High dispersal capacity of *Culicoides obsoletus* (Diptera: Ceratopogonidae), vector of bluetongue and Schmallenberg viruses, revealed by landscape genetic analyses

Antoine Mignotte[1,2*], Claire Garros[1,2], Simon Dellicour[3,4], Maude Jacquot[3,5], Marius Gilbert[3], Laetitia Gardès[1,6], Thomas Balenghien[1,7,8], Maxime Duhayon[1,2], Ignace Rakotoarivony[1,2], Maïa de Wavrechin[1,2] and Karine Huber[1]

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 flow 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 were genotyped using 13 newly designed microsatellite markers.

Results: We found low genetic differentiation and a weak 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 did detect significant anisotropic isolation by distance on a north-south axis. We employed a multiple regression on distance matrices approach to investigate the correlation between genetic and environmental distances. Among all the environmental factors that were tested, only cattle density seems to have an impact on *C. obsoletus* gene flow.

Conclusions: The high dispersal capacity of *C. obsoletus* over land found in the present study 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 disease control.

Keywords: *Culicoides obsoletus*, Landscape genetics, Microsatellite, Dispersal, Palearctic region

Background

During a vector-borne disease outbreak, a precise understanding of the dispersal capacity of the vector species is key to implementing 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") the latter can be categorized as a conductance factor [1]. On the contrary, when dispersal is limited by a landscape characteristic, the latter can be categorized as a resistance factor [1, 2]. However, some types of dispersal allow organisms to overcome resistance

*Correspondence: antoine.mignotte@cirad.fr
[1] ASTRE, Univ Montpellier, Cirad, INRAE, Montpellier, France
Full list of author information is available at the end of the article

Mignotte *et al. Parasites Vectors*   (2021) 14:93

Page 2 of 14

factors, as 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 Schmallenberg virus (SBV) in 2011–2012 [4]. Both viruses have spread very quickly and widely 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 the western Palearctic region [5–7]. *Culicoides obsoletus* is a very widespread species, which is 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 [10–14]. For example, high ambient temperatures reduce the incubation period of the virus [15], whereas high precipitation and wind speeds can reduce *Culicoides* flight activity [16–18].

Previous mark-release-recapture studies on *Culicoides* species showed that the post-release dispersal distance traveled 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 traveled distance recorded was 6 km for *Culicoides mohave* in a particular desert landscape [20]. Yet, the potentially high dispersal capacity of *Culicoides* and the high mortality of adults when they are manipulated represent practical limitations in the context of mark-release-recapture procedures. Many studies have reported a correlation between disease movement and wind-borne transport of *Culicoides* during outbreaks [3, 18, 21–26]. In 2017, Sanders et al. attempted to quantify *Culicoides* dispersal over land, and demonstrated through a capture-mark-recapture study that 84.4% of flights of more than 1 km took place downwind, while only 15.6% of flights were made upwind [27]. The introduction of BTV serotypes by wind-borne infected *Culicoides* has been demonstrated from northern Africa to southern Europe [28], from Kenya to Southwest Indian Ocean islands [29], from Sardinia to the Balearic Islands [25], from France to Corsica [30], in Ireland [31], and in the UK [32]. Most of these studies were supported by modeling analyses in which dispersal trajectories were evaluated by atmospheric dispersion models over water bodies [31, 33–37]. *Culicoides* dispersal over land has been investigated less than its dispersal over water [38], and little attempt has been made to link the former to environmental factors. Given the 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 studying the impact of landscape on gene flow generally use population-based sampling and are often based on a limited number of sampled populations [29, 39]. However, *Culicoides*, like other dipteran species, are not spatially structured into separate populations and must be considered as a continuum of individuals heterogeneously distributed across a landscape. In order to identify the environmental factors that have an impact on the dispersal of *Culicoides*, an individual approach is more relevant as it avoids the misinterpretation of inter-population genetic distance [40]. Indeed, the individual approach in landscape genetics aims at maximizing the number of sampling sites, and thus brings much greater statistical power to the detection of spatial patterns of genetic differentiation and the environmental factors that cause them [41]. 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 [42]. It is therefore important to also incorporate a visualization approach into the analytical workflow [43]. Mapping averaged pairwise information (MAPI) allows the visual comparison of genetic dissimilarity with some environmental factors, and also the development of working hypotheses. On the other hand, it is also necessary to use statistical analyses of genetic and environmental distances as a complement to MAPI.

The aims of the present study are to determine the inland connectivity of populations of a main vector species at large geographical scales 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 propose a complementary framework integrating multiple approaches which can be applied more generally to the study of gene flow and its links with environmental factors.

Methods

Culicoides obsoletus sampling and morphological identification

The data set analyzed in the present study is composed of 368 individual female biting midges collected in 46 sites in France in April 2011, using the national surveillance network for *Culicoides* populations or local

Mignotte *et al. Parasites Vectors*   (2021) 14:93

Page 3 of 14

complementary collections (Additional file 1: Table S1). Collections were made overnight using Onderstepoort Veterinary Institute light traps set up at farms near cattle or sheep. All insects were stored in 70% ethanol. Morphological identification to species level was carried out in the same way as described in a previous study [44] under a binocular microscope using available identification keys [45].

