Received Date : 08-Jun-2015

Revised Date : 06-Oct-2015

Accepted Date : 09-Oct-2015

Article type : Original Article

Colonization of the Mediterranean Basin by the vector biting midge species *Culicoides imicola*: an old story

S. JACQUET[1,2,28,29*], C. GARROS[1,2], E. LOMBAERT[3], C. WALTON[4], J. RESTREPO[1,2], X. ALLENE[1,2,§], T. BALDET[1,2,§], C. CETRE-SOSSAH[1,2,5,§], A. CHASKOPOULOU[6,§], J-C DELECOLLE[7,§], A. DESVARS[8,§], M. DJERBAL[9,§], M. FALL[10,§], L. GARDES[1,2,§], M. DE GARINE-WICHATITSKY[11,25,§], M. GOFFREDO[12,§], Y. GOTTLIEB[13,§], A. GUEYE FALL[10,§], M. KASINA[14,§], K. LABUSCHAGNE[15,§], Y. LHOR[16,§], J. LUCIENTES[17,§], T. MARTIN[18,§], B. MATHIEU[7,19,§], M. MIRANDA[20,§], N. PAGES[21,22§], I. PEREIRA DA FONSECA[23,§] , D. W. RAMILO[23,§], A. SEGARD[24,25,§], M-L. SETIER-RIO[19,§], F. STACHURSKI[1,2,§], A. TABBABI[26,§], M. TALLA SECK[10,§], G. VENTER[15,§], M. ZIMBA[27,§], T. BALENGHIEN[1,2], H. GUIS[1,2], C. CHEVILLON[28,29], J. BOUYER[1,2,10], K. HUBER[2,1]

[1] CIRAD UMR15 CMAEE 34398 Montpellier France, [2] INRA UMR1309 CMAEE 34398 Montpellier France, [3] INRA, UMR1355 Institut Sophia Agrobiotech, 06903 Sophia Antipolis, France, [4] Computational and Evolutionary Biology, Faculty of Life Sciences, University of Manchester, Manchester, UK, [5] Plateforme de recherche CYROI, CRVOI, Sainte Clotilde, La Réunion, France, [6] USDA-ARS European Biological Control Laboratory, 54623 Thessaloniki, Greece, [7] Institute of Parasitology and Tropical Pathology (IPPTS), Medicine faculty, EA7292, 67000 Strasbourg, France, [8] Department of Clinical Microbiology, Umea University, Umea, Sweden, [9] Regional Veterinary Laboratory of Draa-Ben-Kheda, Tizi-Ouzou, Algeria, [10] ISRA-LNERV, BP 2057, Dakar, Senegal, [11] Cirad, UPR AGIRs, RP-PCP, Harare, Zimbabwe; Cirad, UPR AGIRs, Montpellier, France, [12] Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise 'G. Caporale', 64100 Teramo, Italy, [13] Koret School of Veterinary Medicine, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, [14] Kenya Agricultural and Livestock Research Organization Sericulture, P.O. Box 7816 code 01000 Thika Kenya, [15] Agricultural Research Council-Onderstepoort Veterinary Institute, Parasites, Vectors and Vector-borne Diseases, Onderstepoort 0110, South Africa, [16] Office National de Sécurité Sanitaire des produits Alimentaires (ONSSA), Rabat, Morocco, [17] Departamento de Patología Animal, Facultad de Veterinaria, Universidad de Zaragoza, Zaragoza, Spain, [18] UR Hortsys, Cirad,

Montpellier, France ; Plant Health Department, ICIPE, Nairobi, Kenya, [19] EID Méditerranée, 34184 Montpellier, France, [20,] Laboratory of Zoology, University of Balearics (UIB), Palma de Mallorca, Spain, [21] Centre de Recerca en Sanitat Animal (CReSA), UAB-IRTA, Campus de la Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Spain, [22] CIRAD, UMR CMAEE, 97170 Petit Bourg, Guadeloupe, France INRA, UMR1309 CMAEE, F-34398 Montpellier, France, [23] CIISA, Faculdade de Medicina Veterinária, ULisboa, 1300-477 Lisboa, Portugal, [24] CNRS, Université de Montpellier, UMR 5175 CEFE, [25] EPHE campus CNRS, 1919 route de Mende, 34293 Montpellier Cedex 5 France, [26] Pasteur Institute of Tunis, Lab of Medical Parasitology, Biotechnologies & Biomolecules (LR 11 IPT 06), [27] University of Zimbabwe, Department of Biological Sciences, Entomology, P.O. Box MP 167, Mount Pleasant, Harare, Zimbabwe, [28] CNRS, Université de Montpellier, UMR 5290 Maladies Infectieuses & Vecteurs-Ecologie, Génétique, Ecologie, Contrôle (MIVEGEC), [29] IRD, UR 224 MIVEGEC, BP 64501, Agropolis, 34394 Montpellier Cedex 5
* Corresponding author

$ by alphabetical order. These authors contributed to the work by collecting and/or identifying *C. imicola* samples, and commenting final manuscript.

Keywords: *Culicoides imicola,* colonization, approximate Bayesian computation, microsatellites, mitochondrial genes

Corresponding author: Stéphanie Jacquet, CIRAD UMR Contrôle des Maladies Animales Exotiques et Emergentes, Campus international de Baillarguet, TA A15/G, 34398 Montpellier cedex 5, France, fax number: (33) 4 67 61 41 40, email : stephanie.jacquet@cirad.fr

Running title: Phylogeography of *Culicoides imicola*

Abstract

Understanding the demographic history and genetic make-up of colonizing species is critical for inferring population sources and colonization routes. This is of main interest for designing accurate control measures in areas newly colonized by vector species of economically

important pathogens. The biting midge *Culicoides imicola* is a major vector of Orbiviruses to livestock. Historically, the distribution of this species was limited to the Afrotropical region. Entomological surveys first revealed the presence of *C. imicola* in the south of the Mediterranean basin by the 1970's. Following recurrent reports of massive bluetongue outbreaks since the 1990s, the presence of the species was confirmed in northern areas. In this study, we addressed the chronology and processes of *C. imicola* colonization in the Mediterranean basin. We characterized the genetic structure of its populations across Mediterranean and African regions using both mitochondrial and nuclear markers, and combined phylogeographical analyses with population genetics and approximate Bayesian computation. We found a west/east genetic differentiation between populations, occurring both within Africa and within the Mediterranean basin. We demonstrated that three of these groups had experienced demographic expansions in the Pleistocene, probably because of climate changes during this period. Finally, we showed that *C. imicola* could have colonized the Mediterranean basin in the late Pleistocene or early Holocene through a single event of introduction; however we cannot exclude the hypothesis involving two routes of colonization. Thus, the recent bluetongue outbreaks are not linked to *C. imicola* colonization event, but rather to biological changes in the vector or the virus.

Introduction

Understanding the history of colonization processes and identifying population sources and geographical pathways are critical. This knowledge would help for the construction of predictive models of future colonization and the implementation of effective biological management measures (Simberloff *et al.* 2013). Historical and observational data for colonizing species are often sparse, incomplete and misleading, so acquiring knowledge on the colonization processes using exclusively direct observations is hazardous. The indirect

approaches of population genetics and phylogeography can nevertheless overcome these limitations. The genetic variability of invading populations depends on the history of their source populations as well as on the historical and demographical features of introduction into new areas (Estoup & Guillemaud 2010). Indeed, during the colonization process, colonizing populations will encounter various and complex demographic events such as population size changes, vicariance events and admixture of differentiated populations that will leave a signature on their genetic composition. The study of genetic variation among and within populations can thus help unravel the evolutionary and demographic history of the studied species (Avise 2000).

An illustrative example in this context is the northward expansion of the African biting midge *Culicoides imicola* Kieffer (Diptera: Ceratopogonidae) into the Mediterranean basin. Historically, *C. imicola* is an Afrotropical species, widespread in sub-Saharan Africa and the Middle East, and occasionally recorded in the Far East (i.e. India (Dyce & Wirth 1983); with putative records in Southern China (Yü 2005). Throughout its distribution, *C. imicola* is a well-known vector of economically important livestock viruses such as bluetongue virus and epizootic hemorrhagic disease virus affecting domestic and wild ruminants as well as transmitting African horse sickness virus to equids (Mellor *et al.* 2000). Recurrent reports of such viruses in Mediterranean areas in the 20th century resulted in the suspicion of *C. imicola's* presence in the area (Mellor *et al.* 2009; Mellor *et al.* 2008; Purse *et al.* 2005). There, the bluetongue virus was first reported in 1924 from Cyprus island which was until 1998 the only European country where it was endemic (Gambles 1949), while it was periodically observed in the southern part of the Iberian Peninsula (Mellor *et al.* 1983; Mellor *et al.* 1985) and several Greek islands (Boorman 1986; Boorman & Wilkinson 1983). Entomological surveys lately confirmed the presence of *C. imicola* in Mediterranean areas. The presence of *C. imicola* in Morocco and Algeria was admitted by the 1970's (Bailly-

Choumara & Kremer 1970). At that time, entomological surveys identified the northern distribution edge of *C. imicola* at the latitude 40°N (Mellor *et al.* 2000; Mellor *et al.* 1983; Mellor *et al.* 1985; Rawlings *et al.* 1998). Meanwhile, *C. imicola* populations were reported in Israel, western Turkey (Anatolia) and on several Greek islands (Lesbos, Rhodes and Chios) (Boorman 1986; Boorman & Wilkinson 1983; Braverman & Galun 1973). By contrast, entomological surveys suggested that *C. imicola* remained absent from mainland Greece until 1999 (at least) (Mellor & Wittmann 2002).

The epidemiology of bluetongue disease dramatically changed between 1998 and 2005 with records of massive outbreaks throughout the Mediterranean basin (Mellor *et al.* 2009; Mellor & Wittmann 2002). This reinforced the entomological surveillance, which in turn confirmed the presence of *C. imicola* in Tunisia (Chaker *et al.* 2005) and virtually all Mediterranean islands along the northern Mediterranean seashore: Portugal (Capela *et al.* 2003), Spain (Catalonia) (Monteys & Saiz-Ardanaz 2003; Monteys *et al.* 2005), the Balearic Islands (Miranda *et al.* 2003), France (Corsica, Var department) (Delecolle & De La Rocque 2002; Venail *et al.* 2012), Italy (Sardinia, Sicily, mainland Italy) (Goffredo *et al.* 2003; Goffredo *et al.* 2004) and mainland Greece (Patakakis 2004). As a result, the consensual hypothesis was a northward expansion of *C. imicola* in late 20th century. Modeling analyses confirmed that the global increase in temperature could have opened new suitable habitats to *C. imicola* in the Mediterranean basin in the 20th century, allowing thus the settlement of new and abundant populations (Purse *et al.* 2005) followed by a subsequent increase of bluetongue transmission (Guis *et al.* 2012).

