MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 2456–2466 doi: 10.1111/mec.12264

Climate change and the spread of vector-borne diseases: using approximate Bayesian computation to compare invasion scenarios for the bluetongue virus vector *Culicoides imicola* in Italy

PATRICK MARDULYN,*† MARIA GOFFREDO,‡ ANNAMARIA CONTE,‡ GUY HENDRICKX,§ RUDOLF MEISWINKEL,¶ THOMAS BALENGHIEN,** SOUFIEN SGHAIER,†† YOUSSEF LOHR‡‡ and MARIUS GILBERT§§†

**Evolutionary Biology and Ecology, Université Libre de Bruxelles, av FD Roosevelt 50, 1050 Brussels, Belgium, †Fonds National de la Recherche Scientifique, rue d'Egmont 5, B-1000 Brussels, Belgium, ‡Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise 'G. Caporale', Via Campo Boario, 64100 Teramo, Italy, §Avia-GIS, Risschotlei 33, 2980, Zoersel, Belgium, ¶Central Institute for Animal Disease Control Lelystad, PO Box 2004, NL-8203 AA Lelystad, The Netherlands, **CIRAD, UMR Contrôle des maladies, F-34398, Montpellier, France, ††Institut de la Recherche Vétérinaire de Tunisie, 20 Rue Djebel Lakhdar La Rabta, 1006, Tunis, Tunisia, ‡‡Laboratoire National d'Epidémiologie et Zoonoses (LNEZ), Ministère de l'Agriculture, du Développement Rural et des Pêches Maritimes du Maroc, IAV Hassan II, Rabat, Morocco, §§Biological control and spatial ecology, Université Libre de Bruxelles, av FD Roosevelt 50, 1050 Brussels, Belgium*

Abstract

Bluetongue (BT) is a commonly cited example of a disease with a distribution believed to have recently expanded in response to global warming. The BT virus is transmitted to ruminants by biting midges of the genus *Culicoides*, and it has been hypothesized that the emergence of BT in Mediterranean Europe during the last two decades is a consequence of the recent colonization of the region by *Culicoides imicola* and linked to climate change. To better understand the mechanism responsible for the northward spread of BT, we tested the hypothesis of a recent colonization of Italy by *C. imicola*, by obtaining samples from more than 60 localities across Italy, Corsica, Southern France, and Northern Africa (the hypothesized source point for the recent invasion of *C. imicola*), and by genotyping them with 10 newly identified microsatellite loci. The patterns of genetic variation within and among the sampled populations were characterized and used in a rigorous approximate Bayesian computation framework to compare three competing historical hypotheses related to the arrival and establishment of *C. imicola* in Italy. The hypothesis of an ancient presence of the insect vector was strongly favoured by this analysis, with an associated $P \geq 99\%$, suggesting that causes other than the northward range expansion of *C. imicola* may have supported the emergence of BT in southern Europe. Overall, this study illustrates the potential of molecular genetic markers for exploring the assumed link between climate change and the spread of diseases.

Keywords: insects, invasive species, population dynamics, population genetics—empirical

Received 27 June 2012; revision received 15 January 2013; accepted 19 January 2013

Introduction

Evidence is accumulating that increases in average temperatures are influencing the large-scale distribution patterns of many species (e.g. Parmesan *et al.* 1999; Parmesan 2006). It has been suggested that this could cause the invasion of pests and infectious diseases, in particular vector-borne diseases, in new territories by the shift of their geographic range towards the poles (Dukes & Mooney 1999; Harvell *et al.* 2002; Lafferty 2009). A commonly cited example of a northward expansion

Correspondence: Patrick Mardulyn, Fax: +32 2 6502445;
E-mail: pmarduly@ulb.ac.be

believed to be associated with climate change is the bluetongue virus that has recently increased its range in the Mediterranean Basin (Purse *et al.* 2005). This virus, responsible for the bluetongue (BT) disease in livestock, is widely distributed around the globe, occurring in Africa, Asia, Australia and the Americas, and has been reported from Mediterranean Europe since at least the 1920s (Gambles 1949). In this region, the virus is primarily transmitted by the small biting midge *Culicoides imicola*, whose females feed on livestock mostly at night and often in vast numbers. The last major epidemic in Europe, which commenced in Greece in 1998, invaded areas that had never been infected previously, including Tunisia, Italy, Corsica, the Spanish islands of Menorca and Mallorca, and Bulgaria (Mellor & Wittmann 2002). Following this epidemic, *C. imicola* was sampled and found at latitudes that were more northern than any previous historical records. This gave rise to the suggestion that the increased incursions of BT, especially within the Mediterranean region, could be due to the concomitant changes in the latitudinal range of the principal insect vector *C. imicola*, which have been linked to climate change (Mellor *et al.* 2008).

Because of the strong change in sampling intensity that usually followed bluetongue epidemic, it has in fact been quite difficult to quantify this northward shift. For example, the available data on *Culicoides* sampling carried out in Italy prior to 2000 have been evaluated in terms of trapping conditions (i.e. collection sites, type of traps, period of the year), and this showed that previous sampling would have had very little chance to catch *C. imicola* even today (Goffredo *et al.* 2003). Furthermore, a retrospective study closely examined time series of entomological surveillance data of *C. imicola* across Italy from 2001 to 2007 and showed that populations were stable in space and time (Conte *et al.* 2009). The study failed to find evidence of an ongoing range expansion during that period, which would have been expected if the northward shift of BTV had been caused by the northward range expansion of its vector.

