Table des matières

Table des illustrations : figures

Introduction générale

Chapitre I

Chapitre II

Chapitre III

Article I.

Discussion générale

Table des illustrations : tableaux

Chapitre I

Chapitre II

Chapitre III

Article I.

Préface

Le déplacement d'individus appartenant à une même espèce est un processus normal et naturel, moteur essentiel de l'évolution des espèces car il permet entre autre le maintien de la diversité génétique. Néanmoins, dans un monde qui se transforme sous l'effet de l'accentuation des variations climatiques, de la mondialisation et de l'intensification des échanges entre des régions éloignées de la planète, certaines espèces voient leurs capacités de dispersion rapidement augmenter. Ceci peut conduire à des introductions en dehors des limites contiguës à la zone native de distribution, fréquemment appelées invasions biologiques. Ces invasions biologiques lorsque les populations envahissantes s'installent représentent parfois une menace pour l'équilibre des écosystèmes et peuvent avoir des conséquences économiques ou sanitaires sur les populations humaines. Particulièrement, les déplacements dans de nouveaux territoires d'hôtes, de réservoirs et d'espèces d'arthropodes vectrices avec les pathogènes transmis favorisent l'augmentation des contacts hôtes/vecteurs/pathogènes et contribuent à l'émergence ou la réémergence de maladies à transmission vectorielle d'intérêt médical ou vétérinaire. Les facteurs d'émergence ou de réémergence de ces maladies sont donc multifactoriels, mais sont très souvent liés à la distribution géographique et à la dynamique des populations d'arthropodes vecteurs.

Dans ce contexte, il apparaît essentiel de caractériser et de comprendre les facteurs impliqués dans les processus d'invasions biologiques des arthropodes d'intérêt. Une meilleure connaissance des facteurs historiques, démographiques et évolutifs à l'origine du succès de la colonisation et de l'installation durable des populations en dehors de leurs aires natives permettra (i) d'anticiper des phénomènes d'expansion de populations déjà installées, (ii) de prédire de nouvelles invasions biologiques et (iii) de mettre en place des stratégies de contrôle plus adaptées. Ce travail de thèse s'inscrit dans cette dynamique. Articulé autour de quatre études réalisées à différentes échelles spatiales, il s'intéresse à décrire et comprendre l'histoire évolutive de la colonisation du bassin méditerranéen par le moucheron d'origine afrotropicale *Culicoides imicola* Kieffer, 1913, ainsi que les facteurs expliquant le succès de son expansion. Ce moucheron hématophage est une espèce vectrice majeure responsable entre autre de la transmission de virus d'intérêt en santé animale. Pour la première fois, ces questions sont traitées en s'appuyant sur des données entomologiques collectées sur toute l'aire de

distribution méditerranéenne et africaine de l'espèce, et avec une méthodologie combinant, un génotypage multi-marqueurs des populations (microsatellites, gènes mitochondriaux, gène nucléaire), des analyses statistiques classiques de génétique de populations et de phylogéographie, des analyses basées sur les méthodes *Approximate Bayesian Computation* (ABC) et des simulations mathématiques de la capacité de dispersion passive à une échelle régionale.

Ce travail pris dans son ensemble présente des arguments scientifiques et propose des pistes de réflexions sur le statut invasif de *C. imicola* dans la région méditerranéenne.

Introduction générale

La distribution spatiale d'une espèce est la résultante des événements historiques ayant conduit à la colonisation et à l'installation de groupes d'individus de cette espèce dans de nouveaux milieux. Sous l'effet combiné de modifications géomorphologiques et climatiques du milieu, ces évènements de colonisation modulent en continu les aires de distribution des espèces. Au cours du temps, ces modifications naturelles de la distribution des espèces ont entraîné la création de biotopes[1] régionaux et locaux. De la caractérisation des biotopes et de la description de leurs frontières spatiales naissent les définitions opposant les espèces natives et les espèces non-natives, utilisés pour la première fois en 1835 par le botaniste anglais John Henslow. A cette époque, une espèce native se définit comme un taxon indigène, endémique ou ayant connue une invasion préhistorique (avant la période Néolithique). Par opposition, une espèce non-native à une aire géographique est une espèce qui a connu une expansion naturelle historique de son aire de répartition ou qui a été introduite par l'homme (Carlton 1996; Manchester & Bullock 2000). Ainsi le choix de qualification natif et non-natif peut sembler arbitraire en fonction des auteurs sur l'espèce (Sax *et al.* 2005). D'ailleurs, le nombre et la fréquence des déplacements d'espèces connaissent une augmentation exponentielle depuis le XX^{e} siècle (Roy *et al.* 2011) (Figure 1).

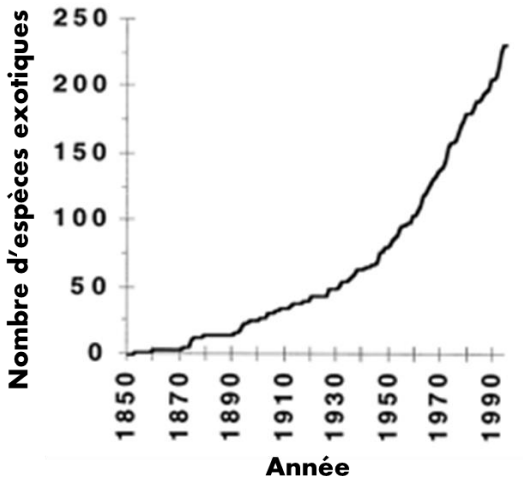


Figure 1. Exemple de dynamique des invasions biologiques. Nombre cumulé d'espèces introduites et établies dans la baie de San Francisco (adapté Cohen & Carlton 1998).

[1] Un biotope est ici désigné comme un milieu biologique déterminé offrant des conditions d'habitat stables à un ensemble d'espèces animales et végétales.

En modifiant leurs environnements, les populations humaines contribuent à la dispersion des espèces hors de leur aire naturelle de répartition à des échelles géographiques et à des vitesses sans précédent (McNeely 2006), suscitant l'intérêt de la communauté scientifique pour décrire et comprendre les patrons et les processus sous-jacents. En exemple, le nombre d'articles de rang A sur le sujet des invasions biologiques ont quasi-quadruplé entre 2004 et 2015 (Figure 2).

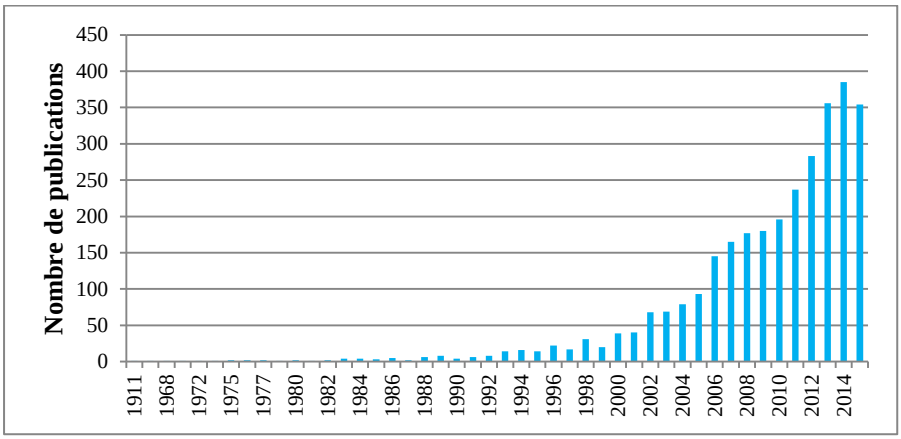


Figure 2. Nombre de publications référencées dans Pubmed sous les mots-clés « biological invasion AND species » (recherche en date du 6/10/15).

Paradoxalement à cet intérêt, la définition d'une invasion biologique n'a toujours pas trouvé de consensus. Cette notion diffère selon les auteurs et les critères qu'ils utilisent pour définir le statut invasif d'une espèce. Classiquement, on retrouve parmi ces critères (i) le rôle des populations humaines dans le déplacement de l'espèce (Lee 2002; Richardson *et al.* 2011), (ii) l'impact de l'espèce dans son nouvel environnement (Davis & Thompson 2000) et (iii) la nécessité d'une déconnexion géographique entre l'aire native et l'environnement colonisé (Colautti & MacIsaac 2004; Mack 2000; Richardson *et al.* 2000). D'autres auteurs considèrent que les invasions biologiques ont toujours eu lieu, et ce bien avant que les populations humaines existent, et préfèrent considérer que les activités anthropiques ont seulement accéléré et augmenté la fréquence de ces évènements (e.g. Mack 2000; Williamson & Fitter 1996). Dans la suite du manuscrit, nous avons choisi de définir la notion d'invasion biologique comme « *le succès d'une espèce à s'établir, proliférer démographiquement et spatialement en dehors de sa zone géographique* » (Facon *et al.* 2006). Nous considérerons l'expression « espèce invasive » comme synonyme d'« espèce exotique envahissante », et pour faciliter la lecture, nous parlerons d'« espèce envahissante ».