Three phylogeographical studies analyzed the sequence polymorphism of the mitochondrial gene cytochrome oxidase I (COI) of *C. imicola* (Calvo *et al.* 2009; Dallas *et al.* 2003; Nolan *et al.* 2008). They all concluded that a matrilineal differentiation between western (Portugal,

Spain, Corsica, Italy, Morocco, Algeria) and eastern (Greece, Turkey, Israel) Mediterranean populations exists (Calvo *et al.* 2009; Dallas *et al.* 2003; Nolan *et al.* 2008). The authors concluded (i) a northward range expansion of *C. imicola* from two or three genetically distinct sources, with North African populations representing the most likely source of the western Mediterranean populations, (ii) the occurrence of two independent routes of colonization under the assumption of a joined colonization of both the bluetongue virus and its vector (Dallas *et al.* 2003; Nolan *et al.* 2008), and (iii) a recent and rapid colonization process in Spain (Calvo *et al.* 2009). The two described routes of colonization were as follows: the first started in North Africa to reach Italy, via Sicily and Corsica; the second connected Israel and Turkey to Greece and Bulgaria. However, these three studies were based on the use of a single genetic marker, which limits the understanding of the evolutionary and demographic history. This was recently complemented by a genetic study that used ten microsatellite markers to characterize the genetic structure of nuclear polymorphism of *C. imicola* in the western Mediterranean basin (North Africa, Italy and France). This study revealed low levels of genetic variation among these populations, indicating that they either share a recent common origin or recurrently exchange genes (Mardulyn *et al.* 2013). Unlike previous studies, the authors indicated an ancient presence of *C. imicola* in Italy submitted to recurrent immigration from North Africa.

Interestingly, none of these studies addressed the issue of the history, population sources and routes of the colonization of the Mediterranean basin in relation to the native area of *C. imicola*. Moreover, they did not fully characterize the timeline of *C. imicola* expansion; a point remaining intensively debated in the literature. We designed the present study to precisely address these points. We thus included populations from both the native range and the Mediterranean basin in a multi-locus study including maternally and bi-parentally inherited markers (i.e., two mitochondrial genes and nine microsatellite markers). More

specifically, we depicted the geographical pattern of genetic variation of *C. imicola* in the Afrotropical region and in the Mediterranean basin by a phylogeographical approach. We then characterized the genetic structure using Bayesian clustering and traditional population genetics tools. We finally performed approximate Bayesian computation analyses (ABC) (Beaumont *et al.* 2002; Bertorelle *et al.* 2010) in order to retrace and date the colonization events.

Materials and Methods

Sampling and PCR amplification

Insects were sampled at 27 sites throughout southern Europe and Africa (Table 1, Fig. 1 circle symbol), including 13 sites from the native range of this species (i.e. sub-Saharan Africa and Indian Ocean). In addition, previously published COI sequences from Turkey and the United Arab Emirates (Nolan *et al.* 2008) were added to the dataset (Table 1, Fig. 1 triangular symbol). Adult midges were collected using black light suction traps placed near livestock or horses. Specimens were preserved in 70% ethanol and *C. imicola* individuals identified and sexed under a binocular microscope using the description references (Delecolle & De La Rocque 2002). Both male and female were used for genotyping and sequencing.

Genomic DNA was extracted from single midges using the NucleoSpin96 Tissue Kit (Macherey-Nagel, Duren, Germany) according to the manufacturer's instructions, starting with an additional step where each individual midge was ground in 200 µL of 1X PBS buffer. Each collected individual (11 to 34 individuals per site) was genotyped at nine microsatellites markers previously developed for *C. imicola* (Mardulyn *et al.* 2013) (Table S1). Seven microsatellite loci (68, 12b, 31, 41b, 88b, 16, 88) were amplified by multiplex PCR with the Type-it-Microsatellites kit (Qiagen, Valencia, CA, USA) according to the protocol described

in the manufacturer's manual and the annealing temperature given in Table S1. Simplex PCR reactions were carried out for two microsatellites markers (3b, 35t), in 20 µl of 1X Qiagen® reaction buffer (Qiagen, Valencia, CA, USA), 0.1 mM each dNTP, 0.2 µM of each primer and 0.6 U of Qiagen Polymerase Taq® and 5 ng/µl of genomic DNA. Standard conditions for PCR amplification included an initial denaturation step of 95°C for 5 min, 35 cycles of denaturation for 30 s at 95°C, annealing for 1 min at variable temperature (Table S1), 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. Among the samples that were successfully genotyped, 201 randomly selected insects (four to nine insects per locality) were sequenced for a portion of the mitochondrial genes cytochrome oxidase subunit I (COI, ~ 474 bp) and cytochrome b (CytB, ~ 633 bp) and the nuclear gene Elongation factor alpha (Efα, ~ 555 bp). PCR fragments and sequences were obtained using, respectively, the primers C1J1718 (CCGGTAAAATTAAAATATAAACTTC) and C1N2191 (GGAGGATTTGGAAATTGATTAGTTCC) (Simon *et al.* 1994), CytB_12329F (GCACCTTCTAATATTTCAATTTGGT) and CytB_13038R (CTGGAATAAAATTATCTGGGTCTCC) and EFα_F1 (CGCCAAGTACTACGTCACCA) and Efα_R3 (GGAGCGAAGACAACAACC-AT) designed in this study. PCR amplification reactions for both mitochondrial genes were performed in a 20 µl total reaction volume containing 1X of Qiagen buffer (Qiagen, Valencia, CA, USA), 2.5 mM each dNTP, 0.2 µM of each primer, 0.6 U of Qiagen Polymerase Taq® and 5 ng/µl of genomic DNA. A single denaturing step at 95°C for 5 min was followed by 5 cycles (denaturation at 95°C for 30 s, annealing at 45°C (COI) or 60°C (CytB) for 40 s and elongation at 72°C for 1 min), then 30 cycles (denaturation at 95°C for 30 s, annealing at 51°C for 40 s and elongation at 72°C for 1 min) and final extension at 72°C for 8 min. For EFα gene, PCR amplification reactions were conducted in a 20 µl total

reaction volume containing 1X of Qiagen buffer (Qiagen, Valencia, CA, USA), 2 mM of $MgCl_2$, 2.5 mM each dNTP, 0.2 µM of each primer, 0.8 U of Qiagen Polymerase Taq® and 5 ng/µl of genomic DNA. A single denaturing step at 95°C for 5 min was followed by 5 cycles (denaturation at 95°C for 30 s, annealing at 52°C for 1 min and elongation at 72°C for 1 min), then 35 cycles (denaturation at 95°C for 30 s, annealing at 45°C for 30 s and elongation at 72°C for 1min) and final extension at 72°C for 8 min.

Sequence analyses

Sequences were aligned with the Clustal W algorithm (Thompson *et al.* 1994) available in the software GENEIOUS v.6.0.5 (Biomatters, http://www.geneious.com).

The nuclear sequences of EFα gene obtained from 204 *C. imicola* individuals were not used for extensive analyses owing to a lack of polymorphism. The haplotype network and results of mismatch distributions and Bayesian clustering analyses are presented as supplementary data (Fig. S1).

Twelve CytB sequences from eight sub-Saharan (Benin, Burkina Faso, Cameroon, Ethiopia, Mali, Mauritius, Mozambique, and Reunion Island) and Israeli populations contained nucleotide uncertainties. We performed all analyses on both markers. For population structure inference (genetic differentiation and Bayesian clustering analyses) the results of CytB are not presented as they were less informative.

Population structure

Population structure was assessed with the Bayesian clustering of genotypes implemented in BAPS v.6.0 (Cheng *et al.* 2013). The analysis was conducted with a series of 50 replicates runs and a maximum number of populations (K) set to 13 (i.e., the number of presently

sampled native sites). Given the limited number of samples in the eastern Mediterranean group and taking geographic locations into account, we grouped the Israeli population with those of the eastern Mediterranean basin cluster (see results) for further analyses. Genetic differentiation between the clusters of sequences inferred by BAPS was tested with ARLEQUIN v.3.5.2 (Excoffier *et al.* 2005).

Genetic diversity and genealogical relationships

Genetic diversity within populations and clusters was evaluated by the haplotype number (H), haplotype diversity (Hd) and nucleotide diversity (π) per site using DNASP v.5.10 (Librado & Rozas 2009). To infer genealogical relationships among populations, we constructed a median-joining network (Bandelt *et al.* 1999) for each gene with NETWORK v.4.6.1.2 (www.fluxus-engineering.com).

Demographic history

The genetic signature of past demographic changes within the inferred clusters was investigated from the COI and CytB concatenated dataset. We performed neutrality tests based on Tajima's D and Fu's Fs statistics with DNASP v.5.10. Significant negative values (i.e., significant rejection of the null hypothesis) are expected in populations that had experienced an increase in effective population size (Fu 1997; Ramos-Onsins & Rozas 2002; Tajima 1989). We also computed a mismatch distribution test with ARLEQUIN v.3.5.2 (Excoffier *et al.* 2005). In populations that have undergone a rapid demographic expansion, the mismatch distribution is expected to have a smooth unimodal curve (Rogers & Harpending 1992). The time of expansion (t) was then estimated using the equation $t= \tau /2u$, where tau (τ) is estimated through the mismatch test and $u = 2 \mu k$, with k describing the

sequence length and μ the mutation rate ranging from 0.0075 to 0.0211 substitutions/site/lineage/Myr (Papadopoulou *et al.* 2010). We further performed a Bayesian skyline analysis implemented in BEAST v.1.8 (Drummond *et al.* 2012) in order to quantify and date the changes in effective population size. The analysis was conducted under a random local molecular clock, the HKY+I substitution model and a mutation rate ranging uniformly from 0.0075 to 0.0211 substitutions/site/lineage/My. We ran 100 million generations sampled every 10,000 steps and used a burn-in of 10%. We used TRACER v.1.6 software to analyze the posterior distributions and plot the graph.

Detection of adaptive selection

Sites under positive or negative selection in COI and Cytb genes were inferred using the single-likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), and random-effects likelihood (REL) methods as implemented in DataMonkey server (http://www.datamonkey.org) (Murrell *et al.* 2012; Pond & Frost 2005a, b). Positive selection for a site was considered to be statistically significant if the *P* value was < 0.1 for the SLAC and FEL methods or the posterior probability was at the ≥ 90% level for the REL method. A mixed-effects model of evolution (MEME) was further used to identify selected sites under conditions of episodic diversifying selection. Selected sites with a *P* value < 0.05 were reported.

Microsatellite analyses

Population structure and genetic diversity

Linkage disequilibrium between locus pairs was tested using FSTAT v2.9.3.2 (Goudet 1995). The same software was also used for performing the following standard genetic analyses. Genetic differentiation among samples and within-samples departures from Hardy-Weinberg

proportions were assessed through the Weir and Cockerham (1984)'s unbiased estimates F_{ST} and F_{IS}. A significant deviation of F_{ST} from zero was tested using the exact G test over 10,000 permutations of genotypes among samples. Significant deviations of F_{IS} from zero were tested through 10,000 allelic permutations among the genotypes belonging to the same samples. The presence of null alleles was tested with the software MICRO-CHECKER v2.2.3 (Van Oosterhout *et al.* 2004). Within-samples estimates in genetic diversity were assessed by computing the allelic richness (Ar) and the mean genetic diversity (He, Nei & Chesser 1983) with FSTAT v2.9.3.2. Observed and expected heterozygosity were computed for each population using the software ARLEQUIN v.3.5.2.