The analysis of molecular genetic data has the potential to shed some light on past invasion histories of populations and can therefore be used to better characterize the causal link between climate change and disease invasions. A recent molecular phylogeographic study of *C. imicola* based on mitochondrial cytochrome oxidase I sequences already confirmed that the North African populations, in contrast to populations from Greece, Turkey and the Middle East, are genetically close to the western European populations (Italy, Spain, Corsica), thereby supporting the view that North African populations provided the source for the colonization of western Europe (Nolan *et al.* 2008). However, the study provided no information as to when this colonization could have occurred. In a more localized study, Calvo *et al.* (2009) analysed the CO1 haplotype structure of *C. imicola* from 58 localities in Spain and concluded that the lack of matrilineal subdivision, and a star-like phylogeny centred around the most frequent haplotype, was best explained by the rapid expansion of *C. imicola* in Spain. However, (i) historical demographic events cannot be confidently inferred from a single DNA fragment, and (ii), more importantly, the short time (a few decades) of the hypothesized invasion is unlikely to be sufficient for the generation of the observed haplotype network. Molecular genetic evidence for the recent colonization of *C. imicola* in Spain was therefore still lacking.

In an attempt to investigate the possible recent colonization of southern Europe by *C. imicola* and hence to better understand the mechanism linking the recent northward spread of BT in Europe to climate change, we examined 10 polymorphic microsatellite loci in specimens obtained from over 60 localities sampled in Italy (including the mainland and the islands of Sardinia and Sicily), Corsica, southern France and North Africa (Tunisia, Algeria and Morocco). These molecular markers were first used to establish the extent of the genetic variation within and among 22 populations of *C. imicola* and to then evaluate two contrastive hypotheses relating to its presence in the west-central part of the Mediterranean Basin: (i) C. *imicola* invaded Italy only recently from out of North Africa (about 30 years ago), and (ii) *C. imicola* has long been established in Italy and has regularly been exchanging migrants with populations from North Africa.

To test these hypotheses, three coalescence-based population evolution models were developed, and the probabilities of each having generated the observed molecular patterns were estimated using the recently developed approximate Bayesian computation (ABC) approach (Beaumont *et al.* 2002; Bertorelle *et al.* 2010). As a Bayesian method, ABC allows for the incorporation of prior biological and historical knowledge, which helps to reduce the size of the parameters-space explored (Beaumont & Rannala 2004). Instead of calculating the full likelihood of a given hypothesis, which can quickly become intractable with moderately complex models, model probabilities are approximated using computer simulations of population evolution. In practice, a large number of simulated data sets are generated and then compared with the observed data through the calculation of several summary statistics. The principal advantage of this approach is the capacity to handle large data sets and relatively complex models of population evolution. Also, just like traditional full-likelihood methods, it is able to employ Markov chain

Monte Carlo (MCMC) methods to increase the efficiency of the parameters-space exploration (Marjoram *et al.* 2003; Ratmann *et al.* 2007; Wegmann & Excoffier 2010; Slater *et al.* 2012). ABC has been used successfully in a number of evolutionary scenarios studies, including human evolution (Fagundes *et al.* 2007), bullhead (*Cottus gobio*) colonization (Neuenschwander *et al.* 2008) and pest beetle (*Diabrotica virgifera*) invasion (Guillemaud et al. 2010).

Methods

Insect sampling

Samples were obtained from four Italian regions where *Culicoides imicola* is well established: Sardinia, Lazio/Tuscany, Calabria and Sicily. Insects were collected using blacklight traps, according to the protocols of the Entomological Surveillance Plan lead by the National Reference Center for Exotic Disease (CESME) (Goffredo & Meiswinkel 2004). In addition, samples were collected in neighbouring regions: Corsica, Var (France), Tunisia, Algeria and Morocco. Figure 1 shows an overview of all sampled localities, and Table S1 (Supporting information) provides more detailed geographic information for all individuals collected.

Development of microsatellite markers

A genomic DNA library was constructed and then enriched for microsatellite loci following a protocol similar to the one described by Glenn & Schable (2005). A brief summary of the protocol used is given in Appendix S1 (Supporting information). Among all identified loci, 24 were tested for polymorphism by genotyping several individuals from different populations. In the end, 10 loci with enough variability and unambiguous amplification patterns were selected for genotyping the collected samples. Primers, annealing temperatures and length of amplified products for these 10 microsatellite loci are given in Table S2 (Supporting information). Selecting the most polymorphic loci from a set of isolated microsatellite sequences can result in introducing a bias in subsequent data analysis, referred to as ‘ascertainment bias’ (e.g. Vowles & Amos 2006; Väli *et al.* 2008). We believe that this potential bias had no impact on the specific purpose of this study, that is, the comparison of three alternative demographic hypotheses. Indeed, while it could result in overestimating genetic diversity within populations, the bias should be equal across populations because individuals from different populations were used for the initial microsatellite screening, and the level of structure among populations detected in this study was extremely weak. Therefore, if an ascertainment bias was introduced by our procedure, it should not have influenced (i) the relative differences in genetic diversity estimated among populations, or (ii) pairwise population differentiation, the parameters used to discriminate among historical scenarios in the ABC analysis (see below).

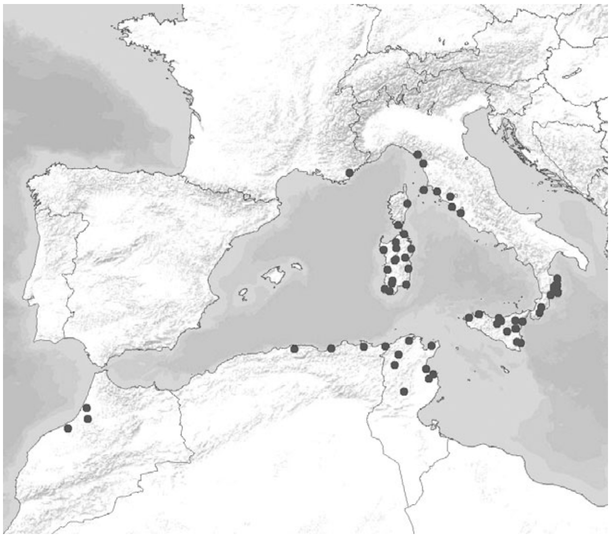

Fig. 1 Map showing the geographic locations of the samples analysed in this study (see Table S1, Supporting information for details).