DNA extraction, amplification, sequencing and sequence analyses

DNA was extracted from a total of 368 individuals belonging to the *C. obsoletus* species complex, and cytochrome *c* oxidase (COI) was amplified and sequenced following the procedure detailed in Mignotte et al. [44]. To ensure that only *C. obsoletus* individuals were included in the final data set [46], sequence assignation was run according to Mignotte et al. [44] (the reference COI sequences used for specific assignment of all individuals are available in Additional file 2: Table S2). All COI sequences were then aligned using the multiple sequence comparison by log-expectation algorithm [47] implemented in the software package GENEIOUS version 6.0.5 (Biomatters, http://www.geneious.com). In order to estimate unbiased gene flow 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. [48]. *Culicoides obsoletus* samples from France and the UK (Additional file 3: Table S3) were used as biological material to prepare a paired-end DNA library (Illumina Nextera technology) by using the Nextera DNA Sample Kit (ref. GAO9115; Illumina, San Diego, CA). The sequencing was performed on a MiSeq Sequencer [2 × 300-base pair (bp) read mode]. In order to maximize the length of the sequences, an assembly of the 2,299,075 reads was performed with the ABySS assembler [49]. The simple sequence repeat (SSR) loci search of di- and trinucleotide motifs was done with the MIcroSAtellite identification tool [50]. Primer design was done with the Primer3 software [51]. Among the 82,276 SSRs identified, primers were designed for 43,676 SSRs, of which 15,888 were dinucleotides, 23,656 were trinucleotides, and 4132 contained complex SSR motifs. For the SSR screening, we selected 3135 primer pairs flanking dinucleotide SSR motifs with a minimum of 12 repeats and amplifying fragments between 150 and 300 bp in length and a half-denaturation temperature 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) according to the protocol described in the manufacturer's manual and the annealing temperature given in Additional file 4: 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 (Additional file 4: 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 (Applied Biosystems, Life Technologies) with double blind reading to limit the potential interpretation bias from the reader.

A first validation of the polymorphisms 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 (Additional file 4: Table S4).

The previously extracted DNA of the 368 *C. obsoletus* was then amplified and genotyped at these 13 microsatellite loci. Genetic variability parameters, such as the number of alleles per locus, the allele size range, and the observed and expected heterozygosity were estimated per locus, and per population over the entire dataset using the R packages ggplot [52], poppr [53] and mmod [54]. Short allele dominance can be a source of heterozygote deficiency in a microsatellites data set [55, 56]. In order to avoid such bias, Spearman's correlations rank between the size of the alleles and their frequencies were calculated in R [55]. Null allele frequencies were estimated using the R package PopGenReport [57], and a Bonferroni correction was applied to all matched tests to take into account potential biases associated with multiple comparisons [58]. Correlation between F_{IS} and null alleles was calculated in R, and loci with significant correlations were removed. The presence of an imbalance in genetic binding between each pair of loci was tested by Fisher's exact test performed with the R package poppr [53]. The same R package was used along with the pegas R package to identify potential gaps in the panmictic matching regime. For this purpose, we performed chi-square tests to compare the observed heterozygosity deficits to the expected heterozygosity under the Hardy-Weinberg equilibrium with a significance threshold of 5%. We also calculated the fixation index (F_{IS}) [59].

Mignotte *et al. Parasites Vectors*   (2021) 14:93

Page 4 of 14

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) [60] and GENELAND (version 4.5.0) [61]. 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 one to (i) assume that allele frequencies are similar among populations; (ii) estimate for each individual the probability that it belongs to each K cluster; and (iii) use sampling locations as prior information to assist the clustering, which is particularly suitable when there is weak genetic structuring [62]. We performed ten independent runs for each value of K varying from 1 to 10. Each run consisted of 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 [64]. 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 from 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 version 4.05.2 [65]. The first ten FCA axes were used to calculate an Euclidean distance between all individuals using the R package ecodist [66], hereafter called the "FCA" distance [67]. The second inter-individual genetic distance is Rousset's distance (a_R), [68], which corresponds to $F_{ST}/(1 - F_{ST})$, but for pairs of individuals. We used the SPAGeDi 1.5 program [69] to compute a_R genetic distance. The third inter-individual genetic metric is Loiselle's kinship coefficient (LKC) [70], which, unlike FCA and a_R that are measures of genetic dissimilarity, is an index of similarity [71]. We used the R package fields [72] 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 [72]. Mantel tests [73] were performed with the R package vegan [74]. Inter-individual genetic distances were mapped using the program MAPI [43]. The method implemented in MAPI allows the visualization of spatial variation in genetic dissimilarity. It can also define areas where genetic dissimilarity is significantly lower or higher than expected by chance [43]. 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 [75, 76] 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) [77], European index of distribution of red deer (C) [77], global cattle, sheep and goat distributions in 2010 (C) [78].

ii. Terrestrial habitat: mean global elevation in 2010 (R) [79] 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 (http://www.eea.europa.eu; resolution: ∼ 100 m) (Additional file 5: 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 means [80]. 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

Mignotte *et al. Parasites Vectors*   (2021) 14:93

Page 5 of 14

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 (Additional file 6: Fig. S2) [81].

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. [84], 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 regression on distance matrices (MRDM) coupled with commonality analyses [82] 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 [66] and yhat [83]. 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 obtained from the linear regression based on environmental distances computed on the null raster were selected for the multivariate analyses (MRDM-commonality analyses) [84]. The same criteria used to identify suppressors were used for the multivariate analysis as in the commonality analyses descriptions [80, 81]. These suppressors allow one to select environmental variables that are more explanatory than the geographical distance alone. An environmental variable is considered to be a suppressor if the regression coefficient and correlation, or the unique contribution and the common contribution, are of opposite signs [85].

Anisotropic isolation by distance
Anisometric isolation by distance is a method used to identify directional gene flow and orientation at which the accumulation of genetic differentiation is the greatest. Thus the notion of anisotropy applied to isolation by distance means that the intensity and significance of isolation by distance varies according to the direction considered. 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 [86] 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 [72]. The projected distance matrix was calculated for each angle between all sampling sites, using the formula below. When calculating 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.

