoides* on land and not only over water. However, studies of landscape genetics remain indispensable and complementary in order to improve the accuracy of predictive models for *Culicoides* dispersal over land through integration of meteorological, landscape and activity-based parameters previously tested and validated in the landscape genetics work flow. In addition, population sampling of the same species at European level, *i.e.* a very large proportion of the known range of the species, would be necessary to observe a more marked structuring at European level and to estimate more precisely the gene flow. For example, *C. imicola* in Europe is structured in two large genetic clusters, “Europe central cluster” and “Europe western cluster” [26]. Moreover, the question of the phylogeographic history of *C. obsoletus* is still very little explored in view of its geographical distribution throughout Europe and North America. On the contrary, the study of the phylogeography of *C. imicola*, the main Afrotropical vector, has shown its distribution range out from the northern part of sub-Saharan Africa to the Mediterranean basin [38]. This type of study would make it possible to estimate the effective size of the populations, a key factor in the dispersal of *C. obsoletus*. In future study, the use of High-throughput sequencing approaches using markers such as ddRadseq (Double-digest restriction-site associated DNA sequencing) can provide greater resolution in view of the large number of SNPs revealed (single nucleotide polymorphism) at a local scale and improve our understanding of the active and passive dispersal of *Culicoides*. It could also be relevant to look at on scale with finer genetic information. This could include more microsatellite markers or SNP to improve genetic resolution and observe the matching and assignment of each individual.

Conclusions

For the first time, this study provided the first complete landscape genetic analysis of *C. obsoletus*, a major vector species of animal viruses in Europe. This study showed that the large scale dispersion is able to homogenize the genetic structure of populations at the scale of a country. Our results demonstrate, for *C. obsoletus*, a very important inland dispersal and vectorization capacity, which has to be considered for further works on vector competence and epidemiological modeling of disease transmission. Wind direction could be a key factor in the dispersal of many vector insect species. Futures studies should increase their geographical extent to cover the entire species distribution area and better apprehend the limits of *Culicoides* gene flow. In addition to this biological knowledge, the study highlights several important areas for methodological improvement that may currently limit the inclusion of wind direction in landscape genetic analyses.

Abbreviations

BTV
bluetongue virus; SBV:Schmallenberg virus; *cox1*:cytochrome *c* oxidase subunit 1; N:number of individuals. SSR:simple sequence repeat. SNP:single nucleotide polymorphisms.

Declarations

Acknowledgement

SD and MG are supported by the *Fonds National de la Recherche Scientifique* (FNRS, Belgium). AM and MJ have received a fellowships from the H2020-727393 PALE-Blu project funded the EU. We are most grateful to H  l  ne Vignes, Ronan Rivallan and Xavier Argout at the GPTR (Grand plateau technique regional de g  notypage) platform in Montpellier, France for their help in development of microsatellite markers and Aurore Manez for her help in genotyping. We would like to thank Sylvain Piry, Karine Berthier and Marie-Pierre Chapuis, for their constructive exchanges which helped to improve the relevance of our conclusions. The authors are grateful to all partners who assisted in *Culicoides* sampling and shipping. We thank William Wint (University of Oxford, PALE-Blu partner) for his continued support in extracting and preparing meteorological data.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its additional files. The newly generated sequences were submitted in the GenBank database under the accession numbers MT828832-MT828844. All *cox1* sequences are available on request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was partially funded by the EU grant H2020-727393 PALE-Blu and by the VectorNet project (OC/EFSA/AHAW/2013/02-FWC1) funded by the European Centre for Disease Prevention and Control (ECDC) and the European Food Safety Authority (EFSA). The contents of this publication are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission, of the ECDC or of the EFSA.

Author contributions

AM, KH and CG designed the study. TB contributed to gather *Culicoides* samples and IR, MD contributed the identification of *Culicoides*. AM, LG, and M de W have performed molecular biology manipulations.

AM, SD and MJ, developed techniques and workflow for landscape genetics analysis. AM analyzed the data. KH and CG contributed to the manuscript firstly written by AM. All authors read and commented the final manuscript version.