Genotyping of sampled individuals

A quick and easy method was used to extract DNA from single midges. Each individual was placed in a 0.2-mL microtube containing 10 μL of a lysis buffer (Tris–HCl 10 mM pH 8.2, KCl 50 mM, MgCl2 2.5 mM, Tween-20 0.45%, gelatin 0.01%, proteinase K 60 μg/mL) and squashed using a 10-μL pipette tip. Each microtube was then incubated at −80°C for 30 min, at 65°C for 1 h and finally at 95°C for 15 min. Microsatellite loci were amplified for each individual by multiplex PCR with the Multiplex PCR Kit (QIAGEN), following the protocol described in the manufacturer’s manual and the annealing temperatures given in Table S2. Amplified products were separated by electrophoresis on an Applied Biosystems 3730 automated sequencer. Allele call was conducted with the software Peak Scanner version 1.0 (Applied Biosystems), and genotypes were manually entered in an Excel sheet to create a data set with the program Microsatellite Analyser (MSA; Dieringer & Schlötterer 2003) that automatically generates input files for several population genetics computer packages.

Data analysis

Before assessing population structure, genotypes were grouped a priori by geographic regions in which they were sampled: Sardinia (Italy), Corsica (France), Lazio/Tuscany (Italy), Sicily (Italy), Calabria (Italy), Tunisia, Algeria, Morocco and Var (France). Hardy–Weinberg equilibrium for each locus and each region, but also across regions (Fisher's method), and linkage disequilibrium between each pair of loci and across regions (Fisher's method) were tested for using GenePop version 4 (Rousset 2008). Intra-population genetic variation was estimated by calculating the number of alleles and the expected heterozygosity for each region sample with Arlequin version 3.5.1.2 (Excoffier & Lischer 2010). Levels of population differentiation among the a priori-defined regions were estimated by calculating the fixation index F_{ST} with GenePop version 4 and by calculating Jost's D (Jost 2008) using SMOGD version 1.2.5 (Crawford 2010). Pairwise population differentiation was tested using the exact *G* test option in GenePop.

To infer population structure from the microsatellite data, we used the program Structure version 2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2003) that implements a Bayesian clustering approach that tries to assign individuals to a predefined set of populations according to their genotypes, while simultaneously estimating population allele frequencies. We performed several structure analyses on the data set (five runs for each value of K ranging from 1 to 10), choosing the admixture model (estimating for each individual the proportion of its genome belonging to each of the K populations), the correlated frequencies model (assuming the allele frequencies are similar among populations due to migration or shared ancestry) and the Locprior model (using sampling locations as prior information to assist the clustering), which is recommended for data sets with a weak signal of structure (Hubisz *et al.* 2009). In each run, a Markov Chain Monte Carlo (MCMC) chain of 10 million steps was launched, after a burn-in of 1 million steps. For these analyses of population structure, each sampled locality was considered to correspond to a separate population, except for the Italian genotypes that were grouped in the four above-defined regions. This was justified by two reasons: (i) these regions were clearly identified in a national surveillance programme as separate geographic entities where *C. imicola* is abundant in Italy and separated by large areas in which the species is virtually absent (see Fig. 2 in Conte *et al.* 2009), and (ii) most individuals sampled from these regions belonged each to a separate location in an attempt to maximize sampling homogeneity. In addition, because the model used by Structure assumes Hardy–Weinberg equilibrium for all loci, we performed the analyses both for the complete data set, but also for a second data set in which four loci (8b1, 41b, 68, 88) for which Hardy–Weinberg equilibrium was significantly rejected across samples were removed. Levels of population differentiation (F_{ST} and D) were re-estimated, this time among five populations defined a posteriori following the results of the Structure analyses (see Results). Moreover, we estimated population size and pairwise migration rates for these five populations using Migrate version 3.1.6 (Beerli & Felsenstein 2001; Beerli 2009), keeping in mind that these estimates are valid only under a stationary model, thereby assuming that population sizes and migration rates have remained constant for a long period of time (i.e. going backward in time, from today to the time of the most recent common ancestor of the gene copies of each locus analysed). Three maximum likelihood MCMC runs were launched to allow for testing convergence, each with the following settings: infinite allele mutation model (a preliminary test using the more realistic stepwise mutation model resulted in an estimated run time of more than 6 months), start parameters: theta and M values generated from F_{ST} calculation, mutation rate assumed constant for all loci, MCMC settings: 10 short chains (2000 recorded steps; increment 100; 200 000 visited genealogies, burn-in 10 000), 3 long chains (20 000 recorded steps; increment 100; 2000 000 visited genealogies, burn-in 10 000), Multiple Markov chains on long chains only, static heating scheme: four chains with temperatures 10, 7, 4, 1 and swapping interval of 1.