Population genetic structure was inferred using the Bayesian approach implemented in STRUCTURE v.2.3.3 (Pritchard *et al.* 2000) which assigns individuals to a defined number of genetic clusters according to their genotypes. We performed 10 independent runs for each value of K varying from 1 to 13 (i.e., the number of presently sampled native sites) under the admixture model and correlated alleles frequencies (Falush *et al.* 2003). We used the sampling locations as priors' information (Locprior model) in order to assist the clustering process. Each run consisted of a burn-in of 10,000 steps followed by a Monte Carlo Markov Chain (MCMC) of 10 million iterations. The accurate number of clusters was inferred with the ΔK method (Evanno *et al.* 2005). The same analysis was performed again within the inferred clusters to assess potential genetic sub-structure. In addition, because the model used by STRUCTURE assumes Hardy-Weinberg equilibrium for all loci, we performed the analyses for the complete data set, but also for the dataset without the three loci for which Hardy-Weinberg equilibrium was significantly rejected across samples (35t, 16, 88). The figures were edited with DISTRUCT v.1.1 (https://web.stanford.edu/group/rosenberglab/distruct.html).

To visualize the genetic structure and relationships between sampled sites, we constructed a neighbor-joining (NJ) tree (Takezaki & Nei 1996) based on the pairwise genetic distances of Cavalli-Sforza and Edwards (1967) using the software POPULATIONS v.1.2.30 (http://bioinformatics.org/~tryphon/populations/). The robustness of nodes was evaluated by performing 1,000 bootstrap replicates.

Colonization scenarios inference

Approximate Bayesian computation (ABC) methods are model-based approaches allowing the inference of complex evolutionary scenarios using summary statistics to compare simulated and real datasets (Beaumont *et al.* 2002). We used DIYABC software v.2.0.4 (Cornuet *et al.* 2014; Cornuet *et al.* 2010) for inferring the possible routes of *C. imicola* colonization. We focused on the questions dealing with the origin and colonization routes toward the western and eastern Mediterranean basin. Table 2 presents the different tested scenarios. The possibilities of incomplete sampling and of genetic sub-structuring within African clusters were taken into account by modeling unsampled populations as described by Lombaert *et al.* (2011). This implies that the colonized population originated from an unsampled population belonging to one African cluster.

The analyses were conducted on microsatellite data by choosing one representative sample displaying the lowest mean F_{ST} per cluster. We also tested the scenarios using the sequence data as well as the combined microsatellite and sequence datasets. However, these simulations never reached convergence for an accurate model comparison even when we fairly increased the number of iterations. We therefore used the microsatellite data only. Prior distributions of demographic, historical and mutational parameters are given in Table S2. For scaling historical parameters, we assumed 10 generations per year (Braverman & Linley

1988) and a divergence time within the last 30,000 generations and starting 330 and 430 generations ago for the western and eastern cluster, respectively (first record of *C. imicola* in the Mediterranean basin (first record of C. imicola in the Mediterranean basin; Bailly-Choumara & Kremer 1970; Szadziewski 1984). The average microsatellite mutation rate prior was set to range from 6.10^{-6} to 10^{-4} substitution per generation on a loguniform distribution.

All observed and simulated data sets were summarized with a set of statistics implemented in DIYABC including the mean number of alleles, the mean expected heterozygosity (Nei 1987), the mean allelic size variance, the Garza-Williamson's M (mean ratio of the number of alleles over the range of allele sizes) (Garza & Williamson 2001), pairwise F_{ST} values (Weir & Cockerham 1984) and the classification index (*mean individual assignment likelihood*) (Pascual *et al.* 2007). We generated 1 million simulated data sets per tested scenario. The posterior probabilities associated with each scenario were calculated by a polychotomous logistic regression (Cornuet *et al.* 2008) performed on the 1% of the simulated data sets closest to the observed data set. The most probable scenario (with the highest probability) was selected.

As a first quality control of the analysis, we performed three replications of our ABC analysis by using site samples belonging to the same genetic unit as replicates of the same evolutionary history (Table S8). Secondly, we tested the robustness of the model divergence time by re-running the model with different starting generation values (100 and 1000 generations).

Confidence in the selected scenario was evaluated by simulating 100 pseudo-observed data sets (pods) of each scenario using parameter values drawn from prior distributions (Table S2). LDA-transformed summary statistics were used to compute posterior probabilities used to calculate type I and II errors. The latter refer respectively to the probability of excluding the selected scenario when it is true and the probability of selecting the scenario when it is false. Mean type II error was calculated over all scenarios.

Finally, we assessed the goodness of fit of the selected scenario by using the model checking option of DIYABC software (Cornuet *et al.* 2010). This allows evaluation of whether the selected scenario matches well with the observed genetic dataset. If the selected scenario fits the observed data correctly, we expect data simulated under this model with parameters drawn from their posterior distribution to be close to the observed data. Our set of statistics included the summary statistics used for the model selection as well as the statistics that were not used for previous ABC treatment.

Results

Population structure and genetic diversity (microsatellite and mtDNA)

All pairs of loci were in linkage equilibrium among the 701 midges collected across the 25 sites (Tunisia and Burkina-Faso were not included due to their low sample sizes) and successfully genotyped at nine microsatellite loci. Significant departures from Hardy-Weinberg equilibrium were noticed in sub-Saharan populations, with F_{IS} ranging from 0.081 to 0.254 (p-value $\leq$ 0.0002) (Table S4). These high F_{IS} values are due to the presence of null alleles observed for three markers (35t, 16, 88) as revealed by the MICROCHECKER analysis; this is not surprising since the primers were designed from European populations (Mardulyn *et al.* 2013). The expected and observed heterozygosity for each population are given in Table S4.

We assessed the genetic structure of *C. imicola* both within its native range and within the presumably colonized area. The clustering analysis reveals a strong geographical structure. The mitochondrial data from a 474-bp-length fragment of the COI mitochondrial gene obtained for 225 individuals showed four major clusters with BAPS v.6.0 (Cheng *et al.* 2013) clustering analysis. Two of these clusters discriminated the western from the eastern Mediterranean populations (Greece and Turkey), defining thus the WMB and EMB clusters, respectively. The third cluster grouped the western African (WA) populations; the fourth and last southeastern African (SEA) cluster merged the samples from central, eastern and southern Africa, Indian Ocean and Middle East (Israel and U.A.E) (Table 1, Fig. 1). The Bayesian clustering performed with STRUCTURE on microsatellite data was slightly different (Fig. 2). At a global scale, ΔK was clearly maximum for $K = 2$, confirming the occurrence of genetic differentiation between the Mediterranean area and the native area of *C. imicola* (Fig. 2b). Within the Mediterranean group, the clustering analysis supports the occurrence of a genetic differentiation between western and eastern Mediterranean populations with $K = 2$ (i.e. the EMB and WMB clusters previously defined; Table 1; Fig. 2b). As previously observed with mitochondrial data, the microsatellite polymorphism failed to detect genetic differentiation between North African and other western Mediterranean populations (Fig. 1 and 2). The slight differences in the results obtained from nuclear and mitochondrial data concerned the native area of *C. imicola*. If microsatellite polymorphisms grouped the western African samples together, they were merged with central and eastern African ones within a central African (CA) cluster. This CA cluster was opposed to a southern African (SA) one grouping southern African and Indian Ocean samples. Interestingly, STRUCTURE analysis revealed admixture events or intermediate frequencies between these CA and SA clusters (Fig. 2b). Three of these four clusters (WMB, EMB, and SA) are consistently retrieved in the STRUCTURE analysis at the global scale (for $K = 4$;

Fig. S2) and the neighbor-joining tree based on microsatellite markers (Fig. 3). Such support is not clearly shown for the CA cluster given the position of the Ethiopian and Kenyan samples in the neighbor-joining tree (Fig. 3). However, the microsatellite polymorphisms tended to group together the western African samples in the Bayesian clustering (for K = 4; Fig. S2) and the neighbor-joining tree (Fig. 3). The defined genetic groups were also obtained with the data set analyzed without the three markers exhibiting the presence of null alleles.

Despite the high geographical distances involved, pairwise F_{ST} estimates based on microsatellite data (Table 3, Table S5) remained relatively low within the WMB cluster ($F_{ST} \leq 0.06$) as well as within and between CA and SA clusters ($F_{ST} \leq 0.07$). By contrast, higher F_{ST} estimates ($0.10 \leq F_{ST} \leq 0.24$) were recorded within the EMB cluster as well as when comparing any EMB sample with any WMB sample ($F_{ST} = 0.12$). Such WMB-EMB genetic differentiation was significantly non-null ($P < 0.05$). This differentiation pattern was also supported by COI sequences (details not shown).

We investigated the genetic diversity within populations. A higher genetic diversity of *C. imicola* was found in the native distribution area than in the Mediterranean basin. Indeed, using microsatellite polymorphism, we obtained an allelic richness ranging from 5.76 to 7.11 alleles per locus in the native area (i.e. CA and SA clusters), 3.43 to 4.43 in western Mediterranean basin and 3.80 to 3.93 in eastern Mediterranean basin (Table S4). The same signal was noticed for the genetic diversity that ranged from 0.67 to 0.77 in the native area (CA and SA), from 0.57 to 0.64 in WMB and was estimated to 0.49 in EMB (Table S4). Mitochondrial data displayed the same picture: 39 COI and 22 CytB haplotypes were discriminated among sub-Saharan African samples but only 13 COI and 20 CytB haplotypes among Mediterranean samples. Moreover, at both mitochondrial genes, the haplotype diversity and nucleotide diversity were higher in sub-Saharan Africa than in the Mediterranean basin (Table S3).

Genealogical relationships

A median-joining network based on COI sequences (Fig. 4) suggested strong relationships between EMB (Greece and Israel) and SEA (southeastern Africa). It also individualized WMB as a genetically distinct cluster sharing no COI haplotype with any of the three other clusters. The pattern of genetic variation within the western Mediterranean basin is characterized by the presence of two widespread dominant COI haplotypes (H39, H40) and few rare haplotypes. This gives thus a star-like shape to the network; i.e. a signature characteristic of populations that have undergone a demographic expansion. In contrast, the genetic variation in eastern Mediterranean basin is not consistent with recent expansion. The same pattern was observed with CytB sequences (Fig. S3).

Demographic history and detection of sites under selction

We explored demographic history across the genetic groups inferred by the clustering analysis. The neutrality and expansion tests based on mitochondrial genes (Table S6) suggest demographic expansion within the WMB and WA clusters, as indicated by the significantly negatives values obtained for Fu's estimates. However, Tajima's D significantly rejected neutrality in the WMB cluster only ($D = -1.972$, p-value < 0.05), indicating thus either the effect of natural selection or past demographic expansion in the WMB ancestors. The Ramos-Onsins & Rozas R_2 and the raggedness r value were significant in WMB and WA, strengthening the hypothesis of past demographic expansions experienced by the western Mediterranean and western African populations. These results are also congruent with the mismatch distributions which are unimodal for WMB and WA clusters (Fig. 5). In contrast, the mismatch distributions in the SEA cluster is not clearly unimodal which can be due to a spatial heterogeneous grouping of multiple different haplotypes (given the large area included in the SEA cluster). However, the global shape of the mismatch distributions in the SEA

cluster does not reject the hypothesis of past demographic expansion. Likewise, the Bayesian skyline plot (BSP) analysis indicated demographic expansion with an increase of the effective population size in the western Mediterranean basin and sub-Saharan Africa during the last 17,000 years and 80,000 years respectively (Fig. 5). These dates are generally consistent with the time of demographic expansion driven from the value of tau (τ). Indeed, such values are estimated at 27,300 – 76,700 years ago for the western Mediterranean basin and at 61,100 - 190,000 years for sub-Saharan/Indian Ocean populations. By contrast, populations from the EMB cluster display a multimodal distribution. This indicates a demographic equilibrium along the last 90,000 years in the EMB cluster with nevertheless a slight signature of expansion ca 20,000 years ago.