References

1. Manel S, Holderegger R. Ten years of landscape genetics. *Trends Ecol Evol.* 2013;28:614–21.
2. Manel S, Berthoud F, Bellemain E, Gaudeul M, Luikart G, Swenson JE, et al. A new individual-based spatial approach for identifying genetic discontinuities in natural populations. *Mol Ecol.* 2007;16:2031–43.
3. Murray MD, Kirkland PD. Bluetongue and Douglas virus activity in New South Wales in 1989: further evidence for long-distance dispersal of the biting midge *Culicoides brevitarsis*. *Australian Veterinary Journal.* 1995;Vol. 72.
4. Carpenter S, Groschup MH, Garros C, Felipe-Bauer ML, Purse BV. *Culicoides* biting midges, arboviruses and public health in Europe. *Antiviral Res.* 2013;100:102–13.
5. Carpenter S, Wilson A, Mellor PS. *Culicoides* and the emergence of bluetongue virus in northern Europe. *Trends Microbiol.* 2009;17:172–8.
6. Hoffmann B, Eschbaumer M, Beer M. Real-time quantitative reverse transcription-PCR assays specifically detecting bluetongue virus serotypes 1, 6, and 8. *J Clin Microbiol.* 2009;47:2992–4.
7. Savini G, MacLachlan NJ, Sanchez-Vizcaino JM, Zientara S. Vaccines against bluetongue in Europe. *Comp Immunol Microbiol Infect Dis.* 2008;31:101–20.
8. Mehlhorn H, Walldorf V, Klimpel S, Jahn B, Jaeger F, Eschweiler J, et al. First occurrence of *Culicoides obsoletus*-transmitted Bluetongue virus epidemic in Central Europe. *Parasitol Res.* 2007;101:219–28.
9. Purse BV, Carpenter S, Venter GJ, Bellis G, Mullens BA. Bionomics of temperate and tropical *Culicoides* midges: knowledge gaps and consequences for transmission of *Culicoides*-borne viruses. *Annu Rev Entomol.* 2015;60:373–92.
10. Guis H, Caminade C, Calvete C, Morse AP, Tran A, Baylis M. Modelling the effects of past and future climate on the risk of bluetongue emergence in Europe. *J R Soc Interface.* 2012;9:339–50.
11. MacLachlan NJ, Guthrie AJ. Re-emergence of bluetongue, African horse sickness, and other orbivirus diseases. *Vet Res.* 2010;41:35.
12. Purse BV, Mellor PS, Rogers DJ, Samuel AR, Mertens PP, Baylis M. Climate change and the recent emergence of bluetongue in Europe. *Nat Rev Microbiol.* 2005;3:171–81.
13. Jacquot M, Nomikou K, Palmarini M, Mertens P, Biek R. Bluetongue virus spread in Europe is a consequence of climatic, landscape and vertebrate host factors as revealed by phylogeographic inference. *Proc. Biol. Sci.* 2017;284.
14. Cuellar AC, Kjaer LJ, Baum A, Stockmarr A, Skovgard H, Nielsen SA, et al. Modelling the monthly abundance of *Culicoides* biting midges in nine European countries using Random Forests machine learning. *Parasit Vectors.* 2020;13:194.

15. Wittmann EJ, Mello PS, Baylis M. Effect of temperature on the transmission of orbiviruses by the biting midge, *Culicoides sonorensis*. *Med Vet Entomol.* 2002;16:147–56.
16. Carpenter S, Szmargad C, Barber J, Labuschagne K, Gubbins S, Mellor P. An assessment of *Culicoides* surveillance techniques in northern Europe: have we underestimated a potential bluetongue virus vector? *J Appl Ecol.* 2008;45:1237–45.
17. Baylis M, Parkin H, Kreppel K, Carpenter S, Mellor PS, McIntyre KM. Evaluation of housing as a means to protect cattle from *Culicoides* biting midges, the vectors of bluetongue virus. *Med Vet Entomol.* 2010;24:38–45.
18. Gloster J, Burgin L, Witham C, Athanassiadou M, Mellor PS. Bluetongue in the United Kingdom and northern Europe in 2007 and key issues for 2008. *Vet Rec.* 2008;162:298–302.
19. Kluiters G, Swales H, Baylis M. Local dispersal of palaeartic *Culicoides* biting midges estimated by mark-release-recapture. *Parasit Vectors.* 2015;8:86.
20. Brenner RJ, Wargo MJ, Stains GS, Mulla MS. The Dispersal of *Culicoides mohave* (Diptera, Ceratopogonidae) in the Desert of Southern-California. *Mosquito News.* 1984;44:343–50.
21. Murray MD. Local Dispersal of the Biting-Midge *Culicoides-Brevitarsis* Kieffer (Diptera, Ceratopogonidae) in Southeastern Australia. *Aust J Zool.* 1987;35:559–73.
22. Eagles D, Deveson T, Walker PJ, Zalucki MP, Durr P. Evaluation of long-distance dispersal of *Culicoides* midges into northern Australia using a migration model. *Med Vet Entomol.* 2012;26:334–40.
23. Hendrickx G, Gilbert M, Staubach C, Elbers A, Mintiens K, Gerbier G, et al. A wind density model to quantify the airborne spread of *Culicoides* species during north-western Europe bluetongue epidemic, 2006. *Prev Vet Med.* 2008;87:162–81.
24. Ducheyne E, De Deken R, Becu S, Codina B, Nomikou K, Mangana-Vougiaki O, et al. Quantifying the wind dispersal of *Culicoides* species in Greece and Bulgaria. *Geospat Health.* 2007;1:177–89.
25. Alba A, Casal J, Domingo M. Possible introduction of bluetongue into the Balearic Islands, Spain, in 2000, via air streams. *Vet Rec.* 2004;155:460–1.
26. Jacquet S, Huber K, Pages N, Talavera S, Burgin LE, Carpenter S, et al. Range expansion of the Bluetongue vector, *Culicoides imicola*, in continental France likely due to rare wind-transport events. *Sci Rep.* 2016;6:27247.
27. Gubbins S, Carpenter S, Baylis M, Wood JL, Mellor PS. Assessing the risk of bluetongue to UK livestock: uncertainty and sensitivity analyses of a temperature-dependent model for the basic reproduction number. *J R Soc Interface.* 2008;5:363–71.
28. Onyango MG, Michuki GN, Ogugo M, Venter GJ, Miranda MA, Elissa N, et al. Delineation of the population genetic structure of *Culicoides imicola* in East and South Africa. *Parasit Vectors.* 2015;8:660.
29. Jacquet S, Huber K, Guis H, Setier-Rio ML, Goffredo M, Allene X, et al. Spatio-temporal genetic variation of the biting midge vector species *Culicoides imicola* (Ceratopogonidae) Kieffer in France. *Parasit Vectors.* 2016;9:141.