To confront two competing hypotheses regarding the presence of *C. imicola* in Italy to the observed pattern of genetic variation, we used an ABC approach with the program ABCtoolbox (Wegmann *et al.* 2010). A population evolution model was developed for three tested scenarios, for the purpose of generating molecular data sets similar to the one obtained by this study, via computer simulations. These models are illustrated in Fig. 2: (i) Italy has been colonized 30 years ago by *C. imicola* individuals from North Africa, through one unique colonization event, with no recurrent migration occurring since then, (ii) Italy has been colonized 30 years ago by *C. imicola* individuals from North Africa, with recurrent migration occurring between both regions since then, and (iii) recurrent migration of *C. imicola* between North Africa and Italy has occurred for a relatively long time (compared with the short period of time during which BTV has been present in Italy). A detailed description of these models (including prior distributions of model parameters) is available in Appendix S1 (Supporting information). For each model, three Markov Chain Monte Carlo (MCMC) runs were launched (100 000 simulations, including 10 000 simulations for the initial calibration of the chain; rangeProp

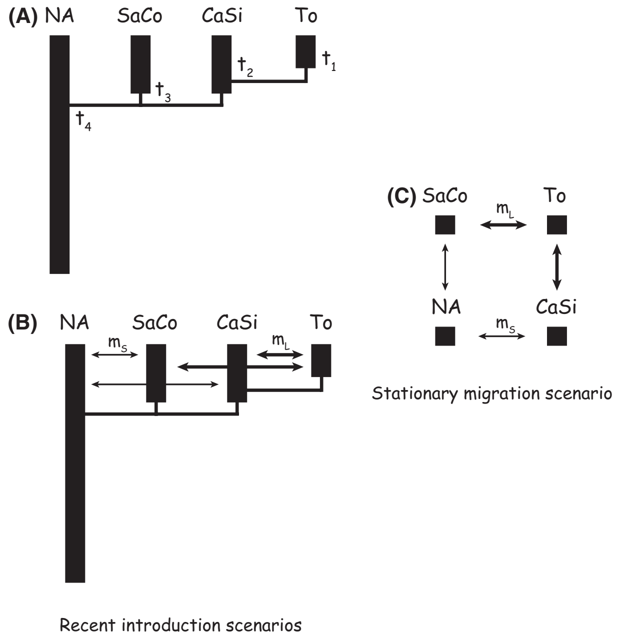

Fig. 2 Three historical scenarios modelled in the Approximate Bayesian Computation (ABC) analyses. (A) Recent introduction without recurrent migration. At time t_4, a small portion of the North African (NA) population colonizes the SaCo (Sardinia + Corsica) and CaSi (Calabria + Sicily) regions. At time t_3, the size of the populations in these two regions is expanded to its contemporary size. At time t_2, a small portion of the CaSi population colonizes the to region (Tuscany). At time t_1, the size of the population in this region is expanded to its contemporary size. Aside from the colonization events, no recurrent migration among regions is implemented. (B) Recent introduction with recurrent migration. Same model as in A except that once a region is colonized, it exchanges migrants with other existing adjacent regions. Two migration rates are defined: a small migration rate, $\mathrm{m_s}$, characterizing gene flow between regions separated by large areas of sea water, and a larger migration rate, m_L, characterizing gene flow between regions separated by shorter areas of sea water. (C) Stationary migration. The number of populations and their size are kept constant during the entire coalescence simulation, and migration occurs among adjacent populations. Migration rates are as defined in B. Note that the range of parameter values used for the ABC simulations were identical across all three scenarios, when applicable.

parameter set to 0.5; sampling interval set to 1; tolerance level set to 0.02) using the ABCsampler program of the ABCtoolbox package in conjunction with the simulation program Simcoal 2 (Laval & Excoffier 2004) and the program Arlsumstat (Excoffier & Lischer 2010) for the computation of summary statistics. The tolerance level was chosen during an initial phase of preliminary runs, so that the acceptance rate of the MCMC chain was reasonably high (>0.20; Wegmann *et al.* 2009). Convergence of the MCMC chain was assessed by comparing the results of the three independent runs. For each simulation, new model parameter values where randomly sampled from specified prior distributions, described in Appendix S1 (Supporting information). Summary statistics characterizing genetic diversity within populations and genetic differentiation among populations were assumed to be the most appropriate to allow discriminating among the compared hypotheses. In all ABC runs, we used as summary statistics: (i) the mean heterozygosity over loci for each population, (ii) the standard deviation over loci of the heterozygosity for each population, but also over populations, and (iii) all pairwise F_{ST} among populations. At the end of each MCMC run, a postsampling regression adjustment (ABC-GLM; Leuenberger & Wegmann 2010) on the retained parameter values was performed with the program ABC estimator (ABCtoolbox package). Models were compared by calculating the Bayes factor as the ratio of marginal densities estimated by ABC estimator for each model (Wegmann *et al.* 2010).

When it became clear that the model depicted in Fig. 2C was much more probable than models 2A and 2B under the conditions described in Appendix S1, we launched additional ABC runs after modifying parameters of models 2A and 2B, attempting to increase their probability. The aim was to determine under what conditions, even if unrealistic, the probabilities of these two models would approach that of model 2C. This was performed by decreasing the intensity of the founder events in models 2A and 2B. More specifically, we decreased the number of generations during which the newly founded populations have a small effective size (from 50 generations to 20, 10 and 5), and we increased the proportion of founding individuals from the population of origin (setting the prior distribution of this proportion as a log-uniform distribution ranging from 0.0001 to 0.01).

Finally, we performed a last comparison of the models, using the parameter ranges of the initial MCMC runs, but this time using a standard rejection procedure as in the study Estoup *et al.* (2004) or Ray *et al.* (2010). For each model, we ran one million simulations and retained the best 5000 (i.e. those with the smallest Euclidean distance between the simulated and observed summary statistics). The Euclidean distances from the 15 000 retained simulations were recomputed from standardized summary statistics (option 'standardizeStats' in ABCtoolbox), and the simulations were sorted by ascending distances. The posterior probability of each model was then taken as the proportion of the first 1000 simulations done under that model. Because model choice under an ABC framework can be biased (Leuenberger & Wegmann 2010; Robert *et al.* 2011), we empirically attempted to evaluate the performances of the specific ABC analysis implemented here. For this purpose, we estimated the rate at which our favoured hypothesis (model C, see results) was chosen when the data were generated under the two alternative hypotheses, model A or B. This was performed by simulating 100 data sets under models A and B and by analysing the resulting 200 pseudo-observed data sets in exactly the same way as the observed data (e.g. Wegmann & Excoffier 2010). In the case of the standard rejection procedure, this was performed relatively easily because the same 10^6 data sets simulated under each model that were used to analyse the observed data could be reused to analyse the pseudo-observed data sets (contrary to the MCMC approach, for which a new MCMC run would have had to be launched for each pseudo-observed data set).