Les invasions biologiques peuvent avoir des conséquences sur la biodiversité et l'équilibre des écosystèmes ou sur un certain nombre d'activités économiques telles que l'agriculture, la pêche ou l'activité forestière avec des pertes économiques importantes (Ehrenfeld 2010; Holmes *et al.* 2009; Pimentel 2011; Pimentel *et al.* 2001). De plus, il y a des enjeux sanitaires majeurs pour les populations humaines, animales ou végétales lorsque les espèces envahissantes sont des espèces vectrices ou réservoirs de pathogènes, qu'elles introduisent en même temps qu'elles s'installent, ou lorsque un pathogène est introduit ultérieure (Hulme 2014; Pimentel 2011). L'encadré 1 détaille et illustre les impacts économiques et sanitaires liés aux espèces envahissantes.

La compréhension des processus d'invasions biologiques et l'étude des facteurs démographiques, géographiques et génétiques impliqués dans le succès invasif des espèces envahissantes est un prérequis essentiel à la mise en place de méthodes de gestion et de contrôle des populations introduites (Kulhanek *et al.* 2011). La prévention avec la limitation des voies d'entrée, le contrôle systématique aux points d'entrée, les mesures de quarantaine permettent une détection précoce de l'introduction et sont des stratégies efficaces et moins couteuses que l'éradication (élimination de tous les individus susceptibles de se reproduire) ou le contrôle d'espèces envahissantes (diminution suffisante de la densité des populations afin que ces dernières n'aient plus d'impact significatif) (Simberloff *et al.* 2013) (Figure 3). L'étude des invasions biologiques permet aussi d'anticiper des expansions secondaires et de prédire les impacts écologiques ou économiques des espèces envahissantes (Kulhanek *et al.* 2011).

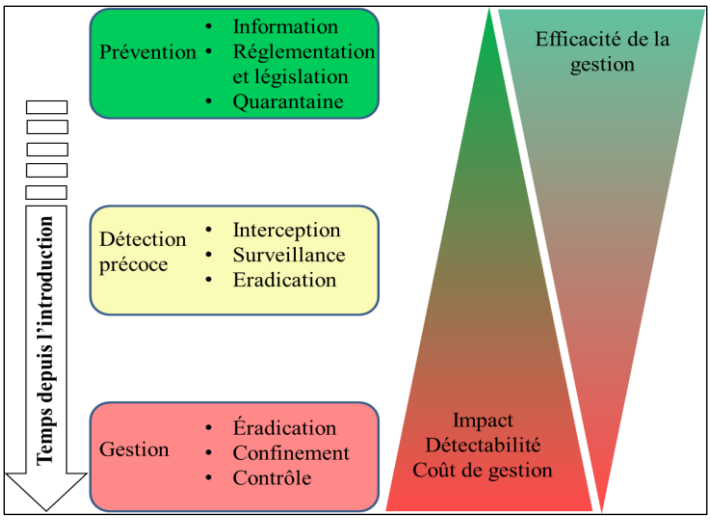


Figure 3. Les stratégies de gestion des espèces envahissantes. La stratégie optimale évolue dans le temps en fonction du temps post-introduction (adapté de Simberloff *et al.* 2013).

Encadré 1. Impacts des invasions biologiques

Les invasions biologiques sont marquées négativement du fait des impacts des espèces envahissantes sur l'environnement, les activités ou la santé des populations humaines. Cette perception relève cependant d'une dimension écologique, économique et socioculturelle (Simberloff *et al.* 2013). De plus, certains impacts peuvent être à peine perceptibles et/ou sur des échelles de temps très longues. C'est par exemple le cas du moustique japonais, *Aedes japonicus* qui a envahi la Belgique, l'Allemagne et la Suisse depuis les années 2000, mais aucun impact sanitaire n'a encore été notifié en dépit de sa compétence vectorielle pour les virus du chikungunya et de la dengue (Medlock *et al*. 2012). Ainsi, il est important de souligner la subjectivité et la difficulté dans la définition du bénéfice ou du coût d'une espèce envahissante.

Des exemples d'impact positif

Le meilleur exemple est certainement celui des espèces introduites à travers le monde par les populations humaines pour se nourrir, telles que les plantes cultivées (tomate, riz, maïs) ou les espèces domestiquées pour l'élevage (Pimentel *et al.* 2001). De même, une grande partie de notre production vestimentaire ou des matériaux de construction provient d'espèces introduites et exploitées en dehors de leurs zones natives (Sax *et al.* 2005). Enfin, le contrôle de certains ravageurs des grandes cultures par l'introduction de prédateurs non-natifs et envahissants permet la réduction durable des pertes de production agricole dans certaines régions (Eilenberg *et al.* 2001).

Des exemples d'impact négatif

Ecologiques. La biodiversité et l'équilibre des écosystèmes sont particulièrement affectés par les invasions biologiques et sont considérées comme la deuxième cause de l'érosion de la biodiversité au niveau mondial (Vitousek *et al.* 1997). A l'arrivée d'une nouvelle espèce dans un écosystème, celle-ci occupe l'espace et utilise les ressources locales. Parfois, l'espèce introduite est accompagnée de ses parasites et pathogènes qui peuvent être néfastes pour les populations indigènes. Ces interactions directes ou indirectes peuvent entrainer la disparition ou la très forte réduction des effectifs des espèces locales (Courchamp *et al.* 2003; Olden *et al.* 2004). Le célèbre cas du frelon asiatique, *Vespa vulitina*, introduit en France et impactant les populations d'abeilles qu'il chasse pour nourrir ses larves (Villemant *et al.* 2011), n'est qu'un exemple parmi d'autres.

Economiques. Les invasions biologiques peuvent être désastreuses pour un certain nombre d'activités économiques telles que l'agriculture, la pêche et l'activité forestière (Born *et al.* 2005). Ces coûts sont difficilement estimables. La Convention sur la Diversité Biologique des Nation-Unies a évalué le coût annuel des dommages et du contrôle des espèces envahissantes aux Etats-Unis à près de 138 milliards de dollars américains. Au niveau mondial, ces coûts s'élèveraient à plus de1 400 milliards de dollars américains (Pimentel et al. 2001). Ainsi, les pertes de production et la mise en place de méthodes de lutte liées à la chrysomèle du maïs *Diabrotica virgifera virgifera* (originaire d'Amérique Centrale) aux Etats-Unis auraient un coût d'environ 1,17 milliards de dollars américains par an (Sappington *et al.* 2006).

Encadré 1. Suite

Sanitaires. Les espèces envahissantes peuvent avoir des effets sur la santé humaine et animale parce qu'elles sont pathogènes ou nuisantes pour ces populations, par la modification des cycles de transmission qu'elles induisent, ou par l'introduction de pathogènes transmis par ces espèces envahissantes (McMichael *et al.* 2000). A titre d'exemple, l'introduction du rat noir, *Rattus rattus* en Europe a été suivie d'épidémies de peste bubonique au XVII^{e} causant de nombreuses victimes sur le continent (Monecke *et al*. 2009). De même, l'introduction du moustique tigre asiatique *Aedes albopictus* en Europe et en Amérique couplée à la circulation de virus pour lesquels l'espèce vectrice est compétente (les virus du chikungunya et de la dengue) ont conduit à l'émergence de ces maladies sur de nouveaux territoires.

I. La colonisation de nouveaux milieux : un processus complexe

Même si le nombre d'invasions biologiques augmentent, il est important de noter que seule une minorité des populations émigrant des aires natives parvient à survivre et à envahir de nouveaux milieux (Williamson & Fitter 1996). Ainsi, on peut se demander pourquoi et comment certaines espèces parviennent à devenir envahissantes.

Le succès d'une invasion biologique implique le franchissement de différentes barrières depuis l'émigration des individus hors de l'aire native (phase d'introduction) jusqu'à la formation et l'installation de populations viables dans l'aire colonisée (phase d'établissement), suivie de leur éventuelle prolifération et expansion géographique (phase d'expansion) (Figure 4). Les barrières peuvent être géographiques, écologiques ou génétiques. La plupart des auteurs s'accordent sur les étapes du processus d'invasion, bien que certains puissent définir des sous-étapes (e.g. Colautti & MacIsaac 2004; Richardson *et al.* 2000; Sakai *et al.* 2001).