30. McGrath G, More SJ, O'Neill R. Hypothetical route of the introduction of Schmallenberg virus into Ireland using two complementary analyses. *Vet Rec.* 2018;182:226.
31. Burgin L, Ekstrom M, Dessai S. Combining dispersion modelling with synoptic patterns to understand the wind-borne transport into the UK of the bluetongue disease vector. *Int J Biometeorol.* 2017;61:1233–45.
32. The U.K. Met Office's Next-Generation Atmospheric Dispersion Model
NAME III
Andrew Jones DT, Hort M, Devenish B. The U.K. Met Office's Next-Generation Atmospheric Dispersion Model, NAME III. 2007.
33. Eagles D, Melville L, Weir R, Davis S, Bellis G, Zalucki MP, et al. Long-distance aerial dispersal modelling of *Culicoides* biting midges: case studies of incursions into Australia. *BMC Vet Res.* 2014;10:135.
34. Durr PA, Graham K, van Klinken RD. Sellers' Revisited: A Big Data Reassessment of Historical Outbreaks of Bluetongue and African Horse Sickness due to the Long-Distance Wind Dispersion of *Culicoides* Midges. *Front Vet Sci.* 2017;4:98.
35. Agren EC, Burgin L, Lewerin SS, Gloster J, Elvander M. Possible means of introduction of bluetongue virus serotype 8 (BTV-8) to Sweden in August 2008: comparison of results from two models for atmospheric transport of the *Culicoides* vector. *Vet Rec.* 2010;167:484–8.
36. Aguilar-Vega C, Fernandez-Carrion E, Sanchez-Vizcaino JM. The possible route of introduction of bluetongue virus serotype 3 into Sicily by windborne transportation of infected *Culicoides spp.* *Transbound Emerg Dis.* 2019;66:1665–73.
37. Burgin LE, Gloster J, Sanders C, Mellor PS, Gubbins S, Carpenter S. Investigating incursions of bluetongue virus using a model of long-distance *Culicoides* biting midge dispersal. *Transbound Emerg Dis.* 2013;60:263–72.
38. Jacquet S, Garros C, Lombaert E, Walton C, Restrepo J, Allene X, et al. Colonization of the Mediterranean basin by the vector biting midge species *Culicoides imicola*: an old story. *Mol Ecol.* 2015;24:5707–25.
39. Shirk AJ, Landguth EL, Cushman SA. A comparison of individual-based genetic distance metrics for landscape genetics. *Mol Ecol Resour.* 2017;17:1308–17.
40. Landguth EL, Cushman SA, Murphy MA, Luikart G. Relationships between migration rates and landscape resistance assessed using individual-based simulations. *Mol Ecol Resour.* 2010;10:854–62.
41. Frantz AC, Cellina S, Krier A, Schley L, Burke T. Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *J Appl Ecol.* 2009;46:493–505.
42. Piry S, Chapuis MP, Gauffre B, Papaix J, Cruaud A, Berthier K. Mapping Averaged Pairwise Information (MAPI): a new exploratory tool to uncover spatial structure. *Methods Ecol Evol.* 2016;7:1463–75.