Results and discussion

Genetic variation and population structure

Patterns of genetic variation across population often give clues about the evolutionary processes that have contributed to create them. Here, we used microsatellite genotypic data to characterize patterns of genetic diversity within and among the sampled populations of *Culicoides imicola*. Gene diversity within populations, as measured through expected heterozygosity (Table 1), appeared similar across all a priori-defined populations, with only Algeria and Morocco populations characterized by slightly higher values. Results of the Hardy–Weinberg equilibrium and linkage disequilibrium tests are provided in Appendix S2 (Supporting information). Population differentiation, as measured by pairwise F_{ST} values (Table 2), was overall extremely low, even between populations that are separated by relatively long distances, or by large areas of sea water (e.g., 0.015 between Morocco and Tuscany populations, or 0.008 between corsican and algerian populations). It has been argued that for genetic markers characterized by a high mutation rate, such as microsatellites, jost's D (Jost 2008) is a better statistic to measure population differentiation. The estimated values of this statistic for pairwise population comparisons (Table 2) were also low, confirming low levels of population differentiation.

Table 1 Intra-population genetic diversity

Population	n	A_e	H_e
Sardinia	48	5.6 ± 2.1	0.56 ± 0.13
Tuscany	40	5.6 ± 2.4	0.58 ± 0.14
Calabria	41	4.4 ± 1.5	0.56 ± 0.21
Sicily	39	4.8 ± 1.8	0.55 ± 0.16
Corsica	19	3.9 ± 1.6	0.57 ± 0.10
France	10	3.3 ± 1.0	0.57 ± 0.13
Tunisia	53	5.1 ± 2.0	0.56 ± 0.19
Algeria	35	4.9 ± 1.5	0.63 ± 0.09
Morocco	31	5.6 ± 1.8	0.62 ± 0.14

n, sample size (number of individuals); A_e, mean number of allele per locus; H_e, expected heterozygosity.

This weak population structure highlighted over such a large area suggests that *C. imicola* is characterized by a strong ability to disperse, and one could even argue that all individuals found in North Africa and Italy should be viewed as a single population. This was further tested using a Bayesian clustering approach, based on the assignment of genotypes to a predefined number of populations, but using sampling locations as prior information to assist the clustering. This analysis unambiguously identified the most probable number of clusters as being equal to 3 ($P > 0.99$ for both data sets, one with all 10 loci, the second with only the 6 loci in Hardy–Weinberg equilibrium), rejecting the hypothesis of a single panmictic population (detailed results of the Structure analyses shown in appendix S3, Supporting information).

Based on the probabilities that each of the 22 analysed populations belongs to each of the three clusters (Appendix S3, Supporting information), but also taking geographic locations into account (i.e. clustering together only populations that form geographically coherent groups), we defined 5 clusters of populations for the comparison analysis of alternative historical hypotheses related to the presence of *C. imicola* in Italy: (i) North Africa including all 15 populations from this area (NA), (ii) a cluster with genotypes from Calabria and Sicily (CaSi), (iii) a cluster with Corsica and Sardinia (SaCo),

Table 2 Mean pairwise F_{ST} across loci (below diagonal) and Jost's D (harmonic mean across loci; above diagonal)

	Sardinia	Tuscany	Calabria	Sicily	Corsica	France	Tunisia	Algeria	Morocco
Sardinia	—	0.004	0.027	0.007	0.000	0.046	0.012	0.013	0.013
Tuscany	0.014*	—	0.006	0.000	0.001	0.029	0.011	0.001	0.005
Calabria	0.043*	0.014*	—	0.001	0.019	0.042	0.029	0.022	0.022
Sicily	0.017*	−0.003	0.006*	—	0.004	0.049	0.011	0.011	0.007
Corsica	0.006*	0.010*	0.040*	0.019*	—	0.031	0.015	0.001	0.006
France	0.092*	0.073*	0.101*	0.099*	0.069*	—	0.024	0.015	0.008
Tunisia	0.029*	0.030*	0.054*	0.036*	0.053*	0.064*	—	0.009	0.003
Algeria	0.023*	0.008*	0.034*	0.016*	0.008*	0.030*	0.025*	—	0.000
Morocco	0.025*	0.015*	0.035*	0.020*	0.020*	0.019*	0.015*	−0.001*	—

*Population pairs characterized by a P-value < 0.05 when testing for differentiation (exact G test).

Table 3 Mean pairwise F_{ST} across loci calculated for the redefined populations

	SaCo	To	CaSi	NA	Va
SaCo	—				
To	0.011*	—			
CaSi	0.028*	0.004*	—		
NA	0.021*	0.014*	0.028*	—	
Va	0.084*	0.074*	0.100*	0.036*	—

*Population pairs characterized by a P-value < 0.05 when testing for differentiation (exact G test).

(iv) Tuscany (To) and (v) the French population from the Var department (Va). Table 3 shows the low F_{ST} values recalculated for these new clusters, still suggesting that migration is strong among them.