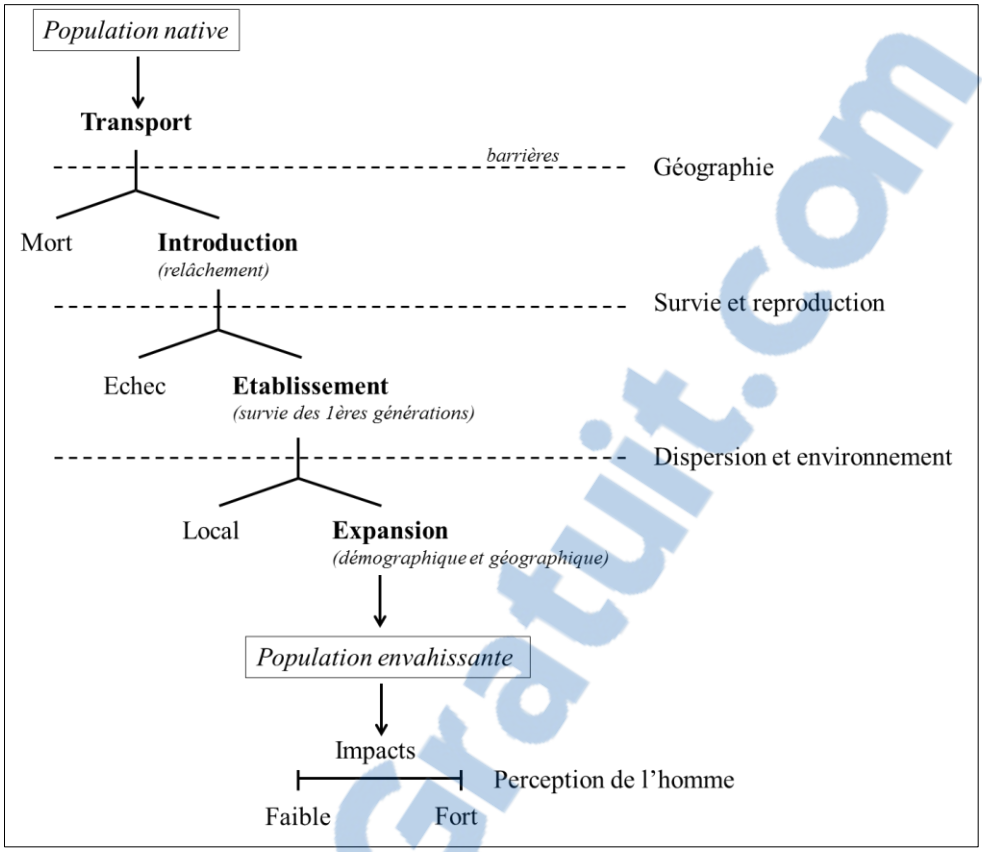


Figure 4. Représentation schématique du processus séquentiel d'invasion biologique (modifié de Lockwood *et al.* 2013). Les quatre étapes du processus d'invasion sont représentées en gras.

I.1. L'étape d'introduction dans le nouveau milieu

La phase d'introduction correspond au franchissement des barrières géographiques qui limitaient initialement l'aire de répartition d'une espèce (Richardson *et al.* 2000). Cette phase est décrite comme étant la plus importante car elle est difficile à réaliser et représente une très bonne cible pour les mesures de gestion (Mack 2000). De ce fait, l'étude des facteurs favorisant le succès d'introduction aux différentes étapes que sont le prélèvement, le transport puis le relâchement des individus dans le nouveau milieu est primordial. Entre autre, le transport d'une population depuis son aire d'origine vers un nouvel environnement requiert une stratégie de dispersion adaptée. La dispersion est un facteur clef du succès de l'invasion en permettant l'introduction d'individus dans de nouveaux environnements et en contribuant à la dynamique des populations installées (Kokko & López-Sepulcre 2006). Les espèces se dispersent en utilisant différents modes de dispersion : naturelle ou anthropique, et sur de courtes ou longues distances.

I.1.1. Les modèles de dispersion

Un des modèles de dispersion les plus simples est une diffusion graduelle par laquelle les individus se déplacent en continu à la limite de l'aire de distribution de l'espèce (Wilson *et al.* 2009). Ce mode de dispersion nécessite le maintien d'une connexion physique étroite entre le milieu d'origine et le milieu nouvellement colonisé. C'est ce type de dispersion que l'on observe dans la recolonisation des milieux après les périodes glaciaires (e.g. François *et al.* 2008; García-Marín *et al.* 1999). Les individus peuvent également atteindre et coloniser de nouveaux milieux en se déplaçant via des couloirs qui connectent physiquement l'aire native et le milieu colonisé (Wilson *et al.* 2009). On peut citer en exemple l'ouverture du canal de Suez qui a permis le passage de la crevette tigrée *Marsupenaeus japonicus*, originaire de la Mer rouge, vers la Méditerranée (Galil 2007).

Si un nouvel habitat potentiellement favorable à l'espèce est éloigné de l'aire de répartition native, une dispersion à longue distance est nécessaire. Wilson *et al*. (2009) ont décrit deux grandes catégories de dispersions à longue distance se différenciant par la présence ou l'absence de flux géniques entre l'aire native et colonisée. Les dispersions à longue distance sont considérées comme des évènements rares et stochastiques, dont il est difficile de quantifier la fréquence et les distances moyennes (Nathan *et al*. 2003, Nathan *et al*. 2008).

I.1.2. Les déplacements naturels et ceux liés aux activités humaines

Si la dispersion naturelle active par la nage, la marche, ou le vol se fait le plus souvent sur de courtes distances, la dispersion passive via les vents ou les courants marins contribue au déplacement depuis les zones natives des espèces à la fois sur de courtes et sur de longues distances (Gillespie *et al.* 2012). La dispersion par le vent est un trait important pour un certain nombre d'insectes (Encadré 2). Les oiseaux constituent aussi un moyen de dispersion efficace pour les plantes et pour un certain nombre d'espèces d'arthropodes (Green & Figuerola 2005; Reynolds *et al.* 2015; Scott *et al.* 2001)

La dispersion en relation avec les activités humaines serait le principal vecteur d'introduction d'espèces non-natives dans de nouveaux milieux (Lockwood *et al.* 2013). En effet, en colonisant l'ensemble de la planète et en développant ses moyens de transports, les populations humaines ont contribué au transport de nombreuses espèces (Hulme *et al.* 2008;

Wilson *et al.* 2009). Ces dispersions à grande échelle spatiale peuvent être classées en introduction intentionnelle ou accidentelle.

L'introduction intentionnelle d'espèces non-natives, c'est-à-dire le déplacement d'organismes dans une nouvelle zone avec une finalité, est très ancienne et a largement accompagné le développement des civilisations (Wilson *et al.* 2009). Lors de ses migrations, les populations humaines ont prélevé de leur milieu d'origine des espèces animales ou végétales, notamment celles servant à l'alimentation, pour les amener avec elles. On peut citer le cas de la pomme de terre, *Solanum tuberosum*, originaire de la Cordillère des Andes aujourd'hui cultivée partout dans le monde (Spooner *et al.* 2005). Le développement de la lutte biologique contre des organismes nuisibles est également à l'origine d'introductions volontaires. Largement développée au cours du XX^{e} siècle, elle consiste à introduire un ennemi naturel d'un nuisible ou d'un ravageur pour réduire ses effectifs et donc les dommages causés par celui-ci (Hoddle 2004). La coccinelle asiatique aphidiphage, *Harmonia axyridis*, originaire de Chine a ainsi été importée en Amérique du Nord au début du 20ème siècle puis en Europe dans les années 80 afin de lutter contre les pucerons (Berkvens *et al.* 2010).

Par opposition, l'introduction accidentelle est le déplacement d'espèces non-natives de manière involontaire, sans finalité, et en lien avec les activités humaines (Hulme *et al.* 2008). Ce type d'introduction date des premières migrations humaines transcontinentales comme l'illustre la dispersion du rat du Pacifique, *Rattus exulans*, qui a suivi les mouvements humains lors de la colonisation de la Polynésie (Matisoo-Smith *et al.* 1998). Ce phénomène a connu une augmentation significative au cours des deux derniers siècles suite à l'intensification des échanges commerciaux et l'amélioration des modes de transport (Tatem & Hay 2007). Des exemples existent d'organismes dispersés à l'échelle intercontinentale via les eaux de ballasts des navires (Ruiz *et al.* 2000) ou les soutes d'avions (Tatem & Hay 2007).