43. Mignotte A, Garros C, Gardes L, Balenghien T, Duhayon M, Rakotoarivony I, et al. The tree that hides the forest: cryptic diversity and phylogenetic relationships in the Palaearctic vector *Obsoletus/Scoticus* Complex (Diptera: Ceratopogonidae) at the European level. *Parasit Vectors*. 2020;13:265.
44. Delecolle JC. Nouvelle contribution à l'étude systématique et iconographique des espèces du genre *Culicoides*, (Diptéra): (Cératopogonidae) du Nord-Est de la France. Strasbourg 1985.
45. Garros C, Balenghien T, Carpenter S, Delecolle JC, Meiswinkel R, Pédarrieu A, et al. Towards the PCR-based identification of Palaearctic *Culicoides* biting midges (Diptera: Ceratopogonidae): results from an international ring trial targeting four species of the subgenus *Avaritia*. *Parasit Vectors*. 2014;7:223.
46. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
47. ggplot2: Elegant Graphics for Data Analysis
Ginestet C. ggplot2: Elegant Graphics for Data Analysis. *J Roy Stat Soc Ser A (Stat Soc)*. 2011;174:245–6.
48. Kamvar ZN. JFTaNJGu. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction.
49. Winter DJ. MMOD: an R library for the calculation of population differentiation statistics. *Mol Ecol Resour*. 2012;12:1158–60.
50. Wattier R, Engel CR, Saumitou-Laprade P, Valero M. Short allele dominance as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Mol Ecol*. 1998;7:1569–73.
51. Manangwa O, De Meeus T, Grebaut P, Segard A, Byamungu M, Ravel S. Detecting Wahlund effects together with amplification problems: Cryptic species, null alleles and short allele dominance in *Glossina pallidipes* populations from Tanzania. *Mol Ecol Resour*. 2019;19:757–72.
52. Adamack AT, Gruber B. PopGenReport: simplifying basic population genetic analyses in R. *Methods Ecol. Evol*. 2014;5:384–7.
53. Rice WR. Analysing tables of statistic tests. *Evolution*. 1989;43(1):223–5.
54. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*. 1984;38:1358–70.
55. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945–59.
56. Guillot G, Mortier F, Estoup A. GENELAND: A computer package for landscape genetics. *Mol Ecol Notes*. 2005;5:712–5.
57. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour*. 2009;9:1322–32.

58. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*. 2011;4:359–61.
59. Belkhir K. Genetix 4.05, logiciel sous Windows TM pour la genetique des populations. <http://www.genetix.univ-montp2.fr/genetix/genetix.htm>. 2004.
60. Goslee SC, Urban DL. The ecodist package for dissimilarity-based analysis of ecological data. *J Stat Softw*. 2007;22:1–19.
61. Kimmig SE, Beninde J, Brandt M, Schleimer A, Kramer-Schadt S, Hofer H, et al. Beyond the landscape: Resistance modelling infers physical and behavioural gene flow barriers to a mobile carnivore across a metropolitan area. *Mol Ecol*. 2020;29:466–84.
62. Rousset F. Genetic differentiation between individuals. *J Evol Biol*. 2000;13:58–62.
63. Hardy O, Vekemans X. SPAGeDI: A versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes*. 2002;2:618–20.
64. Loiselle BA, Sork VL, Nason J, Graham C. Spatial Genetic-Structure of a Tropical Understory Shrub, *Psychotria Officinalis* (Rubiaceae). *Am J Bot*. 1995;82:1420–5.
65. Hardy OJ. Estimation of pairwise relatedness between individuals and characterization of isolation-by-distance processes using dominant genetic markers. *Mol Ecol*. 2003;12:1577–88.
66. Nychka D, Furrer R, Paige J, Sain S. *Fields: Tools for spatial data*. 2017.
67. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res*. 1967;27:209–20.
68. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin P, O'Hara RB, et al. *Vegan: Community Ecology Package*. R Package Version. 2.0–10. CRAN. 2013.
69. *A Tool for Landscape Ecology*
McRae B, Shah VB. *Circuitscape. A Tool for Landscape Ecology*. 2008.
70. McRae BH. Isolation by resistance. *Evolution*. 2006;60:1551–61.
71. Neil A, David M, Jolyon m, Kate S, Willy W. A first attempt at modelling roe deer (*Capreolus capreolus*) distributions over Europe. 2014.
72. Gilbert M, Nicolas G, Cinardi G, Van Boeckel TP, Vanwambeke S, Wint WGR, et al: Global cattle distribution in 2010 (5 minutes of arc). V3 edn: Harvard Dataverse; 2018.
73. Carabajal CC, Harding DJ, Boy JP, Danielson JJ, Gesch DB, Suchdeo VP. Evaluation of the Global Multi-Resolution Terrain Elevation Data 2010 (GMTED2010) Using ICESat Geodetic Control. *International Symposium on Lidar and Radar Mapping 2011: Technologies and Applications*. 2011;8286.
74. Wan Z. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sens Environ*. 2014;140:36–45.
75. Fernández-López J, Schliep K. rWind: download, edit and include wind data in ecological and evolutionary analysis. *Ecography*. 2018;42:804–10.