Hypotheses testing

Comparing the genetic differentiation observed between the Italian populations of *C. imicola* and its most adjacent neighbours gives indications on the evolutionary history of the species. A previous study based on mitochondrial DNA sequences (Nolan *et al.* 2008) showed that Mediterranean populations of *C. imicola* located east of Italy belonged to another lineage than those sampled in South-West Europe and North Africa, and this was interpreted as suggesting a long history of separation between these two geographic entities. In contrast, the highlighted low genetic differentiation among populations analysed in this study rather suggests that these populations are somehow connected, either because they share a recent common origin or because recurrent gene flow occur among them. The continental French population in our sample, the most adjacent population located to the west of Italy, is more genetically distant from the Italian populations than are the populations from North Africa, as judged by pairwise F_{ST} in Table 3. The North African populations of *C. imicola* therefore appear to be the genetically closest populations to the Italian populations, which supports a strong connection between them. This does not mean that a recent colonization of Italy from North Africa following global warming is the only possible scenario to account for the contemporary pattern of genetic variation. In fact, such a recent colonization of Italy (maximum 30 years ago) would have likely involved a modification of allelic frequencies in the newly established population through the strong genetic drift expected to be generated by the associated founder event, and at least some level of differentiation would be expected between the population of origin and the newly founded population. In contrast, if *C. imicola* had been present in Italy for a long time, exchanging recurrent migrants with North Africa, low differentiation between these two regions could be expected.

Using our genotype data, we formally compared these hypotheses in an ABC framework, by developing coalescence-based models of population evolution for three distinct scenarios (Fig. 2). These analyses unambiguously favoured the hypothesis of an ancient presence of *C. imicola* in Italy, with extremely large Bayes factors in favour of the corresponding model ($B_{31} \approx 10^{19}$, $B_{32} \approx 10^{10}$). The ABC analysis conducted using a standard rejection procedure gave similar results: the estimated probability of the third hypothesis was 0.986, compared with 0.009 and 0.005 for the second and first hypotheses, respectively. The standard rejection procedure was further validated by the analyses of pseudo-observed data sets generated under scenarios 1 and 2: the analyses of the pseudo-observed data sets generated under scenario 1 marginally favoured scenario 3 in only 2 cases out of 100 (with an associated estimated probability of only 0.509 and 0.425), and the analyses of the pseudo-observed data sets generated under scenario 2 never favoured scenario 3. With both types of pseudo-observed data sets, scenario 3 was significantly rejected (probability of this scenario < 0.05) in 90% of the cases, showing the potential of this ABC procedure to discriminate model 3 from models 1 and 2. Therefore, our model implementation of the recent colonization of Italy by *C. imicola* is much less likely to have generated the observed pattern of genetic variation than an alternative scenario in which *C. imicola* has been present for a long time in Italy, continuously exchanging migrants with North African populations. Our interpretation of this result is that the intensity of the founder effect imposed by our model of recent colonization of Italy is too strong to account for the low genetic differentiation that was observed between the Italian and North African populations in our data.

Although our models were constrained by ranges of parameter values (prior distributions) that appear reasonable in the light of prior knowledge on the studied system (see Material and Methods), it seemed interesting to investigate under which parameter values, if at all possible, the likelihood of models 1 and 2 (as depicted in Fig 2A and Fig 2B, respectively) could become equivalent to that of model 3. With that goal in mind, we tried to decrease the Bayes factor of the favoured hypothesis (model of Fig. 2C) by relaxing some of the prior constraints on the range of parameter values for the two other hypotheses. More specifically, we decreased the intensity of the founder effect associated with the colonization of Italy in two different ways: (i) by progressively decreasing the number of generations during which the bottleneck was imposed on the newly colonized Italian populations and (ii) by progressively decreasing the bottleneck intensity (inversely proportional to the proportion of the source population that is transferred to the newly colonized area), while maintaining the initial duration of that bottleneck. We were able in both cases to lower the Bayes factor in favour of hypothesis 3 over hypothesis 2 down to a value approaching 1. In the first case, it required that the time from the beginning of the colonization process of each Italian population to the time where it reached its contemporary effective size was set to five generations (±1/2 year). This almost instantaneous colonization would imply an impossibly high rate of reproduction. In the second case, it required that the proportion of the North African population transferred to Italy in a single generation, that is, reflecting the intensity of the founder effect associated with the colonization of Italy, could reach a value as high as 0.01. Because these conditions appear completely unrealistic, we feel it is safe to conclude that the recent colonization of Italy by *C. imicola* is not compatible with the observed data on genetic variation.

Large-scale BTV spread

Our results have implications for our understanding of the mechanism of BTV northward expansion. While a historical and/or contemporary connection between North Africa and Italy was previously highlighted for *C. imicola* using mitochondrial sequences (Nolan *et al.* 2008), our data suggest that the Italian population was established long before the appearance of BTV in Italy. If *C. imicola* has been present in Italy for a long time, its northward spread can no longer be the main cause for the northward spread of BTV and alternative causes, which may also possibly be linked to climate change, need to be considered. This conclusion corroborates results of the study by Conte *et al.* (2009) that showed that *C. imicola* population have not expanded in range in the last 7 years in Italy, and those of Acevedo *et al.* (2010) who suggested that *C. imicola* range in Spain would not be predicted to expand in the future. Rather, it supports the hypothesis that the BTV in Italy spread through populations of indigenous vectors, as was the case with the spread of BTV in northern Europe, where BTV was transmitted by other species of indigenous *Culicoides* vectors. Indeed, in 2006, a bluetongue serotype different from the five that had been circulating within the Mediterranean Basin was introduced adventitiously into north-western Europe at about 51°N near to where the borders of Belgium, Germany and the Netherlands meet (Van Wuijkhuise *et al.*, 2006). From there, it spread rapidly in all directions and over the following 2 years and affected fifteen countries. The advance of the virus to 58°N was unparalleled and for the historical range of bluetongue represented a 1000-km shift northwards. However, as this epidemic involved other vectors and did not involve the northward advance of a southern vector such as *C. imicola*, it became clear that other factors could be involved to explain the northward shift of BT, possibly linked to climate change, but yet not fully understood (Wilson & Mellor 2009; Maclachlan 2011).