None of the sites were detected as being under positive selection, and few sites (13 of 474 sites and 8 of 633 sites of COI and CytB genes, respectively) were under negative selection.

Inference of historical colonization pathways

We tested the colonization pathways of *C. imicola* out from sub-Saharan Africa into the Mediterranean basin. We used four analytic runs differing from one another by the samples chosen as representative of each cluster. All runs provided the same results (Table S8). The posterior probabilities calculated for each scenario provided higher statistical support to the scenarios #5 and 7 without being able to discriminate among these two. These scenarios assumed that the colonization of the Mediterranean basin by *C. imicola* had resulted from emigrants originating from the CA cluster. They differ from one another by the assumed order of colonization between the western and eastern Mediterranean basin (Table 2). The scenario #5 assumed that the western Mediterranean basin was the first colonized area before acting as a population source for creating the EMB populations. The scenario #7 assumed the east-Mediterranean basin to have been the first colonized area and EMB emigrants to have lately founded WMB.

Running the ABC analysis with earlier and later values of starting generation for divergence time did not impact the results: scenarios #5 and #7 were also selected as best-fitting models ($p = 0.32$ and 0.29, for scenarios #5 and #7 respectively, using earlier values and $p = 0.29$ and 0.32 for later values) which testify the robustness of the model.

We calculated type I and II errors in order to evaluate to what extent these scenarios could be trusted. We obtained type I error rate with a mean value of 0.64 for scenario #5 and 0.51 for scenario #7 (Table 2). These high values of type I error were mostly associated to scenario #7 when simulating scenario #5 and reciprocally. This may reflect that our data are not enough informative to discriminate between both scenarios using our ABC approach. Type II errors are relatively low with a mean value of 0.07 for both scenarios. Model checking was carried out for these two selected scenarios. None of the summary statistics (used and unused for ABC inferences) display low probability (i.e. $P < 0.05$), indicating thus that both scenarios fit well the real dataset (Table S8). This was also confirmed by a Principal Component analysis (PCA): the PCA points simulated from the posterior distribution nicely grouped together and are relatively well centered on the target corresponding to the real dataset (Fig. S4). Altogether, these results indicate that both scenarios # 5 and 7 provide a satisfying description of our real dataset.

Discussion

ABC methods have been successfully used to infer colonization pathways of several invasive species (e.g. Brouat *et al.* 2014; Lombaert *et al.* 2011; Miller *et al.* 2005; Pascual *et al.* 2007). However, most of the studies focused on well-documented colonization for which the introduction dates and/or the colonists origin were already known (e.g. Guillemaud *et al.* 2010; Konecny *et al.* 2013; Lombaert *et al.* 2010; Lombaert *et al.* 2014; Lombaert *et al.*

2011; Miller *et al.* 2005). In our study case, the colonization process of the Mediterranean basin by *C. imicola* was unclear. The records of bluetongue outbreaks during the 1990s suggested the presence of *C. imicola* in the northern Mediterranean regions, which was subsequently confirmed by entomological surveys. However, neither these direct observations nor the genetic studies initially carried out (Dallas *et al.* 2003; Mardulyn *et al.* 2013; Nolan *et al.* 2008) provided information on the timing or routes of colonization. In this current study, we addressed these issues over a large geographic area including the native and colonized range of *C. imicola.* Combining standard population genetics with ABC and phylogeographical analyses using both mitochondrial and nuclear markers, allowed us to demonstrate (1) a major genetic structuring of *C. imicola* between its native area and Mediterranean populations, (2) a genetic structuring within the native range and (3) the previously reported west/east genetic subdivision among Mediterranean populations (Nolan *et al.* 2008). Altogether, these results shed a new light on the timing and routes of colonization of the Mediterranean basin by the bluetongue vector.

Genetic structure of C. imicola *within its native range*

Our study investigated the genetic structure of *C. imicola* within its native area. Maternally (mitochondrial) and bi-parentally (microsatellites) inherited markers congruently discovered a genetic sub-structure of *C. imicola* in sub-Saharan Africa. Mitochondrial polymorphism clearly discriminated western African populations from all the others. Bayesian clustering analysis of nuclear polymorphism grouped the populations from southern Africa and the Indian Ocean together within the SA cluster. Such a cluster was differentiated from the central African cluster (CA) which groups all populations located at low latitudes in a west-to-east strip. The Bayesian analysis revealed the signature of admixture events that could

have blurred an ancestral west/east differentiation. This west/east pattern of genetic differentiation could result from isolated populations deriving from the refuges opened in the glaciations of the Pleistocene. Climatic variations during this period have already been suggested to be a factor driving such pattern of differentiation in other African taxa including mammals (e.g. Barlow *et al.* 2013; Barnett *et al.* 2014; Lorenzen *et al.* 2012) and insects (e.g. Sezonlin *et al.* 2006). This hypothesis would also explain the observed signature of past demographic expansion of African populations that we have dated between 60,000 and 200,000 years ago.

Alternatively, the observed genetic substructure could reflect a genetic differentiation of *C. imicola* populations derived from a widely distributed ancestral population, owing to limited gene flow due to geographical barriers such as desert, forest or water bodies. Thus, this would explain the highest genetic differentiation observed between the geographically most distant populations (West Africa and Indian Ocean). The apparent signature of admixture at the intermediate longitudes would be an artefact due to the inability of the STRUCTURE program to assign the individuals to one of the two clusters, and thus only reflects intermediate allele frequencies between central and southern Africa. This is in accordance with the obtained significant pattern of isolation by distance within sub-Saharan populations (data not shown), suggesting a stepping-stone model of migrations compatible with the high passive dispersal capacity through wind of *C. imicola*.

Colonization history of the Mediterranean basin

During the colonization process, complex demographic events may lead to complex genetic patterns in the colonized area (Estoup & Guillemaud 2010). Thus, inferring the history and routes of colonization of species may constitute a major challenge. In our case, the ABC

analyses unambiguously identified the central Africa cluster as the source of the Mediterranean populations of *C. imicola*. More specifically, confidence analyses showed that the most probable scenarios involved a single introduction event of insects of central African origin into the Mediterranean basin and then a secondary colonization event of Mediterranean insects into new Mediterranean habitats. However, the ABC failed to discriminate the best scenario of this secondary event within the Mediterranean basin. The hypothesis of a unique event of colonization out from the native area is consistent with the microsatellite clustering analysis at the global scale suggesting two main clusters: one genetic cluster within the Mediterranean basin and one in sub-Saharan Africa. Within the Mediterranean basin, our results are congruent with those of previous studies through the support given to west/east genetic differentiation of *C. imicola* (Calvo *et al.* 2009; Dallas *et al.* 2003; Nolan *et al.* 2008). This subdivision is sharp in both microsatellite and mitochondrial data suggesting, under the hypothesis of equilibrium, a long term isolation of these two genetic groups. Two hypothetic scenarios could explain such results: (1) a northward spread of *C. imicola* from sub-Saharan Africa to North African coast via the Sahara followed by an allopatric divergence within Mediterranean basin or (2) a colonization of eastern Mediterranean basin from colonists of sub-Saharan African origin which passed by the Arabian Peninsula followed by subsequent spill-over toward the western Mediterranean basin.

Based on mitochondrial data, the historical demographic analyses suggest the occurrence of demographic expansion in western Mediterranean populations between 27,000 and 77,000 years ago. By contrast, the eastern Mediterranean populations of *C. imicola* seemed to have remained demographically stable over the last 90,000 years, with a slight demographic expansion 20,000 years ago. Interestingly, these estimates overlap the wet phases of the Sahara desert during which it was vegetated, included permanent lakes and was probably

occupied by humans and wild animals (Castaneda *et al.* 2009). These phases are considered as key periods for the migration of fauna, flora and human populations out of sub-Saharan Africa (Castaneda *et al.* 2009; Hooghiemstra *et al.* 1992; Osborne *et al.* 2008). Despite the fact that demographic analyses may be affected by the complex demographic events occurring during the colonization process, our results are consistent with an expansion of *C. imicola* distribution taking place during humidification of the Sahara in the Late Pleistocene and Holocene. These climate changes could have opened new suitable habitats to this species allowing its expansion toward the North African coast.

The clear west/east genetic structure observed in the Mediterranean basin has also been reported in many taxa (e.g. Arnaud-Haond *et al.* 2007; Horn *et al.* 2006; Kousteni *et al.* 2014). Hewitt *et al*. (2000) suggested that during periods of glaciations, many animals and plants species have evolved into different genetic groups in the Mediterranean basin. Three main Mediterranean refuges have been described including the Iberian Peninsula, Italian Peninsula and the eastern Mediterranean basin (Balkan Peninsula, Middle East). Subsequent to the range expansion of *C. imicola* in the North African coast, a geographical isolation in western and eastern refuges during glacial Pleistocene periods could have created the current west/east differentiation pattern. This process is often associated with population contraction/expansion or range reduction (Hewitt 2001). Thus, the discrimination of a WMB cluster with low genetic diversity could result from a population contraction followed by an allopatric differentiation associated to a distinct glacial refuge. By contrast, the higher genetic diversity in eastern Mediterranean basin (EMB cluster) could either reflect a larger effective population size in this second Mediterranean refuge, or the recurrence of gene flows between native and the eastern Mediterranean populations. Alternatively, the presence of deserts in Egypt and Libya could have acted as a barrier to gene flows for *C. imicola*, thus maintaining this west/east differentiation among *C. imicola* Mediterranean populations.

An alternative hypothesis would assume past *C. imicola* spread from sub-Saharan Africa to eastern Mediterranean basin via the Arabian Peninsula. This is supported by the close genealogical relationships observed from the mitochondrial haplotype network between the native populations and the eastern Mediterranean populations. The strong affinity with southeastern African populations likely reflects the involvement of a large founding effective population size consisting of widespread sub-Saharan haplotypes and/or a genetic connectivity with recurrent gene flow between these two areas. The latter hypothesis is consistent with observations of Persian air streams responsible of midges transport from the Arabian Peninsula to Israel (Braverman & Chechik 1996). Under that hypothesis, the western Mediterranean populations could have been colonized by migrants of east-Mediterranean origin via a contact zone located in Egypt and/or Libya. If so, one would expect a gradient in genetic diversity along the Mediterranean coast (i.e. Middle East > northeastern Africa (Egypt/Libya) > northwestern Africa (Morocco/Algeria/Tunisia)) as well as shared haplotypes and/or admixture within the contact zone. Unfortunately, our sampling design is not accurate for testing such a hypothesis because of the paucity of samples collected in the Middle East and in northeastern Africa.