$$log\,(|d_{AB} * cos(a_{AB})|) \tag{1}$$

A linear regression of the genetic distances on the projected geographical distances obtained for each angle was then performed. The angle that maximizes the R^2 of this regression with a positive regression coefficient was considered as the angle maximizing the isolation by distance signal [87]. 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 is described in Additional file 7: 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 (Additional file 3: Table S3), and 368 *C. obsoletus* from 46 populations were genotyped with 13 microsatellite loci (Additional file 4: Table S4). A filter was also 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 (Additional file 8: Table S5). The expected heterozygosity, which varied from 0.918 (OBSms5) to 0.404 (OBSms29), was always higher than the observed heterozygosity, which varied from 0.872 (OBSms13) to 0.197 (OBSms25), reflecting a heterozygous deficit (Additional file 8: Table S5). To ensure that all markers provided independent information, linkage disequilibrium between each pair of loci was tested. After Bonferroni correction, no significant linkage disequilibrium was observed between any loci. The OBSms25 locus was also the only locus with more than 5% missing data. Thus, only markers with a null percentage lower than 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

Mignotte *et al. Parasites Vectors*   (2021) 14:93

Page 6 of 14

rate of null alleles ($r = 0.947$, $p > 0.001$), 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 (Additional file 8: Table S5). The genetic resolution of microsatellite markers depends on the number of markers and their polymorphisms. All 11 selected loci were therefore polymorphic (Additional file 8: Table S5), and the number of alleles present at a given locus varied from 40 for OBSms5 to nine for OBSms29. No significant isolation by distance was detected with the different inter-individual genetic distances, a_R ($R^2 < 0.001$, $p = 0.439$; Additional file 9: Fig. S4), LKC ($R^2 < 0.001$, $p = 0.992$; Additional file 9: Fig. S4), and FCA ($R^2 < 0.001$, $p = 0.361$; Additional file 9: 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 (Additional file 10: 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 was well supported ($Q > 0.90$), no clear association between spatial clustering and distribution was observed, i.e. there was no association between *Culicoides* sampling location and the genetic cluster distributions. However, a west-east transect seemed to emerge, as samples collected on this transect were more frequently assigned to cluster 1, contrary to the north-south transect which was more frequently represented by cluster 2. Genetic clusters inferred by GENELAND and STRUCTURE ($K = 2$) were consistent with these results, both clusters being found at all the sampling sites (Fig. 1a, b). Results for the second optimal K value ($K = 4$) of the STRUCTURE program gave a similar impression, reinforcing the conclusion that the populations are genetically weakly structured (Additional file 11: Fig. S6). Additional STRUCTURE analysis was carried out with only loci OBSms 4, 11 and 28, which are at Hardy Weinberg equilibrium, and no additional structure was detected using only these loci (data not shown). Genetic dissimilarity resulting from MAPI analyses with the a_R genetic distance showed a non-homogeneous pattern of dissimilarity. The maximum inter-individual genetic dissimilarity was 0.182 and the lowest was 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 the Alsace (northeastern France) showed low genetic dissimilarity (Fig. 1c, d). MAPI analyses with FCA inter-individual genetic distance showed a homogeneous genetic dissimilarity. However, almost none of the areas were 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 $k = 100$ ($Q = 0.009$, $p = 0.001$) and $k = 1000$ ($Q = 0.029$, $p = 0.001$), with sheep distributions for $k = 100$ ($Q = 0.002$, $p = 0.002$) and for $k = 1000$ ($Q = 0.006$, $p = 0.001$), with elevation for $k = 100$ ($Q = 0.001$, $p = 0.005$) and for $k = 1000$ ($Q = 0.001$, $p = 0.006$), and with grassland for $k = 100$ ($Q = 0.009$, $p = 0.001$) and for $k = 1000$ ($Q = 0.011$, $p = 0.001$) distances (Additional file 12: Table S6). Similarly, a_R inter-individual genetic distances were significantly associated with distances computed from elevation for $k = 100$ ($Q = 0.00018$, $p = 0.006$) and for $k = 1000$ ($Q = 0.00038$, $p = 0.007$), with grassland for $k = 100$ ($Q = 0.001$, $p = 0.001$) and for $k = 1000$ ($Q = 0.002$, $p = 0.001$) and with urban areas for $k = 100$ ($Q = 0.004$, $p = 0.001$) and for $k = 1000$ ($Q = 0.007$, $p = 0.001$) rasters (Additional file 12: Table S6). However, no significant association was found between LKC genetic distance and the environmental distances (Q values always < 0).
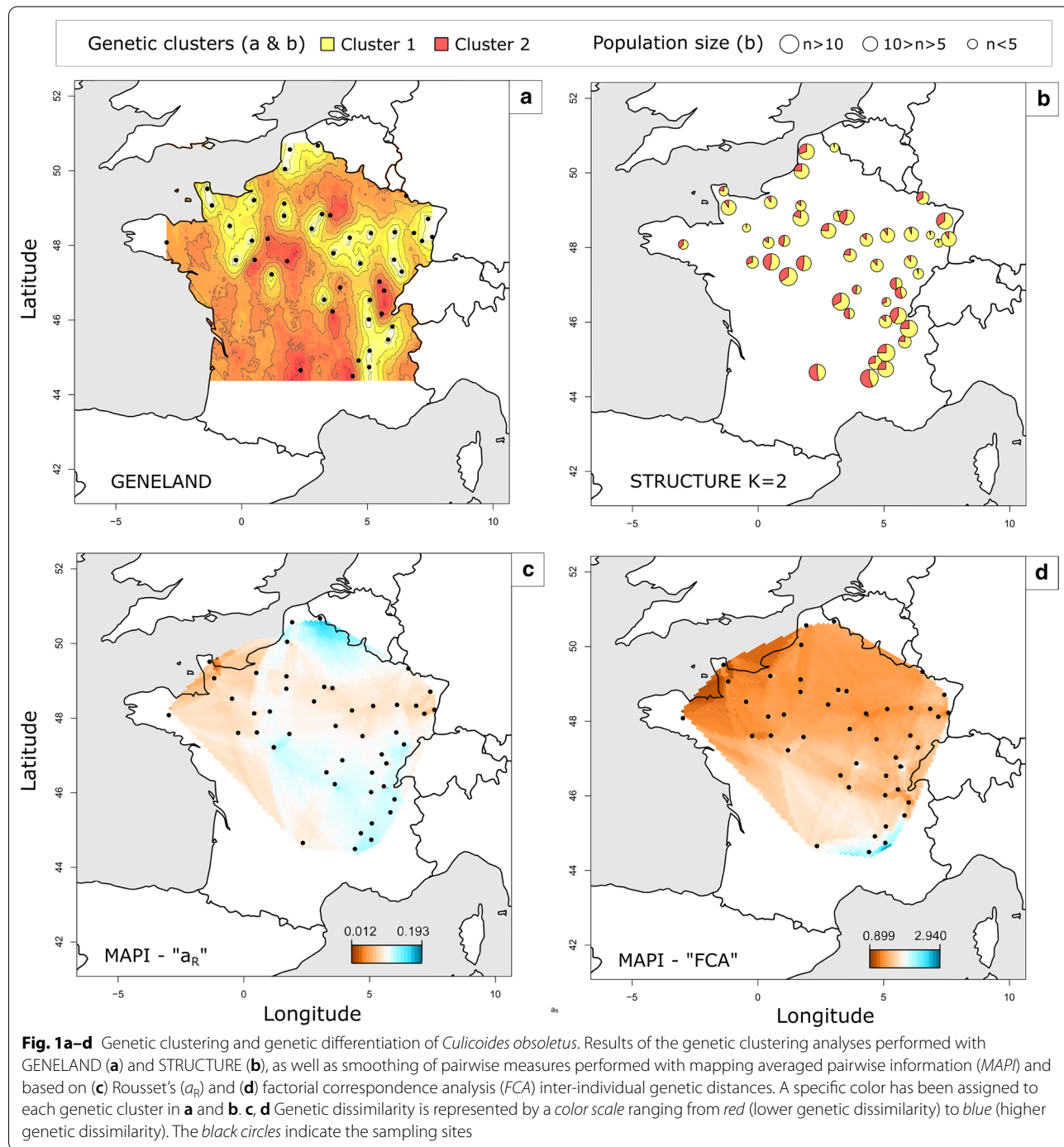
Multivariate regressions revealed a significant positive association of FCA inter-individual genetic distance with cattle distributions for $k = 1000$ ($r = 0.1733$, $p < 0.001$) and with grassland for $k = 1000$ ($r = 0.1080$, $p < 0.001$) (Table 1; Additional file 12: Table S6). Also, multivariate 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% ($U = 0.02$) (Additional file 12: 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 the north–south orientation and a significant negative correlation between genetic distances and orientational distances computed along the west-east orientation (Fig. 2a,