To explain the appearance of BT in Italy, it could be hypothesized that climate change has influenced the interaction between the virus, the vector and the environment and hence result in changes of vector capacity and epidemic conditions. Guis *et al.* (2012) have closely examined the effects of climate change on the emergence of BT in Europe, by combining high-resolution climate observations with a mechanistic model of BT transmission risk. Their results indeed suggest that climate explains many aspects of BT's recent emergence and spread and that the drivers of emergence differ somewhat between the south and the north. The vector capacity is strictly linked to the vector abundance and to susceptible hosts availability, alongside to the vector competence. The climate change could have contributed to augment the abundance of local populations of *C. imicola*, and higher temperatures could have facilitated the BTV replication itself. On the other hand, the availability of vertebrate hosts as blood source (i.e. sheep, cattle, horses) could also influence the growing of the local *C. imicola* populations, and the increasing of BT susceptible ruminants could largely promote the BTV spreading. For example, according to FAOSTAT (FAOSTAT 2010, http://faostat.fao.org), the output/input ratio of sheep production in Europe was ranging between 15 and 20 kg of outputs (meat and milk) per animal per year. In the period between 1991 and 2002, this value increased up to a range of 30–35, and a similar pattern of increasing productivity over time is observed in most Mediterranean countries. In Italy, the sheep and goats population increased from 10 millions in 1982 to 12 millions in 1999 (Serie storiche, L'Archivio della statistica italiana. ISTAT - http://seriestoriche.istat.it/), and when BTV reached Sardinia in 2000, it found in the island 3 million sheep and a huge local population of *C. imicola*.

Furthermore, the animal trade and the globalization could also have facilitated the BTV introduction through viraemic cattle in new places, where the local vector population could have then spread the virus and lead to BTV outbreaks.

Finally the intensified grazing over smaller areas of more intensively managed pastures could have contributed to favour more numerous artificial larval habitats for *C. imicola*, hence increasing the local populations of the vector and facilitating the spread of BTV upon introduction.

Overall, this study illustrates the potential of molecular genetic data for exploring the assumed link between climate change and the spread of diseases. More specifically, patterns of genetic variation highlighted by molecular markers can be tested against alternative scenarios of invasion, using rigorous statistical methods, to better characterize the specific history of the invasion. This seems necessary to confirm or infirm presumed associations between biological invasion and climate change, as well as to improve our understanding of the mechanism of this association.

Acknowledgements

Marie-Anne Vaesen provided technical assistance with the development of the microsatellite markers. We are grateful to Thierry Baldet and Jean-Claude Delécolle for providing the Algerian samples and to Bruno Mathieu for the Var samples. The comments of two anonymous reviewers and the subject editor (Roman Biek) greatly improved a previous draft of our manuscript. This project was partially funded by the research programme for earth observation Stereo II of the Belgian Science Policy (research project EPISTIS, contract N° SR/00/102).

References

Acevedo P, Ruiz-Fons F, Estrada R *et al.* (2010) A broad assessment of factors determining *Culicoides imicola* abundance: modelling the present and forecasting its future in climate change scenarios. *PLoS ONE*, **5**, e14236.

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics*, **5**, 251–261.

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Beerli P (2009) How to use migrate or why are Markov chain Monte Carlo programs difficult to use? In: *Population Genetics for Animal Conservation* (eds Bertorelle G, Bruford MW, Hauffe HC, Rizzoli A, Vernesi C), pp. 42–79. Cambridge University Press, Cambridge, UK.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences USA*, **98**, 4563–4568.

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.

Calvo JH, Calvete C, Martinez-Royo A *et al.* (2009) Variations in the mitochondrial cytochrome c oxidase subunit I gene indicate northward expanding populations of Culicoides imicola in Spain. *Bulletin of Entomological Research*, **99**, 583–591.

Conte A, Gilbert M, Goffredo M (2009) Eight years of entomological surveillance in Italy show no evidence of Culicoides imicola geographical range expansion. *Journal of Applied Ecology*, **46**, 1332–1339.

Crawford NG (2010) smogd: software for the measurement of genetic diversity. *Molecular Ecology Resources*, **10**, 556–557.

Dieringer D, Schlötterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.

Dukes JS, Mooney HA (1999) Does global change increase the success of biological invaders? *Trends in Ecology & Evolution*, **4**, 135–139.

Estoup A, Beaumont M, Sennedot F, Moritz C, Cornuet J (2004) Genetic analysis of complex demographic scenarios: spatially expanding populations of the cane toad. *Bufo marinus, Evolution*, **58**, 2021–2036.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Fagundes NJ, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences USA*, **104**, 17614–17619.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Gambles RM (1949) Bluetongue of sheep on Cyprus. *Journal of Comparative Pathology*, **59**, 176–190.

Glenn TC, Schable NA (2005) Isolating microsatellite DNA loci. *Methods in Enzymology*, **395**, 202–222.

Goffredo M, Meiswinkel R (2004) Entomological surveillance of bluetongue in Italy: methods of capture, catch analysis and identification of *Culicoides* biting midges. Proceeding of the

Third International Symposium on Bluetongue, Taormina, Italy, 26-29 October. *Veterinaria Italiana*, **40**, 260–265.

Goffredo M, Conte A, Cocciolito R, Meiswinkel R (2003) Distribution and abundance of *Culicoides imicola* in Italy. *Veterinaria Italiana*, **47**, 22–32.

Guillemaud T, Beaumont MA, Ciosi M, Cornuet JM, Estoup A (2010) Inferring introduction routes of invasive species using approximate Bayesian computation on microsatellite data. *Heredity*, **104**, 88–99.

Guis H, Caminade C, Calvete C, Morse AP, Tran A, Baylis M (2012) Modelling the effects of past and future climate on the risk of bluetongue emergence in Europe. *Journal of the Royal Society Interface*, **9**, 339–350.