Discrepancies were observed between mitochondrial and microsatellite polymorphisms. The COI haplotype network indicated strong genealogical relationships between south-eastern African and eastern Mediterranean populations as well as between West African and western Mediterranean populations suggesting two routes of introduction, while the microsatellites support a unique introduction event. Although we cannot exclude the hypothesis of two separate introductions into the Mediterranean basin, our results globally favored a unique introduction event. Indeed, the Bayesian clustering results for both microsatellites and nuclear gene EFα support two major groups at the global scale, one in sub-Saharan Africa

and one in the Mediterranean basin (Supplementary data), consistent with a single introduction event from sub-Saharan Africa into the Mediterranean basin. According to the haplotype network, eastern Mediterranean populations are genetically connected to one of the widespread haplotypes distributed throughout sub-Saharan Africa which includes samples that originated from West Africa. This pattern could result from an ongoing gene flow between Middle-East and sub-Saharan Africa following the introduction, while western Mediterranean populations could have been more isolated and experienced a stronger genetic drift. It is worth noting that there is a relative imbalance in the sampling effort between western and eastern Mediterranean regions owing to field accessibility issues. Incomplete sampling can induce genetic bias and lead to incorrect interpretations and conclusions (Muirhead *et al.* 2008). Therefore, a more extensive sampling within eastern Mediterranean and Middle-East is needed to further uncover of the pattern of variation within and among populations as well as the connectivity between populations.

Discrepancies between mitochondrial and nuclear markers are likely to reflect their different sensitivities to demographic changes. Due to their maternal inheritance, the effective population size of mitochondrial genes is fourfold lower than that of nuclear autosomal genes (Birky *et al.* 1983). As a first consequence, mitochondrial markers are much more susceptible to stochastic processes such as genetic drift. In other words, they will tend to exhibit stronger differentiation patterns than nuclear genes over comparable evolutionary time scales (Hare 2001). As a second consequence, a longer coalescence time is observed for autosomal than for mitochondrial markers (Hare 2001).

It would also be interesting in a future study to include *C. imicola* samples from the Far East in order to investigate the genetic relationships between oriental, African and eastern Mediterranean populations. A recent study of the COI sequence variation included specimens from India and China. It showed that far-east populations shared COI haplotypes with South African and Israeli populations (Harrup, personal communication), supporting hence the hypothesis of a genetic connectivity among these areas. Interestingly, such connectivity was also supported by examining the variation in BTV serotypes (Maan *et al.* 2004; Nomikou *et al.* 2009).

Colonization of C.imicola *in southern Europe*

As in previous studies (Mardulyn *et al.* 2013; Nolan *et al.* 2008), our genetic dataset indicated the North African populations of *C. imicola* as the most likely source of colonists for Europe. Moreover, our mitochondrial dataset allowed dating of a demographic expansion in western Mediterranean basin during the last 17,000 years. Although we could not estimate the period of colonization of southwestern Europe, this present study suggests that *C. imicola* might have been present in some Mediterranean territories for a long time. Unlike the western group, the eastern Mediterranean basin displays a strong population genetic structure. These regions are highly mountainous so we could hypothesize geographical barriers to limit *C. imicola* long dispersal there.

The observed geographic subdivision of *C. imicola* populations within the Mediterranean basin matches well the genetic clusters previously described clusters for bluetongue virus (BTV) serotypes. Indeed, phylogenetic studies have identified BTV lineages belonging to either an “eastern” source (BTV1, 9, 16) or to a “western” source (BTV 2) (Nomikou *et al.* 2009; Purse *et al.* 2005). Moreover, two different strains of BTV serotype 4 have been shown to occur from north-west and east Africa (Nomikou *et al.* 2009). The genetic consistency regarding the genetic structure of *C. imicola* and BTV in the Mediterranean basin is likely to

result from a similar demographic history involving similar environmental and/or geographical constraints. However, given the inferred timing of divergence between western Mediterranean and the native populations, *C. imicola* seems to have been established there long before the first report of BTV outbreaks (i.e. 1924 in Cyprus). In other words, other factors than the presence/absence of the vector have driven the spread of BTV in the Mediterranean basin (e.g. increases in vector abundance and/or vector competence, a reduced extrinsic incubation period, etc). Further studies should address these points.

Conclusion

The current study illustrates how molecular data can provide insight into the evolutionary and historical processes underlying colonization. The use of genetic data indicated that *C. imicola* have expanded its distribution range out from the northern part of sub-Saharan Africa to the Mediterranean basin. Discrepancies between nuclear and mitochondrial markers suggest that the species populations could have colonized the Mediterranean basin through a single or two events of introduction. However, our results globally support a unique introduction. The estimated timescales of demographic expansion in Mediterranean populations highlight the potential role of Pleistocene and/or early Holocene climate change in shaping the geographical distribution of this species and do not support the recent colonization of *C. imicola* of the Mediterranean basin. However, a precise divergence time between the sub-Saharan and Mediterranean populations would help to better understand the factors underlying the range expansion of *C. imicola* in the Mediterranean basin.

Acknowledgements

We wish to express our sincere thanks to C. Coste and S. Falala, (Cirad UMR15 CMAEE and INRA UMR1309 CMAEE in Montpellier) for computer support. We are grateful to R. Rivallan for genotyping assistance and M. Chapuis for advises on gene sequencing. This work was supported by the high performance cluster of the UMR AGAP Cirad. SJ was granted by the Chercheur d'Avenir 2011 call from the Languedoc-Roussillon region, France, Cirad and the Ministère de l'Agriculture. This study was partially funded by EU grant FP7-613996 Vmerge and is catalogued by the VMERGE Steering Committee as Vmerge006 (http://www.vmerge.eu). The contents of this publication are the sole responsibility of the authors and don't necessarily reflect the views of the European Commission. MdGW and MZ were supported by the Ministère Français des Affaires Etrangères through the FSP-RenCaRe project within the framework of the Research Platform Production and Conservation in Partnership/RP-PCP.

References

Arnaud-Haond S, Migliaccio M, Diaz-Almela E, *et al.* (2007) Vicariance patterns in the Mediterranean Sea: east?west cleavage and low dispersal in the endemic seagrass Posidonia oceanica. *Journal of Biogeography* **34**, 963-976.

Avise JC (2000) Phylogeography: The History and Formation of Species.

Bailly-Choumara H, Kremer M (1970) Epizooties de 1989, 1990 et 1991. Donnees epidemiologiques et strategie de lutte. *Report of the Direction de l'Elevage. Ministere de l'Agriculture at de la Reforme Agraire, Rabat, Maroc.*

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**, 37-48.

Barlow A, Baker K, Hendry CR, *et al.* (2013) Phylogeography of the widespread African puff adder (Bitis arietans) reveals multiple Pleistocene refugia in southern Africa. *Mol Ecol* **22**, 1134-1157.

Barnett R, Yamaguchi N, Shapiro B, *et al.* (2014) Revealing the maternal demographic history of Panthera leo using ancient DNA and a spatially explicit genealogical analysis. *BMC Evol Biol* **14**, 70.

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

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

Birky CW, Maruyama T, Fuerst P (1983) An approach to population and evolutionary genetic theory

for genes in mitochondria and chloroplasts, and some results. *Genetics* **103**, 513-527.
Boorman J (1986) Presence of bluetongue virus vectors on Rhodes. *Veterinary Record* **118**, 21.
Boorman J, Wilkinson P (1983) Potential vectors of bluetongue in Lesbos, Greece. *Veterinary Record* **113**, 395-396.
Braverman Y, Chechik F (1996) Air streams and the introduction of animal diseases borne on Culicoides (Diptera, Ceratopogonidae) into Israel. *Rev Sci Tech* **15**, 1037-1052.
Braverman Y, Galun R (1973) The occurrence of Culicoides in Israel with reference to the incidence of bluetongue. *Refuah Veterinarith* **30**, 121-127.
Braverman Y, Linley J (1988) Parity and voltinism of several Culicoides spp.(Diptera: Ceratopogonidae) in Israel, as determined by two trapping methods. *Journal of Medical Entomology* **25**, 121-126.
Brouat C, Tollenaere C, Estoup A*, et al.* (2014) Invasion genetics of a human commensal rodent: the black rat Rattus rattus in Madagascar. *Mol Ecol* **23**, 4153-4167.
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. *Bull Entomol Res* **99**, 583-591.
Capela R, Purse B, Pena I*, et al.* (2003) Spatial distribution of Culicoides species in Portugal in relation to the transmission of African horse sickness and bluetongue viruses. *Med Vet Entomol* **17**, 165-177.
Castaneda IS, Mulitza S, Schefuss E*, et al.* (2009) Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. *Proc Natl Acad Sci U S A* **106**, 20159-20163.
Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* **19**, 233-257.
Chaker E, Sfari M, Rabhi M*, et al.* (2005) Note faunistique sur les Culicoides (Diptera, Ceratopogonidae) du Gouvernorat de Monastir (Tunisie). *Parasite* **12**, 359-361.
Cheng L, Connor TR, Siren J, Aanensen DM, Corander J (2013) Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol Biol Evol* **30**, 1224-1228.
Cornuet JM, Pudlo P, Veyssier J*, et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*.
Cornuet JM, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics* **11**, 401.
Cornuet JM, Santos F, Beaumont MA*, et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**, 2713-2719.
Dallas JF, Cruickshank RH, Linton YM*, et al.* (2003) Phylogenetic status and matrilineal structure of the biting midge, Culicoides imicola, in Portugal, Rhodes and Israel. *Med Vet Entomol* **17**, 379-387.
Delecolle J-C, De La Rocque S (2002) Contribution à l'étude des Culicoides de Corse. Liste des espèces recensées en 2000/2001 et redescription du principal vecteur de la fièvre catarrhale ovine: C. imicola Kieffer, 1913 (Diptera, Ceratopogonidae). *Bulletin de la Société entomologique de France* **107**, 371-379.
Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-1973.
Dyce A, Wirth W (1983) Reappraisal of some Indian Culicoides species in the subgenus Avaritia (Diptera: Ceratopogonidae). *Int. J. Entomol* **25**, 221-225.
Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Mol Ecol* **19**, 4113-4130.
Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**, 2611-2620.
Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics online* **1**, 47.
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.
Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and