I.1.3. Les facteurs contribuant au succès de transport

Certaines caractéristiques peuvent augmenter ou nuire au succès d'introduction au moment du transport. Les dissémination peuvent être influencées par le degré d'affinité des espèces avec les populations humaines. Ainsi, les espèces adaptées aux milieux anthropisés sont plus enclines à être prélevées et dispersées hors de leur aire d'origine (Hufbauer *et al.* 2012). A titre d'exemple, la distribution cosmopolite de certains rongeurs tel que le rat noir

Rattus rattus, est liée à leur affinité avec les milieux anthropisés. Originaire de la péninsule indienne, ce rongeur a profité des migrations humaines au cours des siècles pour coloniser l'ensemble des continents (Morand *et al.* 2015). Par ailleurs, l'abondance de l'espèce dans son aire d'origine (Hulme 2009), l'étendue de cette aire, et/ou sa localisation (par exemple, proche d'une plaque tournante de transport) (e.g. Tatem & Hay 2007) sont des facteurs contribuant au succès d'introduction. Enfin, une fois prélevées, toutes les espèces ne sont pas adaptées à un transport sur de longues distances ou de longue durée, dont les conditions abiotiques et biotiques (température, humidité, ressources alimentaires) sont souvent peu favorables. La capacité des œufs de certaines espèces de moustiques à supporter la dessiccation leurs permet de survivre pendant les phases de transport longues (Panov *et al.* 2004).

Une fois les barrières franchies et les migrants relâchés dans le nouvel environnement, les nouveaux arrivants doivent s'y acclimater et s'y installer.

I.2. La survie et l'installation dans le nouveau milieu

L'établissement correspond à l'étape pendant laquelle la population nouvellement arrivée doit franchir une série de barrières écologiques et reproductives pour survivre et s'établir durablement dans son nouvel environnement (Lockwood *et al.* 2013; Richardson *et al.* 2000). La densité de la population, les caractéristiques environnementales et le potentiel adaptatif de la population auront un rôle primordial dans le succès ou l'échec de l'invasion.

I.2.1. La pression de propagule

Durant la phase d'introduction, seul un faible pourcentage des individus prélevés dans l'aire d'origine survivront aux conditions de transport et seront relâchés dans le nouvel habitat. De nombreux auteurs estiment que la « pression de propagule », c'est-à-dire le nombre d'introductions (nombre de propagules) et le nombre d'individus par arrivée (taille de propagule) est le principal facteur influençant le succès global de l'invasion (Blackburn *et al.* 2015; Lockwood *et al.* 2005). La probabilité d'établissement d'une espèce dans un nouveau milieu est plus importante si elle est fondée par un grand nombre d'individus et si ces individus sont relâchés à plusieurs endroits différents et à plusieurs reprises (Blackburn *et al.* 2015).

Encadré 2. Dispersion des insectes par le vent

La capacité de dispersion active des petits insectes est généralement limitée par leur petite taille (de quelques centaines de mètres à quelques kilomètres). La dispersion par les vents peut constituer un moyen de dispersion à longue distance efficace, à la condition que ceux-ci présentent des caractéristiques favorables à la survie des organismes (température, vitesse) (Reynolds *et al.* 2006).

Si l'insecte qui se disperse par les vents est aussi vecteur de pathogènes, il peut alors disséminer le pathogène en même temps qu'il étend son aire de distribution (Reynolds *et al.* 2006). Les espèces du genre *Culicoides,* moucherons de la famille des Ceratopogonidés, peuvent ainsi être transportées passivement par les vents sur des distances allant jusqu'à plusieurs centaines de kilomètres au-dessus des masses d'eau, leurs permettant de franchir des barrières géographiques aussi larges que la Méditerranée (Braverman 1991; Sellers *et al.* 1977; Sellers *et al.* 1978).

La dispersion par le vent est souvent vue comme un évènement rare, avec des expansions géographiques qui se limiteraient à l'échelle régionale. Cependant, des auteurs montrent aussi que ces phénomènes pourraient survenir plus régulièrement, avec des fréquences variant fortement en fonction des saisons et des années (Parry *et al.* 2015). Pour certaines espèces d'insectes, ce mode de dispersion n'est pas entièrement passif et peut être initié ou entretenu par des phases de vol actif (Reynolds *et al.* 2006). Tous les types de vents ne sont pas favorables à la dispersion des organismes. Les vents à grandes vitesse (> 100 km/h) comme lors des cyclones ou des tempêtes sont plus à même de transporter des organismes de taille moyenne à grande, tandis que les vents et courants chauds à faible vitesse (20-30 km/h) seront plus favorables au transport des organismes de petite taille (Parry *et al.* 2015).

Ces phénomènes de dispersion atmosphérique risquent de s'accentuer en raison des changements intervenant dans le régime des vents suite aux variations climatiques (Chen *et al.* 2011). Les évènements de dispersion à des altitudes plus élevées seront probablement plus fréquents (Parry *et al.* 2015).

Cependant, au cours de leurs transports et de leurs installations, les populations peuvent être soumises à des goulots d'étranglement, conduisant à une réduction drastique de leurs tailles (Nei *et al.* 1975). Ces populations de petite taille seront plus vulnérables aux processus génétiques et démographiques, ce qui peut considérablement influencer la suite du processus d'invasion. Si la population subsiste sous ce faible effectif pendant plusieurs générations, la diversité génétique sera réduite par l'effet de la dérive génétique[2]. Par conséquent, la population introduite sera caractérisée par une diversité plus faible et des fréquences alléliques différentes de celles de l'aire native. La variance génétique additive[3] sera aussi plus faible (Lee 2002). Cette réduction de la diversité génétique et/ou les changements induits par la dérive vont avoir un effet négatif sur l'évolution adaptative et le succès d'établissement de l'espèce (Bock *et al.* 2015). D'un point de vue démographique, le nombre limité d'individus fondateurs peut impacter leur survie, une population de faible densité étant soumise à de plus grands risques d'extinction en raison de l'effet Allee[4] (Courchamp *et al.* 1999).

Toutefois, les conséquences associées au faible effectif des populations introduites peuvent être contrebalancées par le nombre d'introductions. En effet, la probabilité qu'une population s'établisse peut augmenter avec le nombre de fois et/ou le nombre d'endroits où les individus sont relâchés (Lockwood *et al.* 2009; Simberloff 2009). L'effet des introductions multiples dépend à la fois des caractéristiques temporelles de l'invasion, c'est-à-dire l'accumulation de matériel génétique au cours du temps par la population envahissante (Dlugosch & Parker 2008), mais aussi spatiale, l'introduction multiple à des endroits différents pouvant augmenter les chances que les individus relâchés rencontrent un environnement favorable (Duncan *et al.* 2014). D'un point de vue génétique, les introductions multiples peuvent avoir un effet de « secours génétique » (Carlson *et al.* 2014) enrichissant la diversité génétique particulièrement lorsqu'elles impliquent des populations sources différentes (Sakai *et al.* 2001). En effet, les individus issus de croisements entre les individus installés et les immigrants peuvent présenter un avantage sélectif. De plus, de nouvelles combinaisons génétiques absentes de l'aire native peuvent apparaitre suite à l'introduction récurrente d'allèles provenant de sources génétiquement différenciées et compenser la perte

[2] Variation aléatoire, d'une génération à l'autre des fréquences alléliques dans une population.
[3] Partie de la variance phénotypique pour un phénotype quantitatif qui est due aux effets additifs de tous les allèles en tous les loci.
[4] Diminution de la valeur sélective (définie ici comme le nombre moyen des descendants laissés à la génération suivante par un individu) d'un individu ou du taux de croissance d'une population suite à une réduction d'effectif ou de densité.

de variance génétique additive survenue lors des évènements fondateurs (Anderson & Stebbins 1954; Bock *et al.* 2015).

I.2.2. Les effets de l'environnement sur le succès de l'établissement

Les caractéristiques environnementales du milieu colonisé sont un facteur déterminant du succès de l'établissement. Pour s'établir, une espèce doit parvenir à survivre aux conditions abiotiques et biotiques de son nouveau milieu.

Des environnements spatialement éloignés peuvent posséder des conditions environnementales similaires (biomes). Dans ce cas, certaines espèces sont pré-adaptées à des milieux auxquels elles n'ont pas géographiquement accès, un changement dans leur régime de migration suffit alors pour qu'elles deviennent envahissantes (Facon *et al.* 2006). Cette pré-adaptation au nouveau milieu peut conférer un avantage leur permettant de s'établir à partir d'un nombre réduit d'individus fondateurs et en limitant les conséquences de l'effet Allee.