Harvell CD, Mitchell CE, Ward JR *et al.* (2002) Climate warming and disease risks for terrestrial and marine biota. *Science*, **296**, 2158–2216.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Jost LOU (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Lafferty KD (2009) The ecology of climate change and infectious diseases. *Ecology*, **90**, 888–900.

Laval G, Excoffier L (2004) SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, **20**, 2485–2487.

Leuenberger C, Wegmann D (2010) Bayesian computation and model selection without likelihoods. *Genetics*, **184**, 243–252.

Maclachlan NJ (2011) Bluetongue: history, global epidemiology, and pathogenesis. *Preventive Veterinary Medicine*, **102**, 107–111.

Marjoram P, Molitor J, Plagnol V, Tavaré S (2003) Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences USA*, **100**, 15324.

Mellor PS, Wittmann EJ (2002) Bluetongue virus in the Mediterranean Basin 1998-2001. *Veterinary Journal (London, England: 1997)*, **164**, 20–37.

Mellor PS, Carpenter S, Harrup L, Baylis M, Mertens PP (2008) Bluetongue in Europe and the Mediterranean Basin: history of occurrence prior to 2006. *Preventive Veterinary Medicine*, **87**, 4–20.

Neuenschwander S, Largiadèr CR, Ray N, Currat M, Vonlanthen P, Excoffier L (2008) Colonization history of the Swiss Rhine basin by the bullhead (Cottus gobio): inference under a Bayesian spatially explicit framework. *Molecular Ecology*, **17**, 757–772.

Nolan DV, Dallas JF, Piertney SB, MordueLuntz AJ (2008) Incursion and range expansion in the bluetongue vector Culicoides imicola in the Mediterranean basin: a phylogeographic analysis. *Medical and Veterinary Entomology*, **22**, 340–351.

Parmesan C (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 637–669.

Parmesan C, Ryrholm N, Stefanescu C *et al.* (1999) Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature*, **399**, 579–583.

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

Purse BV, Mellor PS, Rogers DJ, Samuel AR, Mertens PPC, Baylis M (2005) Climate change and the recent emergence of bluetongue in Europe. *Nature Reviews Microbiology*, **3**, 171–181.

Ratmann O, Jørgensen O, Hinkley T, Stumpf M, Richardson S, Wiuf C (2007) Using likelihood-free inference to compare evolutionary dynamics of the protein networks of *H. pylori* and *P. falciparum*. *PLoS Computational Biology*, **3**, e230.

Ray N, Wegmann D, Fagundes NJR, Wang S, Ruiz-linarez A, Excoffier L (2010) A statistical evaluation of models for the initial settlement of the American continent emphasizes the importance of gene flow with Asia. *Molecular Biology and Evolution*, **27**, 337–345.

Robert CP, Cornuet J-M, Marin J-M, Pillai NS (2011) Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences USA*, **108**, 15112–15117.

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Slater GJ, Harmon LJ, Wegmann D, Joyce P, Revell LJ, Alfaro ME (2012) Fitting models of continuous trait evolution to incompletely sampled comparative data using approximate Bayesian computation. *Evolution*, **66**, 752–762.

Väli Ü, Einarsson A, Waits L, Ellegren H (2008) To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology*, **17**, 3808–3817.

Van Wuijckhuise L, Dercksen D, Muskens J, de Bruijn J, Scheepers M, Vrouenraets R (2006) Bluetongue in the Netherlands; description of the first clinical cases and differential diagnosis. Common symptoms just a little different and in too many herds. *Tijdschrift voor Diergeneeskunde*, **131**, 649–654.

Vowles EJ, Amos W (2006) Quantifying ascertainment bias and species-specific length differences in human and chimpanzee microsatellites using genome sequences. *Molecular Biology and Evolution*, **23**, 598–607.

Wegmann D, Excoffier L (2010) Bayesian inference of the demographic history of chimpanzees. *Molecular Biology and Evolution*, **27**, 1425–1435.

Wegmann D, Leuenberger C, Excoffier L (2009) Efficient approximate Bayesian computation coupled with Markov Chain Monte Carlo without likelihood. *Genetics*, **182**, 1207–1218.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, **11**, 116.

Wilson AJ, Mellor PS (2009) Bluetongue in Europe: past, present and future. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, **364**, 2669–2681.

P.M., M.Go., A.C., G.H., R.M. and M.Gi. designed the study. M.Go., A.C., T.B., S.S. and Y.L. collected most of the samples. P.M. genotyped the samples and analysed the data. All authors contributed to writing the manuscript. P.M. has a general interest in the analysis of DNA sequence variation for studying evolution. His current research focuses mainly on comparing multilocus phylogeographic patterns among different temperate and cold-adapted herbivorous insects. M.Go. studies vector-borne diseases, with a particular interest on *Cu-*

licoides and mosquitoes transmitting arboviruses. A.C. works on the statistical and spatial analysis of disease spread, with a particular interest in vector-borne diseases. G.H. is interested in developing distribution and spread models of *Culicoides* sp. in Europe. R.M. specializes in the taxonomy and biology of vector *Culicoides*. T.B. has a general interest in vector ecology. His current research focuses on *Culicoides* distribution and dynamics and on host/vector contact. M.Gi. works on the spatial epidemiology of several animal diseases and on the development of modelling approaches to predict invasion patterns.

Data accessibility

The microsatellite data analysed in this study are available in Appendix S4 (Supporting information).

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Sampling of *C. imicola.*

Table S2 Primers used for the amplification of the microsatellite loci in C. Imicola.

Appendix S1 Methods: (i) Protocol to isolate microsatellite loci and (ii) description of models used in ABC analyses.

Appendix S2 Results of Hardy-Weinberg and linkage disequilibrium tests.

Appendix S3 Results of the structure analysis.

Appendix S4 Microsatellite data set.