background selection. *Genetics* **147**, 915-925.
Gambles RM (1949) Bluetongue of Sheep in Cyprus. *Journal of Comparative Pathology and Therapeutics* **59**, 176-190.
Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Mol Ecol* **10**, 305-318.
Goffredo M, Conte A, Cocciolito R, Meiswinkel R (2003) The distribution and abundance of Culicoides imicola in Italy. *Vet Ital* **47**, 22-32.
Goffredo M, Conte A, Meiswinkel R (2004) Distribution and abundance of Culicoides imicola, Obsoletus Complex and Pulicaris Complex (Diptera: Ceratopogonidae) in Italy. *Vet Ital* **40**, 270-273.
Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.
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 (Edinb)* **104**, 88-99.
Guis H, Caminade C, Calvete C*, et al.* (2012) Modelling the effects of past and future climate on the risk of bluetongue emergence in Europe. *J R Soc Interface* **9**, 339-350.
Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends Ecol Evol* **16**, 700-706.
Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.
Hewitt GM (2001) Speciation, hybrid zones and phylogeography—or seeing genes in space and time. *Mol Ecol* **10**, 537-549.
Hooghiemstra H, Stalling H, Agwu COC, Dupont LM (1992) Vegetational and climatic changes at the northern fringe of the sahara 250,000–5000 years BP: evidence from 4 marine pollen records located between Portugal and the Canary Islands. *Review of Palaeobotany and Palynology* **74**, 1-53.
Horn A, Roux-Morabito G, Lieutier F, Kerdelhue C (2006) Phylogeographic structure and past history of the circum-Mediterranean species Tomicus destruens Woll. (Coleoptera: Scolytinae). *Mol Ecol* **15**, 1603-1615.
Konecny A, Estoup A, Duplantier JM*, et al.* (2013) Invasion genetics of the introduced black rat (Rattus rattus) in Senegal, West Africa. *Mol Ecol* **22**, 286-300.
Kousteni V, Kasapidis P, Kotoulas G, Megalofonou P (2014) Strong population genetic structure and contrasting demographic histories for the small-spotted catshark (Scyliorhinus canicula) in the Mediterranean Sea. *Heredity (Edinb)*.
Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.
Lombaert E, Guillemaud T, Cornuet JM*, et al.* (2010) Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLoS One* **5**, e9743.
Lombaert E, Guillemaud T, Lundgren J*, et al.* (2014) Complementarity of statistical treatments to reconstruct worldwide routes of invasion: the case of the Asian ladybird Harmonia axyridis. *Mol Ecol*.
Lombaert E, Guillemaud T, Thomas CE*, et al.* (2011) Inferring the origin of populations introduced from a genetically structured native range by approximate Bayesian computation: case study of the invasive ladybird Harmonia axyridis. *Mol Ecol* **20**, 4654-4670.
Lorenzen ED, Heller R, Siegismund HR (2012) Comparative phylogeography of African savannah ungulates. *Mol Ecol* **21**, 3656-3670.
Maan S, Samuel A, Maan N*, et al.* (2004) Molecular epidemiology of bluetongue viruses from disease outbreaks in the Mediterranean Basin. *Vet Ital* **40**, 489-496.
Mardulyn P, Goffredo M, Conte A*, et al.* (2013) 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. *Mol Ecol* **22**, 2456-2466.
Mellor PS, Baylis M, Mertens PP (2009) *Bluetongue* Elsevier/Academic Press.
Mellor PS, Boorman J, Baylis M (2000) Culicoides biting midges: their role as arbovirus vectors. *Annu Rev Entomol* **45**, 307-340.
Mellor PS, Boorman JPT, Wilkinson PJ, Martinezgomez F (1983) Potential Vectors of Bluetongue and African Horse Sickness Viruses in Spain. *Veterinary Record* **112**, 229-230.

Mellor PS, Carpenter S, Harrup L, Baylis M, Mertens PP (2008) Bluetongue in Europe and the Mediterranean Basin: history of occurrence prior to 2006. *Prev Vet Med* **87**, 4-20.
Mellor PS, Jennings DM, Wilkinson PJ, Boorman JP (1985) Culicoides imicola: a bluetongue virus vector in Spain and Portugal. *Veterinary Record* **116**, 589-590.
Mellor PS, Wittmann EJ (2002) Bluetongue virus in the Mediterranean Basin 1998-2001. *Vet J* **164**, 20-37.
Miller N, Estoup A, Toepfer S*, et al.* (2005) Multiple transatlantic introductions of the western corn rootworm. *Science* **310**, 992.
Miranda MA, Borras D, Rincon C, Alemany A (2003) Presence in the Balearic Islands (Spain) of the midges Culicoides imicola and Culicoides obsoletus group. *Med Vet Entomol* **17**, 52-54.
Monteys VSI, Saiz-Ardanaz M (2003) Culicoides midges in Catalonia (Spain), with special reference to likely bluetongue virus vectors. *Med Vet Entomol* **17**, 288-293.
Monteys VSI, Ventura D, Pages N, Aranda C, Escosa R (2005) Expansion of Culicoides imicola, the main bluetongue virus vector in Europe, into Catalonia, Spain. *Veterinary Record* **156**, 415-+.
Muirhead JR, Gray DK, Kelly DW*, et al.* (2008) Identifying the source of species invasions: sampling intensity vs. genetic diversity. *Mol Ecol* **17**, 1020-1035.
Murrell B, Wertheim JO, Moola S*, et al.* (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* **8**, e1002764-e1002764.
Nei M (1987) *Molecular evolutionary genetics* Columbia University Press, New York.
Nei M, Chesser RK (1983) Estimation of Fixation Indexes and Gene Diversities. *Annals of Human Genetics* **47**, 253-259.
Nolan DV, Dallas JF, Piertney SB, Mordue Luntz AJ (2008) Incursion and range expansion in the bluetongue vector Culicoides imicola in the Mediterranean basin: a phylogeographic analysis. *Med Vet Entomol* **22**, 340-351.
Nomikou K, Dovas CI, Maan S*, et al.* (2009) Evolution and phylogenetic analysis of full-length VP3 genes of Eastern Mediterranean bluetongue virus isolates. *PLoS One* **4**, e6437.
Osborne AH, Vance D, Rohling EJ*, et al.* (2008) A humid corridor across the Sahara for the migration of early modern humans out of Africa 120,000 years ago. *Proc Natl Acad Sci U S A* **105**, 16444-16447.
Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Mol Biol Evol* **27**, 1659-1672.
Pascual M, Chapuis MP, Mestres F*, et al.* (2007) Introduction history of Drosophila subobscura in the New World: a microsatellite-based survey using ABC methods. *Mol Ecol* **16**, 3069-3083.
Patakakis MJ (2004) Culicoides imicola in Greece. *Vet Ital* **40**, 232-234.
Pond SLK, Frost SD (2005a) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531-2533.
Pond SLK, Frost SD (2005b) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22**, 1208-1222.
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*, et al.* (2005) Climate change and the recent emergence of bluetongue in Europe. *Nat Rev Microbiol* **3**, 171-181.
Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* **19**, 2092-2100.
Rawlings P, Capela R, Pro MJ*, et al.* (1998) The relationship between climate and the distribution of Culicoides imicola in Iberia. In: *African Horse Sickness* (eds. Mellor P, Baylis M, Hamblin C, Mertens PC, Calisher C), pp. 93-102. Springer Vienna.
Rogers AR, Harpending H (1992) Population-Growth Makes Waves in the Distribution of Pairwise Genetic-Differences. *Mol Biol Evol* **9**, 552-569.
Sezonlin M, Dupas S, Le Rü B*, et al.* (2006) Phylogeography and population genetics of the maize stalk borer Busseola fusca (Lepidoptera, Noctuidae) in sub-Saharan Africa. *Mol Ecol* **15**, 407-420.
Simberloff D, Martin JL, Genovesi P*, et al.* (2013) Impacts of biological invasions: what's what and the way forward. *Trends Ecol Evol* **28**, 58-66.
Simon C, Frati F, Beckenbach A*, et al.* (1994) Evolution, weighting, and phylogenetic utility of

mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the entomological Society of America* **87**, 651-701.
Szadziewski R (1984) Ceratopogonidae (Diptera) from Algeria. VI. Culicoides Latr. *Polskie pismo entomologiczne= Bulletin entomologique de Pologne*.
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-595.
Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics* **144**, 389-399.
Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680.
Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.
Venail R, Balenghien T, Guis H*, et al.* (2012) Assessing Diversity and Abundance of Vector Populations at a National Scale: Example of Culicoides Surveillance in France After Bluetongue Virus Emergence. **3**, 77-102.
Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population-Structure. *Evolution* **38**, 1358-1370.
Yü I (2005) Ceratopogonidae of China:(Insecta: Diptera). Military Medical Science Press.

Data accessibility

Nucleotide sequences of COI and CytB generated in this study were deposited in GenBank under Accession Numbers KT0264989 - KT027189 and KT027190 - KT027376, respectively.
Microsatellite genotypes, COI alignment, Cytb alignment and Samples ID are available on Dryad doi:10.5061/dryad.1sm6c

Author contributions

SJ, CG, and KH designed the study. SJ and JR genotyped the samples. SJ, CG, EL and KH analyzed the data. XA, TB, CCS, AC, J-CD, AD, MD, MF, IF, LG, MdG-W, MG, YG, AGF, MK, KL, YL, JL, TM, BM, MM, NP, DR, AS, M-LS-R, FS, MTS, GV and MZ collected and/or identified the *C. imicola* samples. CW, TB, HG, CC and JB contributed to the manuscript firstly written by SJ, CG and KH. All authors read and commented the final manuscript version.

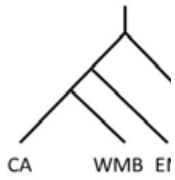
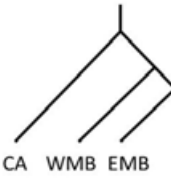
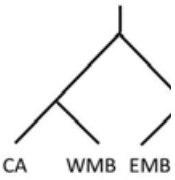
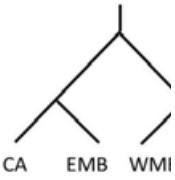
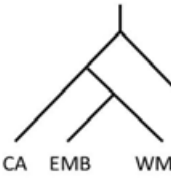
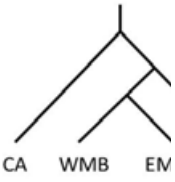
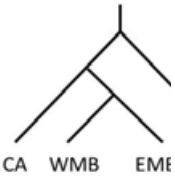
Tables

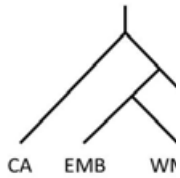
Table 1 Geographical location of sampled sites of *C. imicola*, number of individuals typed for microsatellite analyses (Nmic), number mitochondrial sequences obtained (Nmit) data and sample grouping based on clustering analysis based on microsatellite (Mic) and mitochondrial (mtDNA) data.