A l'inverse, un régime de migration peut permettre à une espèce d'être régulièrement introduite dans des environnements dont les conditions lui sont hostiles et dans lesquels elle ne parviendra pas à s'installer. Un changement des conditions environnementales dans le milieu de destination pourra alors permettre à l'espèce de s'établir et proliférer (Facon *et al.* 2006). Les changements environnementaux associés aux activités anthropiques comme la modification des écosystèmes conduisant souvent à une homogénéisation des territoires ainsi que le réchauffement climatique sont d'excellents exemples de modifications de l'environnement ayant permis l'établissement d'espèces envahissantes (e.g. Gray *et al.* 2009).

I.2.3. L'adaptation à l'environnement colonisé

Le succès des invasions repose également sur la capacité des espèces envahissantes à s'adapter aux pressions de sélection exercées par le nouvel environnement. Une augmentation du taux de croissance et de la capacité reproductive est fréquemment observée sur le terrain notamment chez les plantes (Pandit *et al.* 2014). Cette amélioration de la fécondité peut contribuer à une expansion démographique et géographique rapide dans le milieu colonisé (Bock *et al.* 2015).

Le succès de l'établissement peut également dépendre des nouvelles interactions biologiques. Ainsi, le relâchement de la pression de bioagresseurs («enemy release») peut permettre à la population introduite de réallouer les ressources utilisées pour sa défense vers

des capacités compétitrices, ses parasites et prédateurs habituels étant absents de l'aire colonisée (Keane & Crawley 2002). L'hybridation interspécifique dans le nouvel milieu peut elle aussi agir comme un stimulus évolutif (Abbott *et al.* 2003; Bock *et al.* 2015). Lorsque cette hybridation se produit entre une espèce introduite et une espèce locale (e.g. Ayres *et al.* 1999; Blair & Hufbauer 2010), elle peut fournir une base génétique grâce à laquelle certains gènes de l'espèce introduite s'exprimeront plus facilement (Lee 2002; Wares *et al.* 2005). Enfin, la plasticité phénotypique[5] peut également aider à surmonter les nouvelles conditions environnementales (Knop & Reusser 2012). Elle permet aux individus d'ajuster leurs traits d'histoire de vie aux nouvelles pressions du milieu. Les individus introduits peuvent posséder cette plasticité avant l'introduction, mais ce caractère peut également être sélectionné au sein de la population introduite au cours de l'établissement (Richards *et al.* 2006).

I.3. La prolifération démographique et l'expansion géographique

Lorsqu'une espèce s'établit avec succès dans un nouveau milieu, elle peut ensuite proliférer, se disperser géographiquement et établir d'autres populations viables dans d'autres nouveaux milieux. Cette étape de prolifération et d'expansion géographique contraint donc les individus qui se dispersent à franchir une nouvelle fois les barrières géographiques, reproductives et écologiques décrites précédemment.

Les capacités et les modes de dispersion ont un rôle fondamental dans le succès de cette étape. Selon Sakai *et al.* (2001), l'expansion géographique des populations établies est souvent une conséquence d'une excellente adaptation à la dispersion. Toutefois, bien que les capacités de dispersion soient déterminantes, l'adaptation à l'environnement biotique et abiotique du foyer d'introduction ne garantit pas le succès de la phase d'expansion si l'environnement est spatialement hétérogène. Dans le cas d'un environnement très hétérogène, l'espèce doit s'adapter aux différents milieux rencontrés lors de l'expansion. La plasticité phénotypique (Richards *et al.* 2006) ou un changement évolutif conduisant à l'adaptation génétique de l'espèce à son nouveau milieu ont un rôle important (Prentis *et al.* 2008).

Lors de l'étape de prolifération, une population est confrontée à un déséquilibre spatial caractérisé par des variations de densité et des effets de fondation (Sexton *et al.* 2009). Ceci peut influencer le succès de l'expansion et la structuration génétique de la population

[5] Capacité d'un génotype à développer des phénotypes différents en fonction des conditions environnementales

envahissante. La dérive génétique engendrée par les effets fondateurs entraîne une réduction de la diversité génétique sur le front pouvant entraîner une différenciation génétique entre le foyer et le front (Excoffier *et al.* 2009; Sexton *et al.* 2009). Une seconde conséquence des effets fondateurs est l'accumulation de mutations potentiellement délétères pouvant limiter le succès d'invasion (Peischl *et al.* 2013). Dans la plupart des cas, les nouvelles mutations apparues chez les individus au niveau du front d'expansion disparaissent ou se maintiennent à basse fréquence. Dans quelques cas, elles peuvent voir leur fréquence augmenter et se propager avec le front d'expansion, atteignant des fréquences élevées à distance de leur zone d'apparition (Excoffier *et al.* 2009; Klopfstein *et al.* 2006). Ce phénomène concerne aussi bien les mutations avantageuses que délétères.

Enfin, le déséquilibre spatial peut conduire à de l'homogamie spatiale qui concerne spécifiquement les capacités de dispersion et tous les traits associés à la dispersion (Shine *et al.* 2011). Lors d'une expansion spatiale, le front sera formé par les meilleurs dispersants qui s'accoupleront entre eux. Si les capacités de dispersion sont liées à un polymorphisme génétique, cette homogamie sur le front résultera en une descendance ayant en moyenne des capacités de dispersion plus élevées sur le front que dans le foyer d'introduction (Shine *et al.* 2011). Ce processus peut donc augmenter la vitesse d'expansion (Phillips *et al.* 2007).

II. L'étude des routes d'invasion

L'identification des populations sources et la description des routes d'invasion suivies et de la chronologie des évènements lors de la colonisation sont un premier pas vers une meilleure compréhension des processus écologiques et évolutifs des invasions biologiques (Estoup & Guillemaud 2010). Des méthodes directes et indirectes ont couramment été utilisées pour retracer les routes d'invasions.

Les méthodes directes reposent sur des observations de présence et d'absence des espèces envahissantes. Les routes d'invasion ont donc traditionnellement été reconstruites à partir de données historiques ou d'observations directes, telles que les dates et localités de premières observations. Par exemple, les routes d'invasions du moustique tigre asiatique *Aedes albobictus* ont été retracées à partir de données de circulation aérienne et maritime en relation avec les données climatiques (Tatem *et al.* 2006). Cependant, ces méthodes ne sont pas exhaustives, et sont souvent basées sur des données rares et incomplètes (Estoup &

Guillemaud 2010), rendant difficile la description rigoureuse des routes empruntées par les populations envahissantes. De plus, les données de présence ne reflètent pas forcément le succès d'établissement des individus.

Les approches moléculaires bien qu'indirectes offrent une alternative pour décrire et étudier les populations sources, les routes et les mécanismes d'expansion (Handley *et al.* 2011). Ces méthodes basées sur les patrons génétiques observés entre et au sein des populations permettent d'inférer des liens de parentés entre populations. Traditionnellement ces méthodes sont basées sur des dendogrammes de distances génétiques inter-populationnelles, des statistiques résumées de la diversité (F_{ST}) ainsi que des méthodes de regroupements d'individus ou de populations. Ces approches ne permettent cependant pas de retracer des scenarios complexes en raison de l'effet du hasard (stochasticité) sur les paramètres démographiques et historiques (par exemple sur les effets fondateurs et sur la dérive génétique) et de l'existence de sources multiples et d'évènements d'admixture. La dérive génétique peut par exemple fortement accroitre la différentiation génétique entre une population envahissante et sa source en modifiant aléatoirement les fréquences alléliques (Knowles 2009). Ainsi, une seule population source peut donner naissance à une infinité de populations très différenciées suite à l'échantillonnage aléatoire des allèles dans l'aire native. Ces phénomènes peuvent alors considérablement brouiller les pistes et rendre difficile l'identification des populations sources et des routes de colonisation.