b). For the analysis with LKC, which is a variable of genetic similarity, 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).

Fig. 1a–d Genetic clustering and genetic differentiation of *Culicoides obsoletus*. Results of the genetic clustering analyses performed with GENELAND (**a**) and STRUCTURE (**b**), as well as smoothing of pairwise measures performed with mapping averaged pairwise information (*MAPI*) and based on (**c**) Rousset's (*a_R*) and (**d**) factorial correspondence analysis (*FCA*) inter-individual genetic distances. A specific color has been assigned to each genetic cluster in **a** and **b**. **c**, **d** Genetic dissimilarity is represented by a *color scale* ranging from *red* (lower genetic dissimilarity) to *blue* (higher genetic dissimilarity). The *black circles* indicate the sampling sites

Mignotte *et al. Parasites Vectors*   (2021) 14:93

Page 8 of 14

Discussion

To the best of our knowledge, this is the first 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 scale of France with the set of inter-individual genetic distances used. However, an anisotropic isolation by distance analysis revealed significant distance isolation on the north-south axis (Additional file 13: Figure S7).

High level of gene flow between *C. obsoletus* populations

The low overall inter-individual genetic dissimilarity highlighted by our study reflected the high level of gene flow for *C. obsoletus*, and reinforces what has been described previously for other *Culicoides* species. In particular, high levels of gene flow have already been observed in France for *Culicoides imicola* [30] and in Australia for *Culicoides brevitarsis* [88]. Genetic studies on *C. imicola* in Europe revealed a high level of gene flow, as reflected by the inference of two large genetic clusters: a "central Mediterranean cluster" including Algeria, Sardinia, Corsica, and the Pyrénées-Orientales and Var departments of France, and a "western Mediterranean cluster" including Morocco, Spain, Portugal and Majorca [26]. In North America, similar results were also found for *Culicoides stellifer*. No barriers to gene flow could be identified in this species in the southeastern USA [89]. In addition, no isolation by distance has been observed at the scale of France. Identifying a genetic structure pattern is a great 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 [90], when the differentiation is weaker, such as in the case of highly dispersed organisms, it may encounter difficulties [91]. This underlines the importance of using complementary tools of visualization approaches like MAPI. In addition, the use of an individual approach avoids artificially considering individuals grouped into populations solely due to the effect of the trapping sites [2, 92]. Although our method of analysis seems complementary and coherent for the detection of genetic structure, *C. obsoletus* is not genetically or geographically structured at the scale of France.

Livestock densities as potential drivers of *C. obsoletus* gene flow

It is undeniable that high gene flow homogenizes 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 the environmental factors tested was very small, the contribution of cattle density was the strongest that we detected. Host density as a conductance factor for *Culicoides* is obviously consistent with the biology of the species. These results are in line with those obtained in landscape genetics studies of BTV, which identified distributions of cattle and sheep as key factors in BTV dispersal [13]. In addition, previous studies showed that dairy cattle density was negatively correlated with BTV spread. Although paradoxical with respect to the previous conclusions, this could be explained by the fact that dairy cattle tend to cluster around the milking parlor and move little, and thus form a fixed feed source, which limits the spread of *Culicoides*. On the contrary, beef cattle disperse much more in pastures and thus could encourage the dispersal of *Culicoides* when they seek a source of blood [93]. In view of these results, and of the marked preference of certain species of *Culicoides* for cattle [94], and of BTV emergence or re-emergence events in cattle in the Netherlands in 2006 and in France in 2015 [95, 96], it seems crucial to survey and closely monitor the movements of *Culicoides* in the vicinity of beef cattle farms [94].