Sampling location	Sampling Year	Latitude	Longitude	Nmic	Nmit	Sample Code	Cluster Code (Mic)	Cluster Code (mtDNA)
Algeria	2003	36.8	8.5	32	9	DZ	WMB	WMB
France-Corsica	2008	42.8	9.4	30	8	CO	WMB	WMB
France-Pyrénées-Orientales	2012	42.4	2.8	11	8	PO	WMB	WMB
France-Var	2008	43.2	6.4	27	8	FR	WMB	WMB
Italy-Calabria	2012	39.1	16.9	32	8	IT	WMB	WMB
Italy-Sardinia	2012	39.1	8.5	32	8	SD	WMB	WMB
Italy-Sicily	2012	38.0	12.6	31	8	SI	WMB	WMB
Morocco	2004	34.4	-6.4	32	8	MA	WMB	WMB
Portugal	2010	39.9	-7.4	31	8	PT	WMB	WMB
Spain-Andalusia	2012	37.3	-6.9	30	8	ES	WMB	WMB
Spain-Balearic Islands	2012	39.5	3.1	28	8	BA	WMB	WMB
Tunisia	2013	36.0	10.0	0	5	TN	WMB	WMB
Greece	2013	41.0	24.7	31	8	GR	EMB	EMB
Turkey (Nolan *et al*. 2008)	2001	38.5	27.7	0	21	TR	EMB	EMB
Israel	2010	29.9	35.1	31	8	IL	EMB	SEA*
Benin	2009	11.9	3.4	29	8	BJ	CA	WA
Burkina Faso	2004	11.2	-4.3	0	7	BF	CA	WA
Cameroon	2009	9.3	13.5	29	7	CM	CA	SEA
Mali	2010	11.0	-6.6	23	7	ML	CA	WA
Senegal	2012	12.6	-12.2	32	8	SN	CA	WA
Ethiopia	2004	8.8	40.7	32	6	ET	CA	SEA
Kenya	2013	0.1	37.1	19	7	KE	CA	SEA
Mozambique	2013	-25.9	32.5	20	8	MZ	SA	SEA
South Africa	2013	-33.9	25.5	20	8	ZA	SA	SEA
Zimbabwe	2013	-21.9	31.6	30	8	ZW	SA	SEA
Madagascar	2012	-18.5	47.4	23	7	MG	SA	SEA
Mauritius	2007	-20.2	57.5	34	4	MU	SA	SEA
France-Réunion Island	2005	-21.3	55.4	32	6	RE	SA	SEA
U.A.E. (Nolan *et al*. 2008)	2005	25.2	55.3	0	3	AE	SA	SEA

Clusters, obtained with STRUCTURE for microsatellite and BAPS for mitochondrial data, are coded as: WMB, western Mediterranean Basin, EMB, eastern Mediterranean Basin, CA, central Africa, WA, western Africa, SA, southern Africa, SEA, southeastern Africa.
(*) The Israeli population was grouped with the EMB cluster for all analyses

Table 2 Description of the scenarios tested by approximate Bayesian computation analyses (ABC) on microsatellite data attempting to retrace the routes of colonization of *C. imicola* and confidence in scenario selection based on posterior probabilities, 95% confidence intervals and type I and II errors. Type I error is the probability of selecting another scenario when the chosen scenario is true. Type II error is the mean probability of selecting the chosen scenario when it is false. The selected (most probable) scenario is highlighted in bold.

Scenarios	Description of tested scenarios		Posterior Probability	95% Credibility interval	Type I Error	Type II error
1	Introduction out from CA independently to WMB and EMB		0.0494	[0.0136, 0.0852]		
2	Introduction out from SA independently to WMB and EMB		0.0026	[0.0000, 0.0409]		
3	Introduction out from CA to WMB and from SA to EMB		0.0096	[0.0000, 0.0474]		
4	Introduction out from CA to EMB and from SA to WMB		0.0039	[0.0000, 0.0422]		
5	**Introduction out from CA to WMB then introduction from WMB to EMB**		**0.4269**	**[0.3768, 0.4770]**	**0.64**	**0.07**
6	Introduction out from SA to WMB then introduction from WMB to EMB		0.0591	[0.0242, 0.0941]		
7	**Introduction out from CA to EMB then introduction from EMB to WMB**		**0.4103**	**[0.3645, 0.4562]**	**0.51**	**0.07**

8	Introduction out from SA to EMB then introduction from EMB to WMB		0.0381	[0.0023, 0.0740]

ABC analyses were performed using one representative population from each cluster: Ethiopia, Zimbabwe, Morocco and Greece. Clusters are coded as: WMB, western Mediterranean Basin, EMB, eastern Mediterranean Basin, CA, central Africa, WA, western Africa, SA, southern Africa, SEA, southeastern Africa.

Table 3 Pairwise *FST* values across loci between the genetic clusters inferred by STRUCTURE v.2.3.3 of *C. imicola*. Population differentiation was assessed with the exact G test implemented in FSTAT v2.9.3.2. Significant values, at the adjusted nominal level (5%) for multiple comparison of 0.0083, are highlighted in bold.

	southern Africa	western Mediterranean Basin	eastern Mediterranean Basin
central Africa	**0.0240**	**0.0746**	**0.1223**
southern Africa	-	**0.0891**	**0.1593**
western Mediterranean Basin	-	-	**0.1247**

Figures

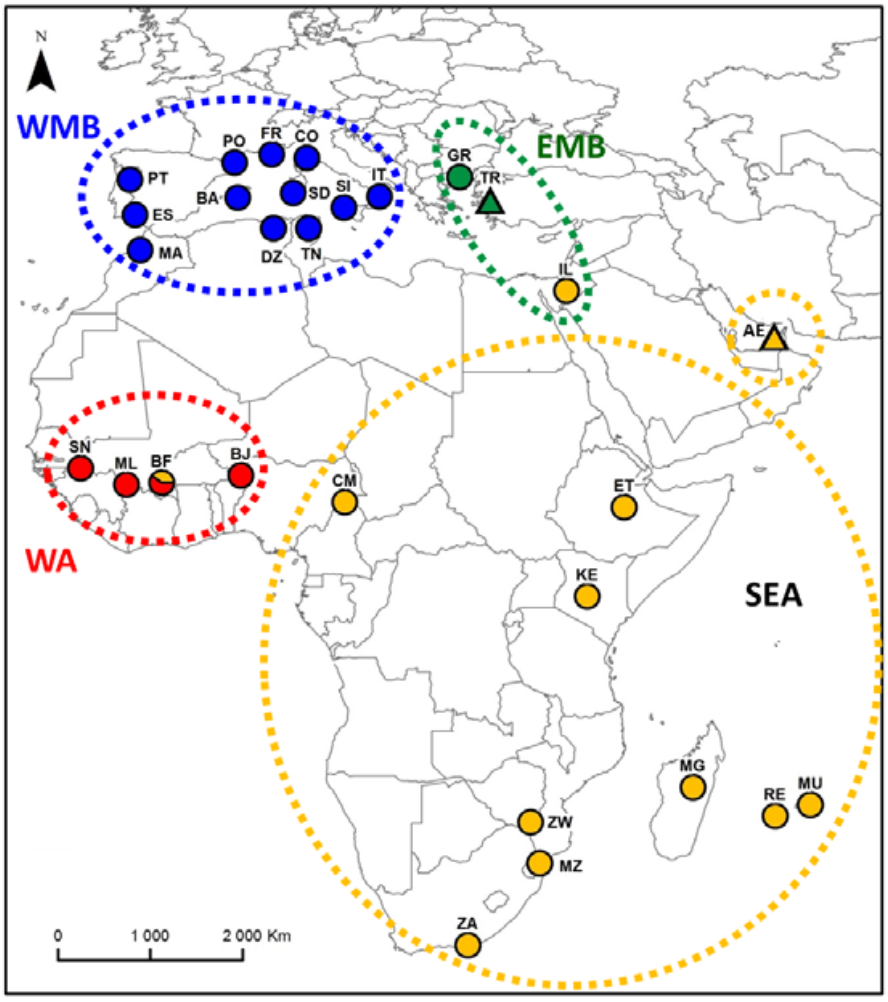


Fig. 1 Sampling locations for *C. imicola* specimens (see sample codes in Table 1) and genetic clustering of *C. imicola* sampled populations based on COI sequences. Genetic groups were assessed by the spatial group clustering method of Corander *et al.* (2004) implemented in BAPS v.6.0. Sample sites with the same color belong to the same cluster.

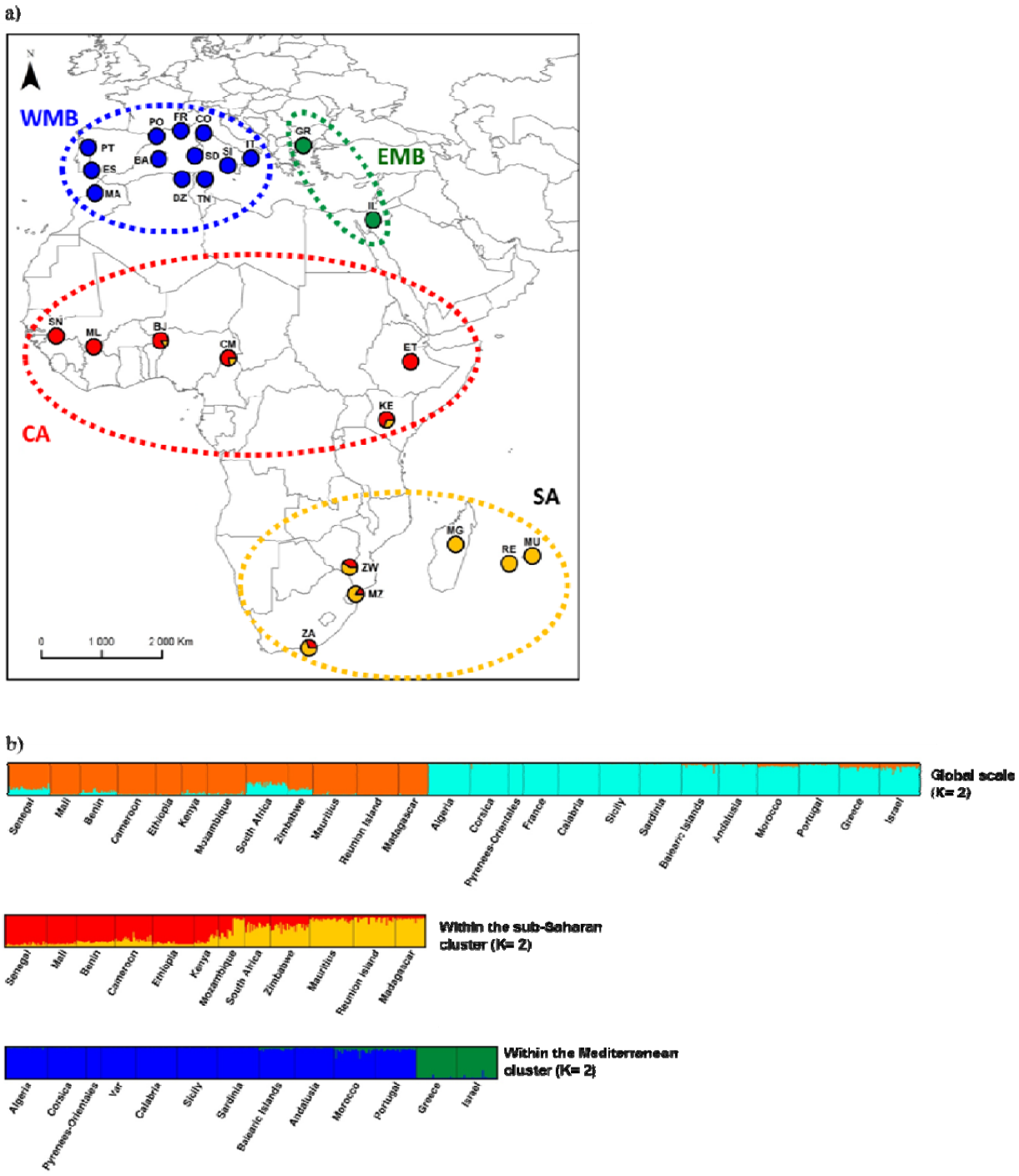


Fig. 2 Genetic clustering of *C. imicola* sampled populations. (a) Spatial Bayesian clustering based on microsatellite data. (b) Ancestry estimation assuming two population clusters at the global scale (upper part) and four population clusters at the genetic groups scale (K=2 within sub-Saharan Africa and Indian Ocean area (center part) and K=2 within the Mediterranean basin (lower part) based on the Bayesian clustering method implemented in STRUCTURE v.2.3.3. Each vertical line represents an individual, and each color represents a cluster. Individuals are grouped by sample site.