L'application aux données moléculaires des nouvelles approches d'*Approximate Bayesian Computation* (ABC) offre maintenant l'opportunité de tester des scénarios complexes et d'estimer des paramètres démographiques sous un modèle d'évolution donné (Beaumont *et al.* 2002; Estoup & Guillemaud 2010) (Encadré 3). L'ABC est une méthode bayésienne qui consiste à faire une approximation des vraisemblances de scénarios évolutifs via la simulation d'un grand nombre de jeux de données. Plus précisément, les probabilités postérieures de différents modèles et/ou les distributions postérieures des paramètres démographiques générés sous un modèle donné sont déterminées en comparant des statistiques calculées à partir des données réelles obtenues à partir des populations étudiées, à celles des données simulées (Beaumont 2010; Bertorelle *et al.* 2010; Csillery *et al.* 2010; Lopes & Beaumont 2010). Cette méthode permet d'estimer des paramètres historiques (temps de divergence), évolutifs (taux de mutation, influence de la sélection, hybridation) et démographiques (taille efficace de la population, goulot d'étranglement) à l'aide de données génétiques, et ceci même dans les cas où l'ensemble des populations source n'auraient pas été

échantillonnées (populations fantômes). Les analyses réalisées avec la méthode ABC constituent des outils puissants et efficaces dans le cadre de l'inférence des routes d'invasions de nombreuses espèces (e.g. Brouat *et al.* 2014; Lombaert *et al.* 2011; Miller *et al.* 2005; Pascual *et al.* 2007).

Cependant, la méthode ABC connait aussi un certain nombre de limites. Le choix de la nature des statistiques résumées et de leurs nombres utilisés pour répondre à une question donnée est une tâche délicate pour laquelle il n'existe pas de règle. Celles-ci doivent être choisies de manière à représenter de manière pertinente les caractéristiques des données, être en nombre suffisant, mais non redondantes. Des solutions pour aider au choix de ces statistiques ont été proposées par différents auteurs (Blum & François 2010; Joyce & Marjoram 2008; Wegmann *et al.* 2009). De plus, la nature bayésienne contraint à fixer des distributions de paramètres *a priori* dans un intervalle qui doit être suffisamment large pour inclure toutes les valeurs jugées probables. Enfin, seul un nombre limité de modèles pourra être exploré (Templeton 2010). Il est donc nécessaire de procéder avec prudence et d'avancer pas à pas au cours d'une analyse basée sur les méthodes ABC. L'utilisation de données historiques et des méthodes indirectes mentionnées précédemment peuvent augmenter l'efficacité de la méthode (Estoup & Guillemaud 2010).

Outre la description des routes d'invasions, l'utilisation des données moléculaires a permis des avancées dans la compréhension des processus gouvernant les invasions. Elles ont ainsi révélé que les processus d'invasions impliquaient fréquemment des introductions multiples, ce qui expliquait des niveaux de diversité similaires voire plus élevés dans l'aire envahie par rapport à l'aire native (Dlugosch & Parker 2008). Ceci est aussi avancé pour expliquer le succès paradoxal des populations envahissantes à surmonter les effets fondateurs associés à la colonisation (e.g. Roman & Darling 2007). Un autre phénomène, l'effet « tête de pont », a été mis en évidence : une population établie et envahissante sert à son tour de population source lors d'invasions secondaires de territoires potentiellement isolés (e.g. Lombaert *et al.* 2010). D'un point de vue évolutif, un unique changement évolutif dans la population introduite (la population « tête de pont ») est requis, alors que plusieurs changements sont nécessaires dans le cas d'introductions multiples (Estoup & Guillemaud 2010).

D'un point de vue général, l'étude des processus d'invasions biologiques nous apportent de nombreux enseignements sur la dynamique des processus évolutifs et écologiques sur des

échelles de temps très courtes (Lee 2002; Sax *et al.* 2007). Elle fournit ainsi des informations sur la vitesse d'adaptation et le rôle des goulots d'étranglement dans l'évolution des espèces (Bock *et al.* 2015). De plus, d'un point de vue plus appliqué, décrire les populations sources et les routes de colonisation permet de renforcer la vigilance face aux espèces envahissantes, via par exemple des suivis spécifiques des populations sources identifiées ou des populations « tête de pont » (Estoup & Guillemaud 2010). D'autre part, les connaissances sur les populations sources peuvent aider à définir les caractéristiques écologiques des populations envahissantes et par conséquent, prédire l'étendue de la distribution de l'espèce dans le nouvel environnement (Kolar & Lodge 2001).

Encadré 3. Inférence des routes d'invasion par la méthode ABC

Dans cet encadré, nous ne détaillerons pas les caractéristiques statistiques de l'approche ABC qui ont été détaillées dans des revues récentes (Beaumont 2010; Bertorelle *et al.* 2010; Csillery *et al.* 2010; Lopes & Beaumont 2010). Nous présenterons, en revanche, le principe et les grandes étapes de l'analyse basée sur la méthode ABC.

Le principe général

Grâce aux statistiques bayésiennes, il est possible de construire des modèles et/ou d'estimer des paramètres sous un modèle donné à partir de données observées telles que des génotypes multi-locus (Beaumont & Rannala 2004). Par conséquent, il est possible de déterminer la distribution *a posteriori* d'un modèle ou du paramètre θ en connaissant les données observées D en suivant l'équation suivante : $P(\theta|D) \propto P(D|\theta)P(\theta)$ où $P(\theta)$ est la distribution *a priori* du modèle ou du paramètre.

La probabilité $P(D|\theta)$ d'observer les données D étant donné le paramètre (ou le modèle) θ correspond à la vraisemblance de θ. Cette vraisemblance peut être estimée, mais lorsque les modèles sont très complexes, cette estimation devient difficile voire impossible. Dans ce cas, des méthodes sans calcul de la vraisemblance (*likelihood free*) permettent de calculer une distribution *a posteriori* en remplaçant la vraisemblance par une approximation en utilisant les statistiques résumées sur des jeux de données simulés selon différents modèles. La méthode ABC est une de ces méthodes particulièrement bien adaptée à l'analyse de données génétiques (Beaumont 2010; Beaumont *et al.* 2002; Bertorelle *et al.* 2010; Csillery *et al.* 2010; Lopes & Beaumont 2010).

Encadré 3. Suite

Les étapes de l'analyse

Le déroulement d'une analyse basée sur les méthodes ABC peut se décomposer en 5 étapes. A chacune de ces étapes, des phases de validations sont effectuées avant de passer à l'étape suivante :

Etape 1 : Définition des scénarios et *priors*. Il s'agit d'établir des scénarios d'introduction et d'associer des valeurs ou des distributions *a priori* à chacun des paramètres génétiques (taux de mutation des marqueurs), démographiques (tailles efficaces des populations), historiques (dates de première observation de chacune des populations envahissantes) et à la fréquence des modèles.

Etape 2 : Choix des statistiques résumées et simulations de données génétiques. Il s'agit de simuler des données génétiques selon chacun des scénarios à partir d'un modèle stochastique de mutation et de dérive liant démographie et génétique. Les paramètres sont tirés aléatoirement dans la distribution des *priors*. Chacun des jeux de données est ensuite résumé à l'aide des statistiques précédemment choisies pour décrire les variations génétiques intra et inter populationnelles (nombre moyen d'allèles par locus, hétérozygotie attendue, F_{ST} par paire de populations).

Etape 3 : Rejet des jeux de données les moins informatifs. Les distances euclidiennes entre les statistiques simulées et observées sont calculées, puis les simulations les plus éloignées des observations, au-delà d'un certain seuil, sont rejetées.

Etape 4 : Sélection du scénario le plus probable. Le calcul de la probabilité *a posteriori* de chaque scénario par une régression logistique sur les jeux de données simulés conservés lors de l'étape 3 va permettre d'identifier le scénario. Il est ensuite possible d'estimer des distributions *a posteriori* des paramètres du scénario gagnant.

Etape 5 : Evaluation de la puissance d'analyse et de la pertinence des inférences. Il important (i) de vérifier la puissance de l'analyse effectuée en calculant les erreurs de type I et les erreurs de type II à l'aide de données simulées, utilisées comme des données qui auraient été observées (jeu de données pseudo-observées), pour déterminer si l'analyse permet de bien distinguer les différents scénarios et (ii) de contrôler la concordance entre le scénario sélectionné et les données observées en simulant des données à partir de valeurs de paramètres tirées dans les distributions *a posteriori* (« model checking ») qui permet de déterminer si le scénario sélectionné et les distributions *a posteriori* des paramètres inférées reproduisent convenablement les données observées. En cas d'incohérence, il peut être nécessaire de repasser à l'étape 1.

III. Colonisation et maladies émergentes : exemple de *Culicoides imicola*

Avec près d'un million d'espèces décrites (Hill 2012), les insectes dominent la faune mondiale et occupent toutes les niches des habitats terrestres. Une grande majorité d'entre eux sont essentiels aux fonctionnements des écosystèmes. Les insectes jouent un rôle écologique majeur en tant que pollinisateurs, décomposeurs et prédateurs, et interviennent dans le recyclage des matières organiques, l'élimination des déchets et la bonne santé des sols (Hill 2012). Ils sont aussi une ressource alimentaire pour de nombreux animaux, et parfois pour certaines populations humaines. De plus, certaines espèces sont exploitées par l'homme (abeille pour le miel ou la cire, vers à soie pour la soie) et sont à classer comme des productions animales.