76. Newton RG, Spurrell DJ. Examples of the Use of Elements for Clarifying Regression Analyses. *Appl Stat.* 1967;16:165–72.
77. Nimon K, Lewis M, Kane R, Haynes RM. An R package to compute commonality coefficients in the multiple regression case: an introduction to the package and a practical example. *Behav Res Methods.* 2008;40:457–66.
78. Dellicour S, Prunier JG, Piry S, Eloy MC, Bertouille S, Licoppe A, et al. Landscape genetic analyses of *Cervus elaphus* and *Sus scrofa*: comparative study and analytical developments. *Heredity.* 2019;123:228–41.
79. Prunier JG, Colyn M, Legendre X, Flamand M-C. Regression commonality analyses on hierarchical genetic distances. *Ecography.* 2017;40:1412–25.
80. Hijmans RJ. *geosphere: Spherical Trigonometry.* 2019.
81. Jay F, Sjodin P, Jakobsson M, Blum MG. Anisotropic isolation by distance: the main orientations of human genetic differentiation. *Mol Biol Evol.* 2013;30:513–25.
82. Onyango MG, Beebe NW, Gopurenko D, Bellis G, Nicholas A, Ogugo M, et al. Assessment of population genetic structure in the arbovirus vector midge, *Culicoides brevitarsis* (Diptera: Ceratopogonidae), using multi-locus DNA microsatellites. *Vet Res.* 2015;46:108.
83. Shults P, Ho A, Martin EM, McGregor BL, Vargo EL. Genetic Diversity of *Culicoides stellifer* (Diptera: Ceratopogonidae) in the Southeastern United States Compared With Sequences From Ontario, Canada. *J Med Entomol.* 2020;57:1324–7.
84. Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE. Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conserv Genet.* 2006;7:295–302.
85. Duchesne P, Turgeon J. FLOCK provides reliable solutions to the "number of populations" problem. *J Hered.* 2012;103:734–43.
86. Prunier JG, Kaufmann B, Fenet S, Picard D, Pompanon F, Joly P, et al. Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: towards a better assessment of functional connectivity using an individual-based sampling scheme. *Mol Ecol.* 2013;22:5516–30.
87. Pioz M, Guis H, Crespin L, Gay E, Calavas D, Durand B, et al. Why did bluetongue spread the way it did? Environmental factors influencing the velocity of bluetongue virus serotype 8 epizootic wave in France. *PLoS One.* 2012;7:e43360.
88. Viennet E, Garros C, Gardes L, Rakotoarivony I, Allene X, Lancelot R, et al. Host preferences of Palaearctic *Culicoides* biting midges: implications for transmission of orbiviruses. *Med Vet Entomol.* 2013;27:255–66.
89. Elbers AR, Backx A, Mintiens K, Gerbier G, Staubach C, Hendrickx G, et al. Field observations during the Bluetongue serotype 8 epidemic in 2006. II. Morbidity and mortality rate, case fatality and clinical recovery in sheep and cattle in the Netherlands. *Prev Vet Med.* 2008;87:31–40.
90. Sailleau C, Breard E, Viarouge C, Vitour D, Romey A, Garnier A, et al. Re-Emergence of Bluetongue Virus Serotype 8 in France, 2015. *Transbound Emerg Dis.* 2017;64:998–1000.

91. Diaz-Sanchez S, Hernandez-Jarguin A, Torina A, Fernandez de Mera IG, Estrada-Pena A, Villar M, et al. Biotic and abiotic factors shape the microbiota of wild-caught populations of the arbovirus vector *Culicoides imicola*. *Insect Mol Biol*. 2018;27:847–61.
92. Elbers AR, Meiswinkel R. *Culicoides* (Diptera: Ceratopogonidae) host preferences and biting rates in the Netherlands: comparing cattle, sheep and the black-light suction trap. *Vet Parasitol*. 2014;205:330–7.
93. Ayllon T, Nijhof AM, Weiher W, Bauer B, Allene X, Clausen PH. Feeding behaviour of *Culicoides spp.* (Diptera: Ceratopogonidae) on cattle and sheep in northeast Germany. *Parasit Vectors*. 2014;7:34.
94. Dijkstra E, van der Ven IJ, Meiswinkel R, Holzel DR, Van Rijn PA, Meiswinkel R. *Culicoides chiopterus* as a potential vector of bluetongue virus in Europe. *Vet Rec*. 2008;162:422.
95. Sedda L, Brown HE, Purse BV, Burgin L, Gloster J, Rogers DJ. A new algorithm quantifies the roles of wind and midge flight activity in the bluetongue epizootic in northwest Europe. *Proc. Biol. Sci*. 2012;279:2354-62.
96. Sedda L, Morley D, Brown HE. Characteristics of Wind-Infective Farms of the 2006 Bluetongue Serotype 8 Epidemic in Northern Europe. *EcoHealth*. 2015;12:461–7.
97. Dlugosch KM, Parker IM. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol Ecol*. 2008;17:431–49.
98. White BP, Pilgrim EM, Boykin LM, Stein ED, Mazor RD. Comparison of four species-delimitation methods applied to a DNA barcode data set of insect larvae for use in routine bioassessment. *Freshwater Science*. 2014;33:338–48.
99. Bock DG, Caseys C, Cousens RD, Hahn MA, Heredia SM, Hubner S, et al. What we still don't know about invasion genetics. *Mol Ecol*. 2015;24:2277–97.
100. Girod C, Vitalis R, Leblois R, Freville H. Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the Msvar method. *Genetics*. 2011;188:165–79.