In addition, it has been shown that BTV spread is facilitated at low elevation, i.e. up to 300 m [39, 93].

Table 1 Results of multiple regression on distance matrices and additional parameters derived from commonality analysis
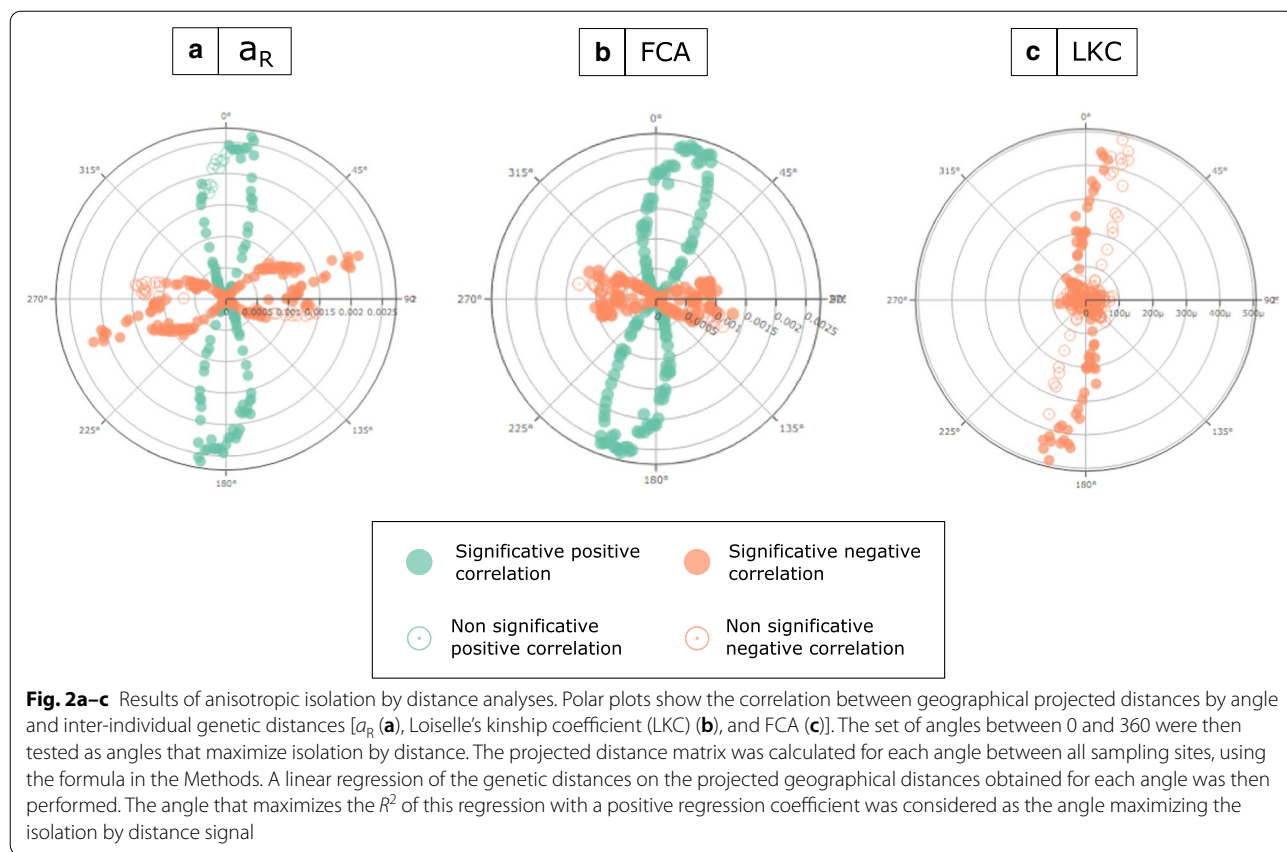
	k^a	r	β	U	C
FCA					
Cattle density [C][b]	1000	0.1733	0.1551	0.0221	0.0079
Grassland [R][c]	1000	0.1080	0.0637	0.0037	0.0079
a_R					
Elevation [R][c]	100	0.0254	−0.9945	0.00044	0.0002
Elevation [R][c]	1000	0.0291	0.9860	0.00048	0.0004
Grassland [R][c]	100	0.0448	0.2279	0.00005	0.0020
Grassland [R][c]	1000	0.0539	−0.1483	0.00002	0.0029
Urban areas [R][c]	100	0.0725	−0.1279	0.00014	0.0051
Urban areas [R][c]	1000	0.0879	0.2156	0.00049	0.0072

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 are shown; *FCA* factorial correspondence analysis, a_R Rousset's distance

[a] Re-scaling parameter used to transform the initial raster file

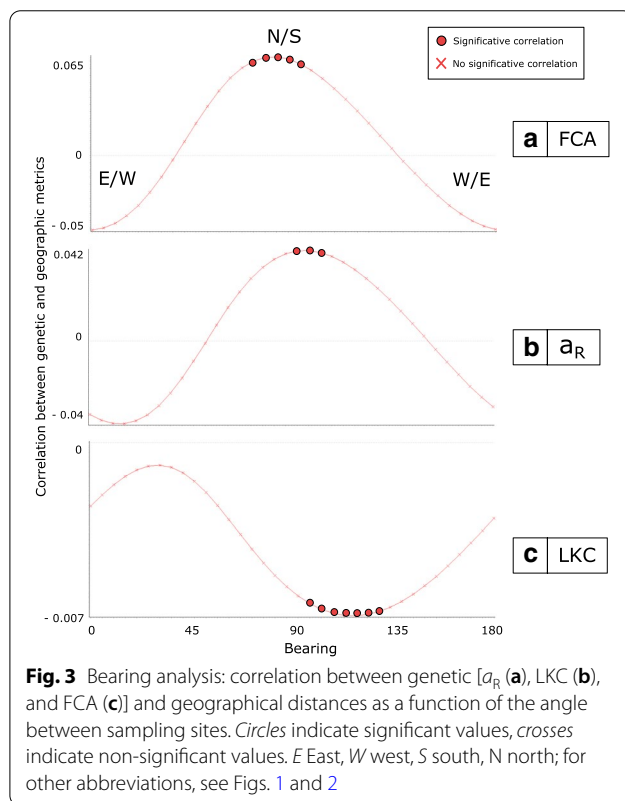[b] The considered environmental raster was treated as a conductance [C] factor