D'autres espèces sont considérées comme nuisibles pour les activités humaines et les écosystèmes. Certaines sont des ravageurs de cultures, comme le criquet pèlerin, *Schistocerca gregaria*, responsable de pertes économiques estimées à plus de 400 millions de dollars (Hill 2012; Lomer *et al.* 2001). D'autres sont pathogènes pour les vertébrés par leurs effets allergisants, urticants, vésicants ou venimeux ; on peut citer les hyménoptères aculéates (abeilles, guêpes, frelons) et certains papillons nocturnes. Enfin, parmi les insectes nuisibles pour les vertébrés, certains sont hématophages et peuvent avoir un impact direct et/ou indirect. Les impacts directs sont liés aux piqûres ou morsures infligés aux hôtes lors des repas de sang. Ces piqûres peuvent être extrêmement douloureuses et provoquer des allergies sévères à cause de la salive inflammatoire et immunogène des insectes. En Ecosse, les piqûres du moucheron *Culicoides impunctatus* Goetghebuer, moucheron inféodé aux zones humides et aux tourbières de la région paléarctique, sont si nombreuses qu'une protection est indispensable pour toute activité à l'extérieur limitant les industries agroforestières pendant le pic d'abondance des populations (Hendry & Godwin 1988). Si le rôle nuisible des insectes n'est pas à négliger, il reste moindre par rapport aux impacts sanitaires et économiques liés à la transmission de pathogènes par les insectes vecteurs.

Les maladies à transmission vectorielle constituent un problème majeur en santé publique et vétérinaire. En plus de leur importance sanitaire, elles ont un coût économique significatif pour la société et sont un frein au développement des pays où elles sont en forte prévalence.

Ce n'est qu'au XIX^{e} siècle qu'il est démontré que certains insectes sont impliqués dans la transmission de pathogènes (Chernin 1983). Aujourd'hui, on distingue deux grands groupes de vecteurs selon le mode de transmission : transmission mécanique et transmission biologique. Lors d'une transmission mécanique, le vecteur se contente simplement de transporter l'agent infectieux sans qu'il y ait une multiplication obligatoire du pathogène (Rhodain & Perez 1985). C'est le cas des mouches qui peuvent transporter des agents infectieux présents à l'extérieur de leurs corps et les transmettre par contact physique à un nouvel endroit. Lors de transmission biologique, une partie du cycle de développement du pathogène a obligatoirement lieu à l'intérieur de l'insecte vecteur (Rhodain & Perez 1985). La transmission est active lorsque le vecteur établit de lui-même le contact entre l'agent infectieux et l'hôte vertébré. Les vecteurs biologiques actifs regroupent essentiellement les insectes et les tiques, et transmettent une grande diversité de pathogènes incluant des virus, des rickettsies, des bactéries, des protozoaires et des helminthes. Dans le reste de ce manuscrit, nous considèrerons la définition retenue par l'expertise collégiale sur la lutte anti-vectorielle qui définit un vecteur comme : « *tout arthropode hématophage qui assure la transmission biologique active d'un agent pathogène d'un vertébré à un autre vertébré* » (Fontenille *et al.* 2009).

Depuis une trentaine d'années, on assiste à l'émergence ou à la réémergence de maladies à transmission vectorielle à l'échelle mondiale (Kampen & Werner 2011; Kilpatrick & Randolph 2012). Ceci s'explique principalement par une augmentation des déplacements des hôtes et par des phénomènes d'invasions biologiques d'arthropodes vecteurs. Dans certains cas, l'émergence des maladies à transmission vectorielle résulte de l'arrivée d'un hôte vertébré infecté dans un nouvel environnement. Un exemple marquant est l'introduction du virus de la fièvre du Nil occidental par des oiseaux migrateurs en 1999 dans le district de New York – maladie transmise par des moustiques, principalement du genre *Culex*, et impliquant des hôtes aviaires dans son cycle sauvage et des hôtes équins et humains dans son cycle péri-domestique ou domestique – qui s'est rapidement adapté aux oiseaux et aux populations locales de moustiques, se propageant en moins de quatre ans dans l'ensemble des Etats-Unis et affectant plus de 1.8 millions de personnes (Bowman 2014). Dans d'autre cas, l'émergence est liée à l'introduction puis l'installation d'une espèce vectrice non-native compétente pour des pathogènes sur de nouveaux territoires avec des hôtes sensibles (Mazza *et al.* 2014). Si un pathogène transmis par cette espèce est introduit par un hôte infecté, alors les conditions pour la transmission du pathogène et l'émergence de la maladie sont réunies. L'invasion mondiale

du moustique tigre asiatique, *Aedes albopictus*, durant les trois dernières décennies, a ainsi propagé les virus de la dengue et du chikungunya dans des territoires initialement indemnes, causant leur émergence et des millions d'infections (Bonizzoni *et al.* 2013).

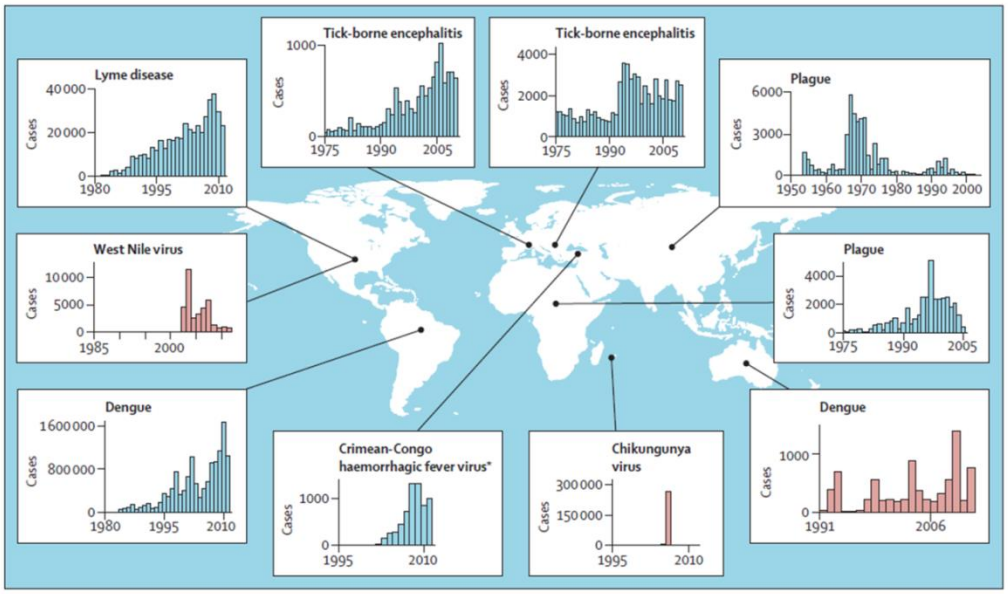


Figure 5. Dynamique épidémiologique de pathogènes à transmission vectorielle introduits (rouge) et pour des maladies endémiques ou établies depuis longtemps dans une zone (bleu) (Kilpatrick & Randolph 2012).

L'augmentation de l'incidence des maladies à transmission vectorielle peut aussi s'expliquer par certains facteurs socio-économiques (arrêt des réseaux de surveillance ou des actions de lutte, sensibilisation moindre des populations, augmentation des contacts hôte/vecteur), des changements évolutifs (modifications génétiques des agents pathogènes), ou environnementaux (anthropisation du milieu, variations climatiques) (Kampen & Werner 2011). Particulièrement, l'augmentation des températures à la surface du globe a des conséquences sur la distribution des espèces d'arthropodes et sur l'épidémiologie des maladies à transmission vectorielle. Les arthropodes vecteurs sont des organismes ectothermes et donc leurs cycles de développement, leurs reproductions, leurs comportements et leurs dynamiques de populations sont sujets aux fluctuations climatiques (Gage *et al.* 2008). De plus, des variations de température peuvent affecter l'interaction entre le pathogène et son vecteur (taux de multiplication du pathogène) et modifier la dynamique de la maladie ou de sa transmission (Baylis & Githeko 2006; Kovats *et al.* 2001). Toutefois, quantifier l'effet du changement climatique sur le niveau de transmission d'une maladie est une tâche difficile (Kovats *et al.* 2001).