Tables

Table 1: Results of multiple regressions on distance matrices (MRDM) and additional parameters derived from commonality analysis (CA). Pearson's correlation coefficient (r), significant regression coefficient (β), as well as unique (U) and common (C) contributions of environmental distances to the variance in the dependent variable. [C]/[R] indicate if the considered environmental raster was treated as a conductance or as a resistance factor, and k corresponds to the re-scaling parameter used to transform the initial raster file.

	k	r	β	U	C
FCA					
Cattle density [C]	1000	0.1733	0.1551	0.0221	0.0079
Grassland [R]	1000	0.1080	0.0637	0.0037	0.0079
a_R	k	r	β	U	C
Elevation [R]	100	0.0254	-0.9945	0.00044	0.0002
Elevation [R]	1000	0.0291	0.9860	0.00048	0.0004
Grassland [R]	100	0.0448	0.2279	0.00005	0.0020
Grassland [R]	1000	0.0539	-0.1483	0.00002	0.0029
Urban areas [R]	100	0.0725	-0.1279	0.00014	0.0051
Urban areas [R]	1000	0.0879	0.2156	0.00049	0.0072

Figures

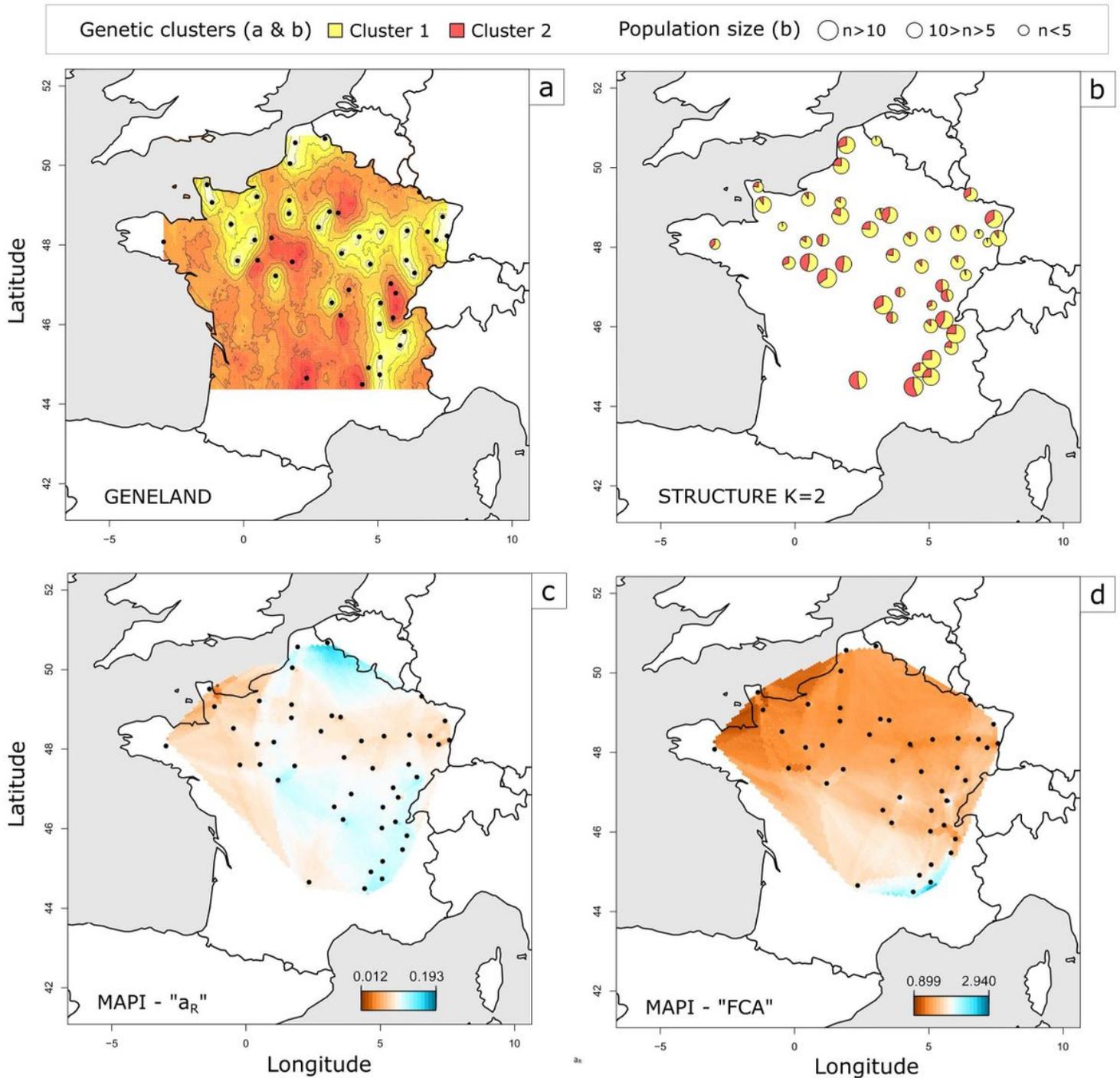


Figure 1

Genetic clustering and genetic differentiation of *C. obsoletus*. Results of the genetic clustering analyses performed with GENELAND (a) and STRUCTURE (b), as well as smoothing of pairwise measures performed with MAPI and based on (c) Rousset's (a_R) and (d) FCA inter-individual genetic distances (see the text for further details). One specific color has been assigned to each genetic cluster in panels (a) and (b). In panels (c) and (d), genetic dissimilarity is represented by a color scale ranging from red (lower genetic dissimilarity) to blue (higher genetic dissimilarity).

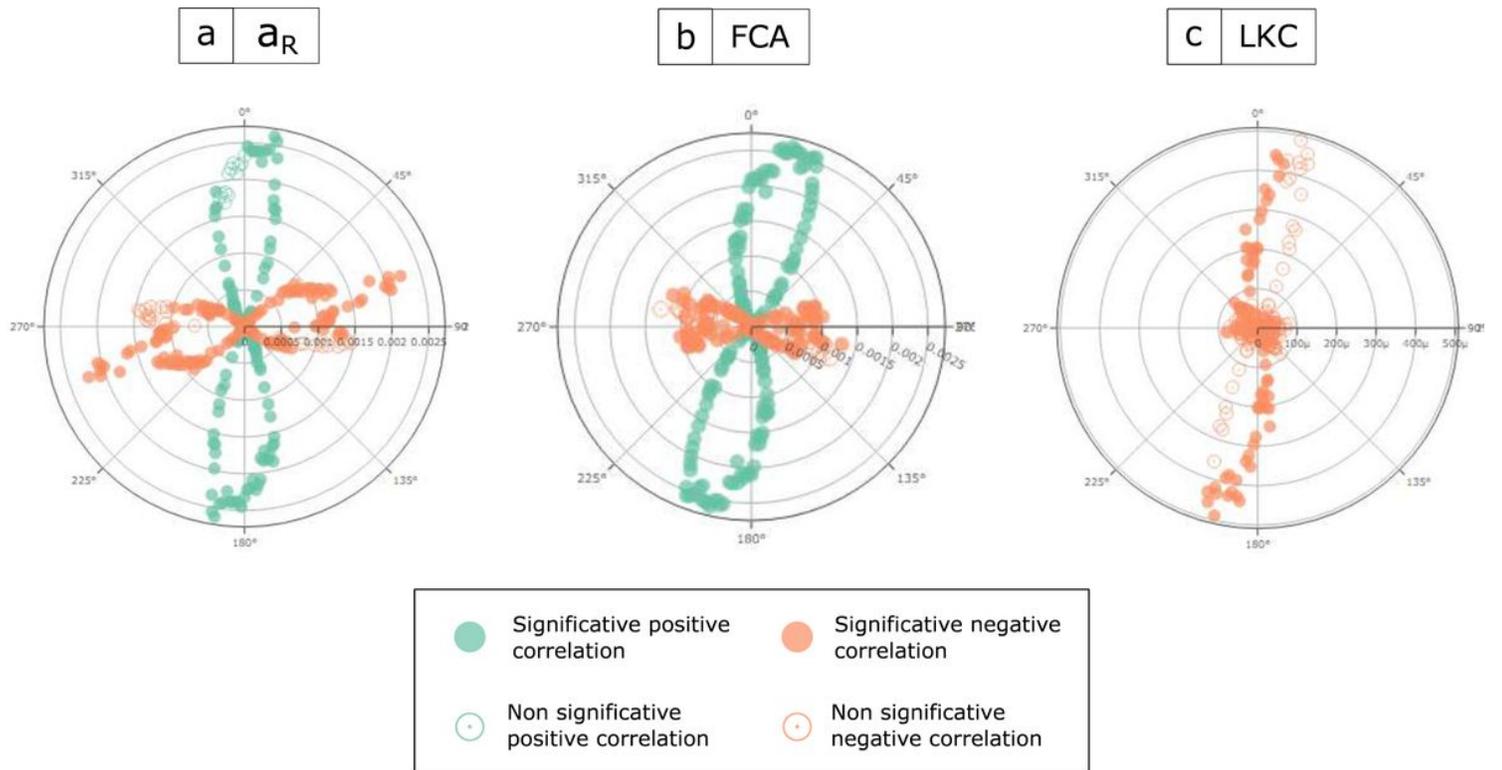


Figure 2

Results of anisotropic isolation-by-distance analyses. Polar plots show the correlation between geographical projected distances by angle and inter-individual genetic distances (aR, LKC, and FCA).