[c] The considered environmental raster was treated as a resistance [R] factor

Fig. 2a–c Results of anisotropic isolation by distance analyses. Polar plots show the correlation between geographical projected distances by angle and inter-individual genetic distances [a_R (**a**), Loiselle's kinship coefficient (LKC) (**b**), and FCA (**c**)]. The set of angles between 0 and 360 were then tested as angles that maximize isolation by distance. The projected distance matrix was calculated for each angle between all sampling sites, using the formula in the Methods. A linear regression of the genetic distances on the projected geographical distances obtained for each angle was then performed. The angle that maximizes the R^2 of this regression with a positive regression coefficient was considered as the angle maximizing the isolation by distance signal

Altitude, an environmental factor tested in this study, failed to explain the inter-individual genetic differentiation of *C. obsoletus*. Altitude, therefore, does not act as a barrier between sampling sites located at relatively low altitudes. It would thus be interesting to integrate into future studies sampling sites at high altitude, i.e. above 1000 m, which are feasible habitats for *C. obsoletus* as it has a large altitudinal range [97]. However, it is possible that 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 an extreme generalist and can take a blood meal from a wide range of hosts [98]. The study of phylogenetically close but ecologically very different species of *Culicoides*, such as *Culicoides chiopterus* [99], might then allow the identification of very different dispersal patterns and the importance of bovine density, *C. chiopterus* being exclusively dependent on cattle for egg-laying [100].

Long-range dispersal of *C. obsoletus*

The most genetically dissimilar individuals were mainly from the southernmost populations of the sampling area. Multiple non-exclusive lines of argument might explain the significant anisotropic isolation by distance observed on a north–south axis in France.

Fig. 3 Bearing analysis: correlation between genetic [a_R (**a**), LKC (**b**), and FCA (**c**)] and geographical distances as a function of the angle between sampling sites. *Circles* indicate significant values, *crosses* indicate non-significant values. *E* East, *W* west, *S* south, *N* north; for other abbreviations, see Figs. 1 and 2

First, the significant anisotropic isolation by distance could be explained by wind dispersal. Dispersion phenomena caused by wind currents, mainly over seas, have already been established for *Culicoides* [3, 18, 21–25, 30]. The resultant dispersal events can be described as passive and active, respectively, because recapture was achieved downwind and upwind of the prevailing wind direction [19], although the activity of *Culicoides*, and thus its active dispersal, is reduced when the wind speed is higher than 3 m/s [16]. Examination of the map of the average wind directions in France over the last 10 years provides potential explanations for this. It can be seen that the southernmost sampling sites (with the most dissimilar individuals) are located in an area where the wind direction is different from that in the rest of France. It should be noted that the diffusion of BTV, and thus the dispersion of *Culicoides*, has already been associated with wind direction. For example, 2% of BTV infections occurred at distances greater than 31 km [101, 102] during the 2006 epizootic. The south of France was initially sampled; however, poor conservation of *Culicoides* DNA did not allow us to perform barcoding and microsatellite genotyping of these populations. A study of this geographical area could thus complete the findings of the present study and potentially identify a stronger genetic structure. This would make it possible to decide whether the premise of differentiation observed in the present study is due to an older phylogeographic structure, different wind dispersion in this geographical area, or a random pattern. It is also questionable whether the indications of very weak genetic structure could be due to differences in genetic diversity between the southern and the northern populations. Lower genetic diversity in the north could then be a sign of an expansion process that contradicts the interpretation of large-scale dispersion [103, 104]. Indeed, the spatial expansion of a population is generally accompanied by gradients of reduction in the genetic diversity of the population during the expansion process, caused by serial founder effects, which create a genetic bottleneck through a founding event [105, 106]. 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 of the sampling methods used. Indeed, if the extent of the sampling varies depending on direction, the distances projected from the angles may represent different distance distributions and lead to the over-representation of the values of strong genetic metrics in the direction of the scatter plot 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 with which we have achieved 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.

Considerations for future work

Our results underline the importance of the methodological development of wind dispersal models for *Culicoides* not only over water but also on land. However, studies of landscape genetics remain indispensable and complementary in order to improve the accuracy of predictive models for *Culicoides* dispersal over land through the integration of meteorological, landscape and activity-based parameters previously tested and validated in landscape genetics workflows. In addition, population sampling of the same species at the European level, i.e. sampling in a very large proportion of the known range of the species, is necessary to observe more marked structuring at the European level and to estimate more precisely the gene flow. For example, *C. imicola* in Europe is structured into two large genetic clusters, the European central cluster and the European western cluster [26]. Moreover, the phylogeographic history of *C. obsoletus* in view of its geographical distribution throughout Europe and North America has only been little explored. On the contrary, the study of the phylogeography of *C. imicola*, the main Afrotropical vector, has shown that its range is from the northern part of sub-Saharan Africa to the Mediterranean Basin [39]. 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 studies, the use of high-throughput sequencing approaches using markers such as double-digest restriction-site associated DNA sequencing can provide greater resolution in view of the large number of single nucleotide polymorphisms (SNPs) revealed at a local scale, and improve our understanding of the active and passive dispersal of *Culicoides*. It could also be relevant to include more microsatellite markers or SNPs to improve genetic resolution and observe the matching and assignment of each individual.