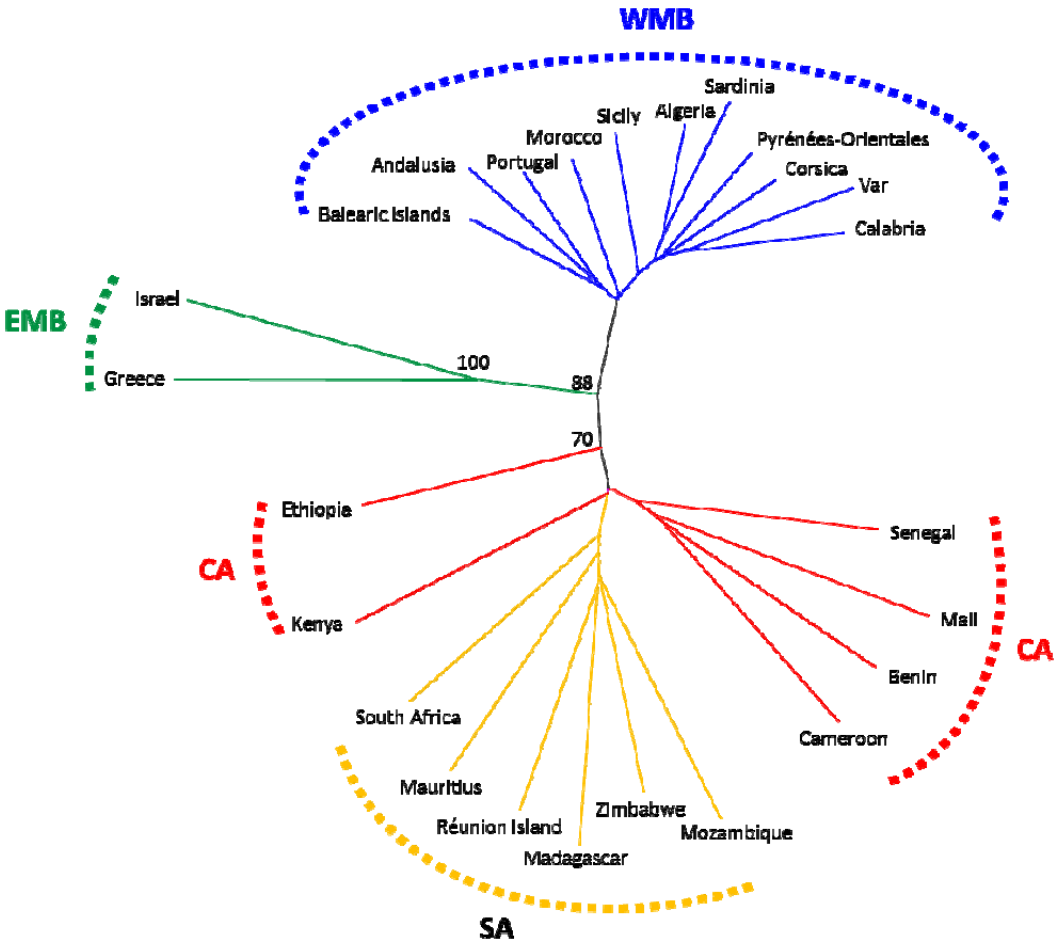


Fig. 3 Neighbor-joining tree of *C. imicola* population samples based on the chord distance of Cavalli-Sforza & Edwards (1967) computed on microsatellite polymorphism. Central African populations are shown in red, southern African populations in yellow, eastern Mediterranean populations in green and western Mediterranean populations in blue. Bootstraps values were calculated over 1000 replicates and are represented as percentage. Only values > 60% are reported.

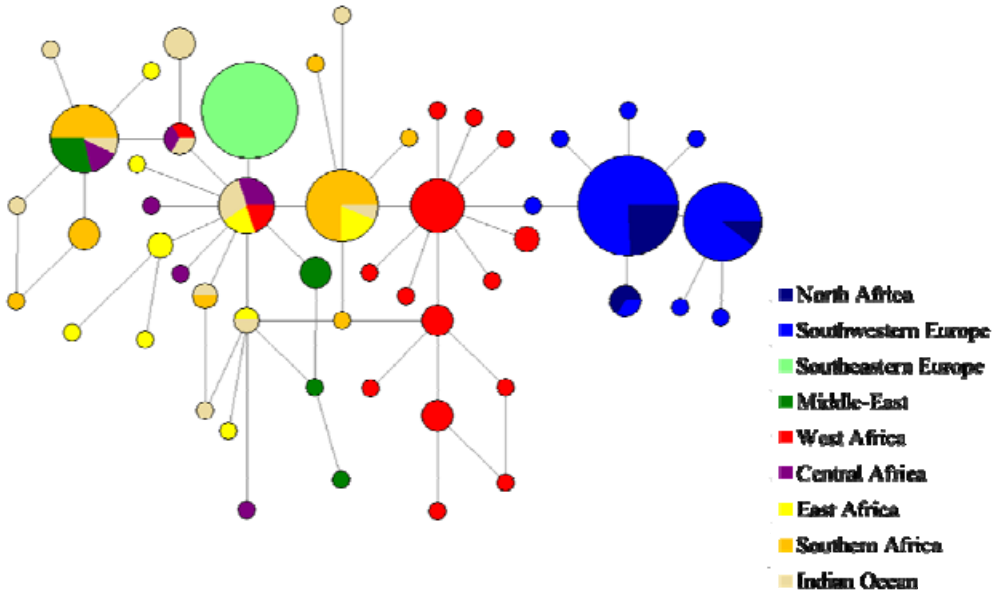


Fig. 4 Median-joining haplotype network of COI mitochondrial sequences of *C. imicola*. The size of the circles is proportional to the number of individuals with that haplotype. The length of the branches separating haplotypes is proportional to the number of mutational steps between them. Haplotype networks were constructed using NETWORK v.4.6.1.2 Colours represent the geographical region of sampled specimens. North Africa: Algeria, Morocco, Tunisia; Southwestern Europe: France-Corsica, France-Pyrénées Orientales, France-Var, Italy-Sardinia, Italy-Sicily, Italy-Calabria, Portugal,Spain-Andalusia, Spain-Balearic Islands, Southeastern Europe: Greece, Turkey; Middle East: Israel, Uniated Arab Emirates; West Africa: Senegal, Benin, Mali, Burkin Faso; Central Africa: Cameroon, East Africa: Kenya, Ethiopia; Southern Africa: Mozambique, Zimbabwe, South Africa; Indian Ocean: Mauritius, Madagascar, France-Réunion Island.

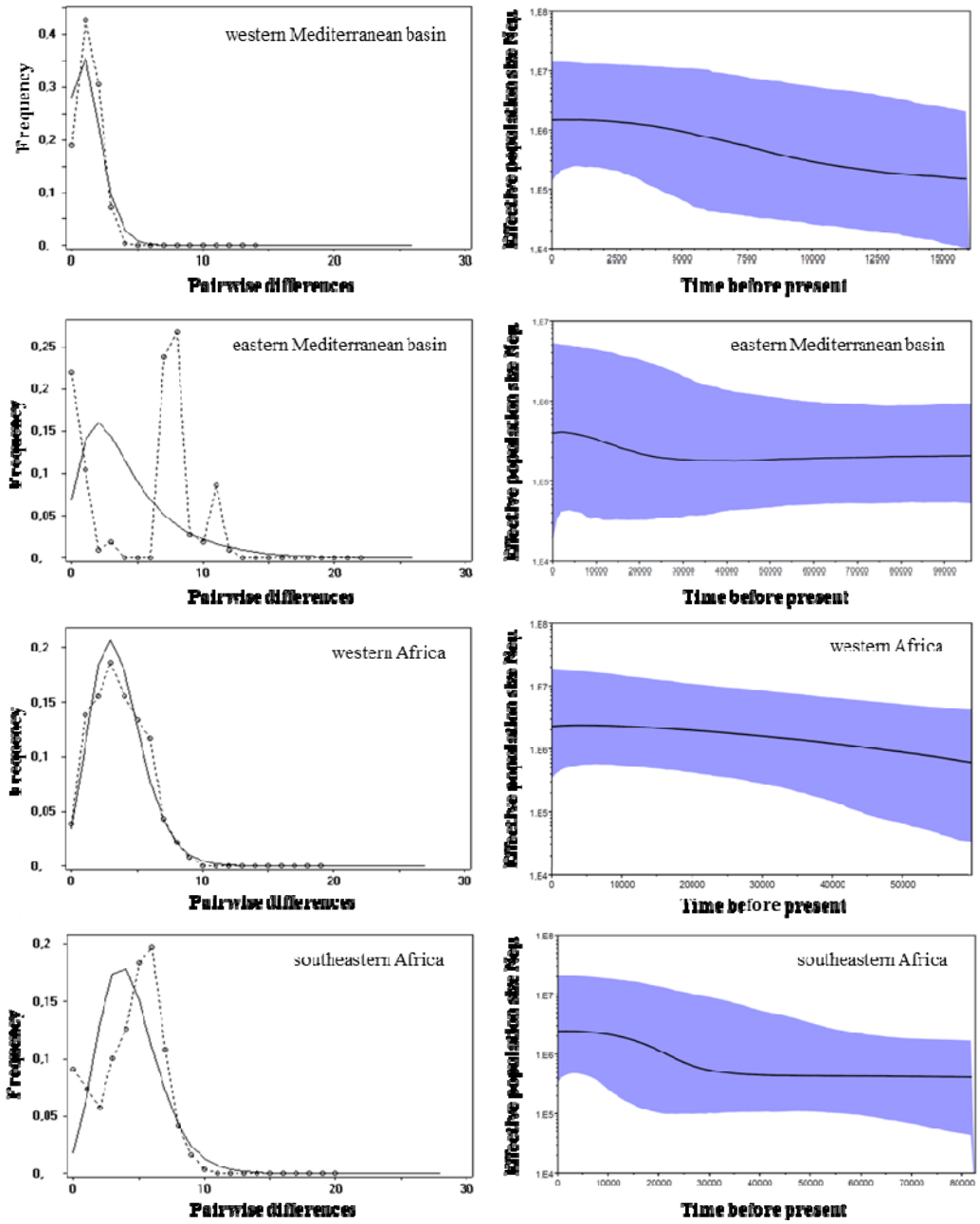


Fig. 5 (a) Mismatch distribution among pairwise differences among haplotypes and (b) Bayesian skyline plot based on COI and Cytb combined dataset in *C. imicola* different geographical groups. Mismatch analyses were conducted according to a growth-decline model. Observed data and theoretical expected distributions are represented by discontinuous and solid line, respectively. Bayesian Skyline plot were performed with a mutation rate of 0.0075 - 0.0115 substitutions/site/Myr and a random local molecular clock. The x-axis indicates times in years before present and the y-axis shows the effective population size. The black line represents the median population size and the grey outline indicated 95% posterior intervals for the population size change.