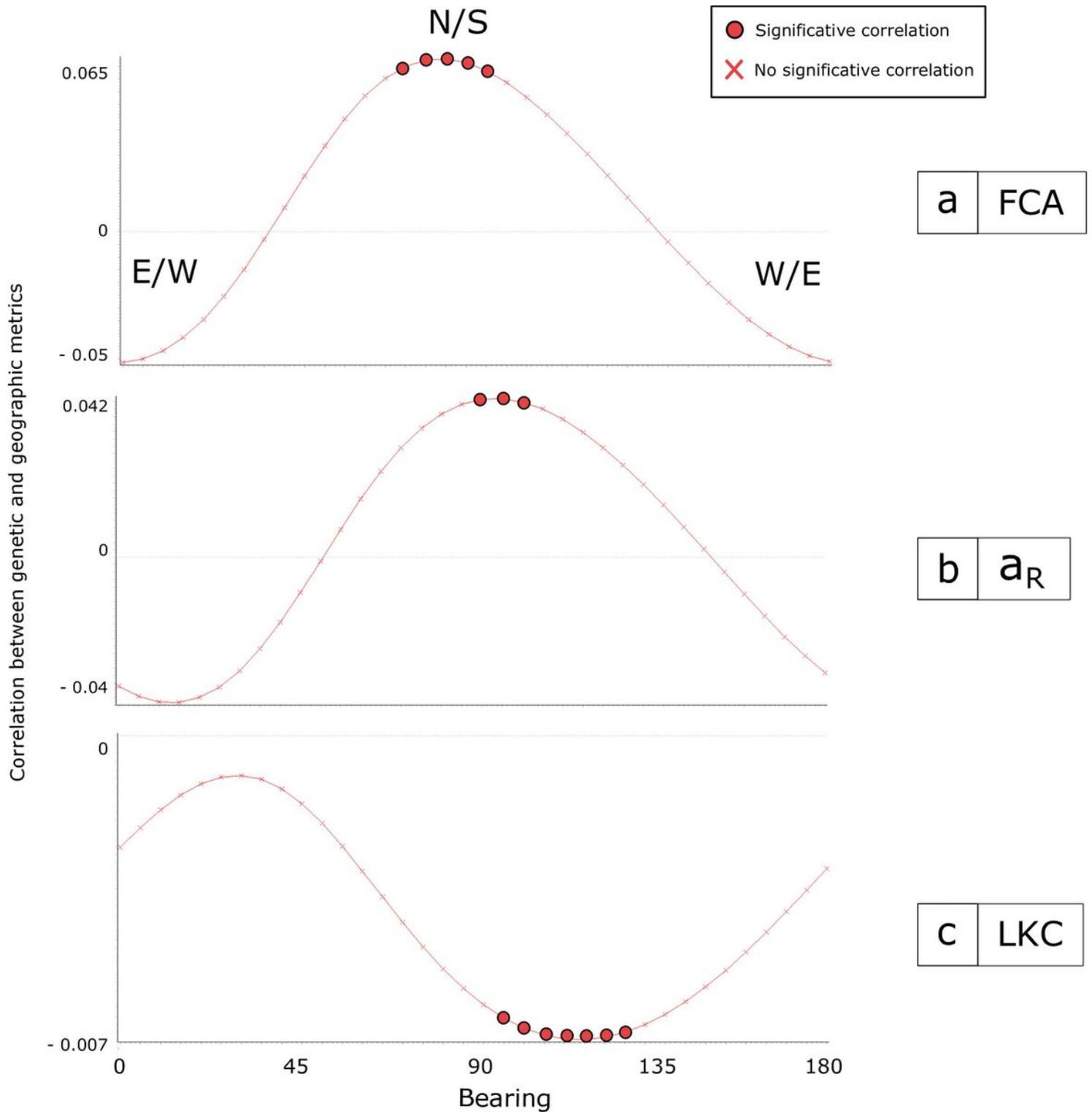


Figure 3

Bearing analysis: correlation between genetic (a_R, LKC, and FCA) and geographical distances as a function of the angle between sampling sites. Circles and crosses indicate significant and non-significant values, respectively. Abbreviations “E”, “W”, “S” and “N” refer to “East”, “West”, “South” and “North”, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS6structurefrk4.eps](#)
- [FigS5resultstrharvester.eps](#)
- [FigS4IBDIND.eps](#)
- [FigS3workflow.eps](#)
- [FigS2windmaptodir.eps](#)
- [FigS1rasterused.eps](#)
- [TabS6univarfcaar.xlsx](#)
- [TabS5divloc.xlsx](#)
- [TabS4microsatmarkers.xlsx](#)
- [TabS3originmarkers.xlsx](#)
- [TabS2refseq.xlsx](#)
- [TabS1tabledivbypop.xlsx](#)