Conclusions

To the best of our knowledge, this study provides the first complete landscape genetic analysis of *C. obsoletus*, a major vector species of animal viruses in Europe. This study shows that the genetic structure of populations at the scale of a country can become homogeneous through large-scale dispersion. Our results demonstrate, for *C. obsoletus*, a very high inland dispersal and vectorization capacity, which has to be taken into consideration in further work on vector

Mignotte *et al. Parasites Vectors*   (2021) 14:93

Page 11 of 14

competence and epidemiological modeling of disease transmission. Wind direction could be a key factor in the dispersal of many insect vector species. Futures studies should increase their geographical extent to cover the entire area of species distribution and enable a better understanding of the limits of *Culicoides* gene flow. In addition to the biological information presented here, this study highlights several important areas for the improvement of methodologies that may currently limit the inclusion of wind direction in landscape genetic analyses.

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1186/s13071-020-04522-3.

Additional file 1: Table S1. Sampling sites and associated genetic diversity. *n* Sample size, *Ho* observed heterozygosity, *Hs* expected heterozygosity, *FIS* fixation index.

Additional file 2: Table S2. Reference sequences used for specific assignation.

Additional file 3: Table S3. Origin and numbers of individuals used to build up the DNA library necessary for the development of the microsatellite markers.

Additional file 4: Table S4. Primers of the 13 microsatellite markers used to genotype *C. obsoletus* populations. *DYE* Fluorochrome, *TFm* half denaturation temperature in degrees, *bp* base pairs.

Additional file 5: Figure S1. Environmental variables tested as potential factors that could impact inter-individual genetic differentiation of *C. obsoletus* in France (raster cell resolution: 0.04 arcmin).

Additional file 6: Figure S2. Map of wind direction averaged from 2000 to 2010. Sampling sites are represented by *black points*. The *color scale* represents the wind direction from 0 to 360 ° from north. The *arrowheads* indicate the exact wind direction of each raster pixel.

Additional file 7: Figure S3. Analytical workflow.

Additional file 8: Table S5. Genetic diversity by locus. *Ho* Observed heterozygosity, *Hs* expected heterozygosity, *FIS* fixation index, *SAD* short alleles dominance.

Additional file 9: Figure S4. Isolation by distance analyses: density plots, Mantel tests and linear regressions performed with each inter-individual genetic distance considered in this study.

Additional file 10: Figure S5. Identification of the optimal number of genetic clusters (*K*) inferred by STRUCTURE using the δ(*K*) and *L'*(*K*) methods.

Additional file 11: Figure S6. Population genetic structure results by clustering analyses performed STRUCTURE. A specific color has been assigned to each inferred genetic cluster.

Additional file 12: Table S6. Results of univariate analyses: determination coefficients (R^2) estimated from univariate regressions between genetic and environmental distances. [C] indicates that the considered environmental raster was treated as a conductance factor for the computation of environmental distances with circuit theory. [R] indicates that the considered environmental raster was treated as a resistance factor for the computation of environmental distances with circuit theory. *Q* Difference between environmental multiple regression on distance matrices (MRDM) R^2 and null raster MRDM R^2. The results for $K = 10$ are not shown as they were non-significant.

Additional file 13: Figure S7. Bar plots of population genetic structure results by clustering analyses performed by STRUCTURE for $K = 2$ and $K = 4$. A specific color has been assigned to each inferred genetic cluster.

Author contributions
AM, KH and CG designed the study. TB contributed to the collection of *Culicoides* samples, and IR and MD contributed to the identification of *Culicoides*. AM, LG, and MW performed the molecular biology manipulations. AM, SD and MJ developed the techniques and workflow for the landscape genetics analysis. AM analyzed the data. KH and CG contributed to the manuscript, the first draft of which was written by AM. All the authors read and approved the final manuscript.

Availability of data and materials
All data generated or analyzed during this study are included in this article and its additional files. The newly generated sequences have been submitted to the GenBank database under accession numbers MT828832-MT828844. All the cytochrome *c* oxidase subunit 1 sequences are available on request.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] ASTRE, Univ Montpellier, Cirad, INRAE, Montpellier, France. [2] Cirad, UMR ASTRE, 34398 Montpellier, France. [3] Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, CP160/12, 50, av. FD Roosevelt, 1050 Bruxelles, Belgium. [4] Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Herestraat 49, 3000 Leuven, Belgium. [5] UMR EPIA, Université Clermont Auvergne, INRAE, VetAgro Sup, 63122 Saint-Genès-Champanelle, France. [6] Cirad, UMR ASTRE, 97170 Petit-Bourg, Guadeloupe, France. [7] Cirad, UMR ASTRE, 10100 Rabat, Morocco. [8] Unité Microbiologie, immunologie et maladies contagieuses, Institut Agronomique et Vétérinaire Hassan II, 10100 Rabat-Instituts, Morocco.

Mignotte *et al. Parasites Vectors*   (2021) 14:93

Page 12 of 14

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*. Aust Vet J. 1995;72:23.
4. Carpenter S, Groschup MH, Garros C, Felippe-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 sero-types 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;284:8.
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. Parasites 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, Szmaragd 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;45:22.
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 Palaearctic *Culicoides* biting midges estimated by mark-release-recapture. Parasites 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, Ceratogonidae) 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*
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. Sanders CJ, Harrup LE, Tugwell LA, Brugman VA, England M, Carpenter S. Quantification of within- and between-farm dispersal of *Culicoides* biting midges using an immunomarking technique. J Appl Ecol. 2017;54:1429–39.
28. 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.
29. 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. Parasites Vectors. 2015;8:660.
30. 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. Parasites Vectors. 2016;9:141.
31. 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.
32. 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.
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. Jones A, Thomson D, Hort M, Devenish B: The UK Met Office's Next-Generation Atmospheric Dispersion Model, NAME III. In: Air pollution modeling and its application XVII: 2007; Boston, MA: Springer US: 580–9.
38. 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.
39. 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.
40. Shirk AJ, Landguth EL, Cushman SA. A comparison of individual-based genetic distance metrics for landscape genetics. Mol Ecol Resour. 2017;17:1308–17.
41. 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.
42. 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.

Mignotte *et al. Parasites Vectors*   (2021) 14:93

Page 13 of 14

43. 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.

44. Mignotte A, Garros C, Gardes L, Balenghien T, Duhayon M, Rako-toarivony 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. Parasites Vectors. 2020;13:265.

45. 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.

46. Garros C, Balenghien T, Carpenter S, Delecolle JC, Meiswinkel R, Pedarrieu 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*. Parasites Vectors. 2014;7:223.

47. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

48. Morillo E, Buitron J, Limongi R, Vignes H, Argout X. Characterization of microsatellites identified by next-generation sequencing in the Neotropical tree *Handroanthus billbergii* (Bignoniaceae). Appl Plant Sci. 2016;4:56.

49. Simpson J, Wong K, Jackman S, Schein J, Jones S, Birol I. ABySS: a parallel assembler for short read sequence data. Genome Res. 2009;19:1117–23.

50. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33:2583–5.

51. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3–new capabilities and interfaces. Nucleic Acids Res. 2012;40:e115.

52. Ginestet C. ggplot2: elegant graphics for data analysis. J R Stat Soc Ser A (Stat Soc). 2011;174:245-6.

53. Kamvar ZN, Tabima JF, Grunwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual repro-duction. PeerJ. 2014.

54. Winter DJ. MMOD: an R library for the calculation of population dif-ferentiation statistics. Mol Ecol Resour. 2012;12:1158–60.

55. Wattier R, Engel CR, Saumitou-Laprade P, Valero M. Short allele domi-nance 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.

56. Manangwa O, De Meeus T, Grebaut P, Segard A, Byamungu M, Ravel S. Detecting Wahlund effects together with amplification problems: cryp-tic species, null alleles and short allele dominance in *Glossina pallidipes* populations from Tanzania. Mol Ecol Resour. 2019;19:757–72.

57. Adamack AT, Gruber B, Dray S. PopGenReport: simplifying basic popula-tion genetic analyses in R. Methods Ecol Evol. 2014;5:384–7.

58. Rice WR. Analysing tables of statistic tests. Evolution. 1989;43(1):223–5.

59. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of popu-lation structure. Evolution. 1984;38:1358–70.

60. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59.

61. Guillot G, Mortier F, Estoup A. GENELAND: a computer package for landscape genetics. Mol Ecol Notes. 2005;5:712–5.

62. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak popula-tion structure with the assistance of sample group information. Mol Ecol Resour. 2009;9:1322–32.

63. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14(8):2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

64. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2011;4:359–61.

65. Belkhir K. Genetix 4.05, logiciel sous Windows TM pour la genetique des populations. 2004. http://www.genetix.univ-montp2.fr/genetix/genetix.htm.

66. Goslee SC, Urban DL. The ecodist package for dissimilarity-based analy-sis of ecological data. J Stat Softw. 2007;22:1–19.

67. Kimmig SE, Beninde J, Brandt M, Schleimer A, Kramer-Schadt S, Hofer H, et al. Beyond the landscape: resistance modelling infers physical and

68. Rousset F. Genetic differentiation between individuals. J Evol Biol. 2000;13:58–62.

69. Hardy OJ, 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.

70. 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.

71. 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.

72. Nychka D, Furrer R, Paige J, Sain S. Fields: tools for spatial data. 2017.

73. Mantel N. The detection of disease clustering and a generalized regres-sion approach. Cancer Res. 1967;27:209–20.

74. 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.

75. McRae B, Shah VB. Circuitscape: a tool for landscape ecology. 2008.

76. McRae BH. Isolation by resistance. Evolution. 2006;60:1551–61.

77. Wint W, Searle K, Morley D, Alexander N, Medlock J. A first attempt at modelling roe deer (*Capreolus capreolus)* distributions over Europe. Open Public Health J. 2014;2:e1.

78. 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.

79. 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. In: International Sympo-sium on Lidar and Radar Mapping 2011: Technologies and Applications. 2011;8286.

80. Wan ZM. New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. Remote Sens Environ. 2014;140:36–45.

81. Fernández-López J, Schliep K. rWind: download, edit and include wind data in ecological and evolutionary analysis. Ecography. 2018;42:804–10.

82. Newton RG, Spurrell DJ. Examples of the use of elements for clarifying regression analyses. Appl Stat. 1967;16:165–72.

83. Nimon K, Lewis M, Kane R, Haynes RM. An R package to compute commonality coefficients in the multiple regression case: an introduc-tion to the package and a practical example. Behav Res Methods. 2008;40:457–66.

84. Dellicour S, Prunier JG, Piry S, Eloy MC, Bertouille S, Licoppe A, et al. Landscape genetic analyses of *Cervus elaphus* and *Sus scrofa*: compara-tive study and analytical developments. Heredity. 2019;123:228-41.

85. Prunier JG, Colyn M, Legendre X, Flamand MC. Regression commonality analyses on hierarchical genetic distances. Ecography. 2017;40:1412–25.

86. Hijmans RJ. geosphere: spherical trigonometry. 2019.

87. 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.

88. 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.

89. 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. Can J Med Entomol. 2020;57:1324–7.

90. 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.

91. Duchesne P, Turgeon J. FLOCK provides reliable solutions to the "num-ber of populations" problem. J Hered. 2012;103:734–43.

92. 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

Mignotte *et al. Parasites Vectors*   (2021) 14:93

Page 14 of 14

connectivity using an individual-based sampling scheme. Mol Ecol. 2013;22:5516–30.

93. 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.

94. 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.

95. 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.

96. 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.

97. 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.

98. 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.

99. 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. Parasites Vectors. 2014;7:34.

100. 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.

101. 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.

102. 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.

103. 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.

104. 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 Sci. 2014;33:338–48.

105. 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.

106. 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.

Publisher's Note