L'émergence de la fièvre catarrhale ovine (FCO) en Europe méditerranéenne est souvent citée comme un exemple du rôle du changement climatique dans l'expansion géographique récente du vecteur afrotropicale du virus de la fièvre catarrhale ovine, *Culicoides imicola.*

III.1. Quelques généralités sur les *Culicoides*

Les *Culicoides* (Diptera: Ceratopogonidae) sont de petits moucherons nématocères (entre 1 et 3 mm) (Figure 6), dont la quasi-totalité des espèces connues sont hématophages. Environ 1 300 espèces de *Culicoides* sont décrites par le monde. Ils sont présents des tropiques à la toundra et du niveau de la mer à 4 200 m d'altitude (au Tibet). Les *Culicoides* sont des organismes holométaboles qui passent par quatre stades larvaires, suivi d'un stade nymphal, et d'une phase adulte (Figure 7). Les stades immatures se développent dans des milieux humides et riches en matières organiques (fèces de ruminants, trous d'arbres, bords de mares). En général, les adultes ont une durée de vie courte qui varie entre 10 et 20 jours, mais celle-ci peut atteindre 44 à 90 jours (Mellor *et al.* 2000).

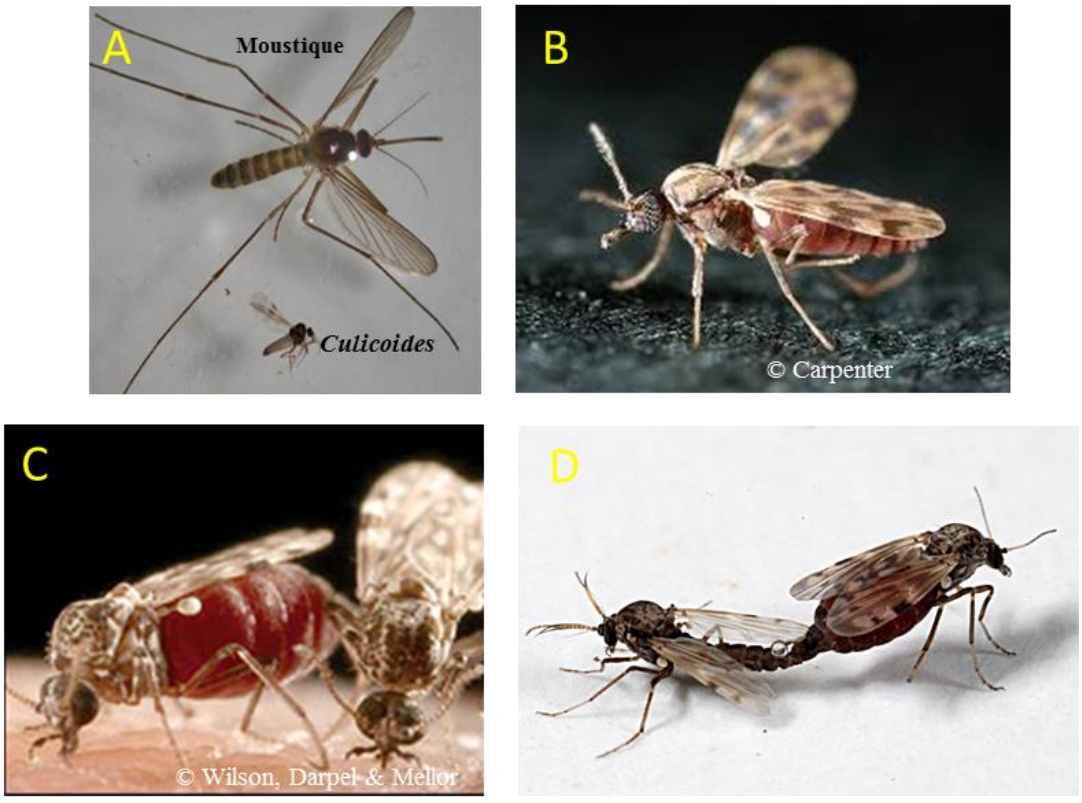


Figure 6. Photos de *Culicoides*. (A) Taille d'un *Culicoide* par rapport à un moustique, (B) *C. imicola*, (C) Prise du repas de sang, (D) Accouplement.

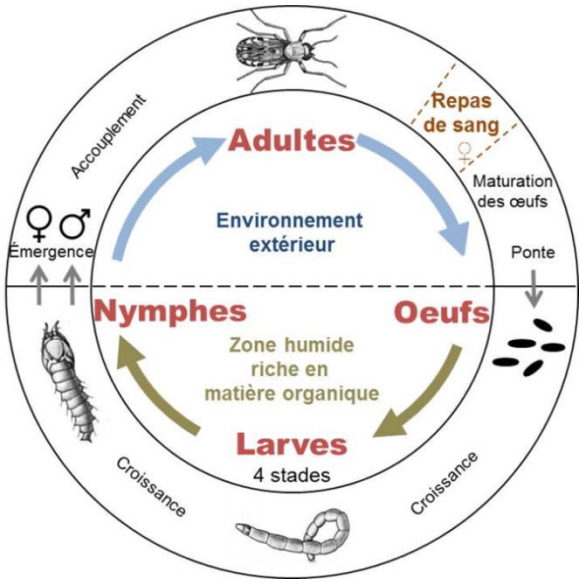


Figure 7. Représentation schématique du cycle de développement des *Culicoides* (Source Thèse Roger Venail 2014).

III.1.1. La dispersion, un trait d'histoire de vie important

La plupart des espèces de *Culicoides* sont crépusculaires ou nocturnes, et seules les femelles sont hématophages (Mellor *et al.* 2000). Les femelles se déplacent activement à différentes étapes de leur cycle de développement pour l'accouplement, la recherche d'un hôte pour le repas de sang, la digestion du repas de sang et la ponte (Sellers 1992). Du fait de leur petite taille, la dispersion active par le vol est décrite dans la littérature comme faible et limitée à quelques centaines de mètres. La plus grande distance parcourue par dispersion active a été estimée à 6 km par jour chez *Culicoides mohave* Wirth dans un désert aux Etats-Unis avec une distribution des hôtes très agrégée (Brenner *et al.* 1984). Récemment, malgré les difficultés méthodologiques (forte mortalité des individus capturés vivants), des études de capture/marquage/recapture ont permis de quantifier les distances parcourues par les *Culicoides* en région paléarctique (Kirkeby *et al.* 2013; Kluiters *et al.* 2015; Sanders & Carpenter 2014). En utilisant du marquage fluorescent, les différents auteurs montrent que les espèces paléarctiques ont des vols actifs omnidirectionnels avec des distances moyennes journalières entre 1.75 et 2.1 km (Kirkeby *et al.* 2013; Kluiters *et al.* 2015). En revanche, la dispersion passive par les vents (Encadré 2) est considérée comme importante dans le cycle biologique des *Culicoides*, et peut atteindre plusieurs centaines de kilomètres au-dessus des masses d'eau (Braverman 1991; Sellers *et al.* 1977).

III.1.2. Le rôle nuisant et vecteur des espèces de *Culicoides*

Les *Culicoides* peuvent attaquer des mammifères, des oiseaux, des reptiles ou même d'autres insectes. Dans certaines zones du monde, leurs attaques peuvent être si intenses, qu'elles causent de fortes nuisances pour les hommes – les *Culicoides* pouvant représenter un frein au tourisme ou aux activités de foresterie – et les animaux, provoquant des affections allergiques, comme la dermatite estivale récidivante chez le cheval.

Outre leur caractère nuisant, un nombre limité d'espèces, environ 60 espèces, est décrit comme impliqué dans la transmission de protozoaires sanguins, de filaires ou de virus d'importance pour la santé humaine ou animale (Mellor 2000). Parmi les virus d'intérêt vétérinaire, on compte les virus de la fièvre hémorragique épizootique (EHD) (qui affecte les ruminants domestiques et sauvages), de Schmallenberg (SBV) (qui affecte les ruminants domestiques), de la fièvre catarrhale ovine (BTV) et le virus d'Akabane (AKAV) (qui affecte les bovins et les ovins) et enfin de la peste équine (AHSV) (qui affecte les équins) (Mellor et al. 2000). Le seul virus décrit d'intérêt médical est le virus Oropouche (OROV), distribué en Amérique du Sud (Figure 8). L'importance des *Culicoides* en santé animale tient surtout à leur capacité à transmettre les virus de la fièvre catarrhale ovine et de la peste équine, deux maladies réglementées, responsable d'épizooties avec des conséquences sanitaires et économiques majeures (Diouf *et al.* 2013; Velthuis *et al.* 2010; Velthuis *et al.* 2011).

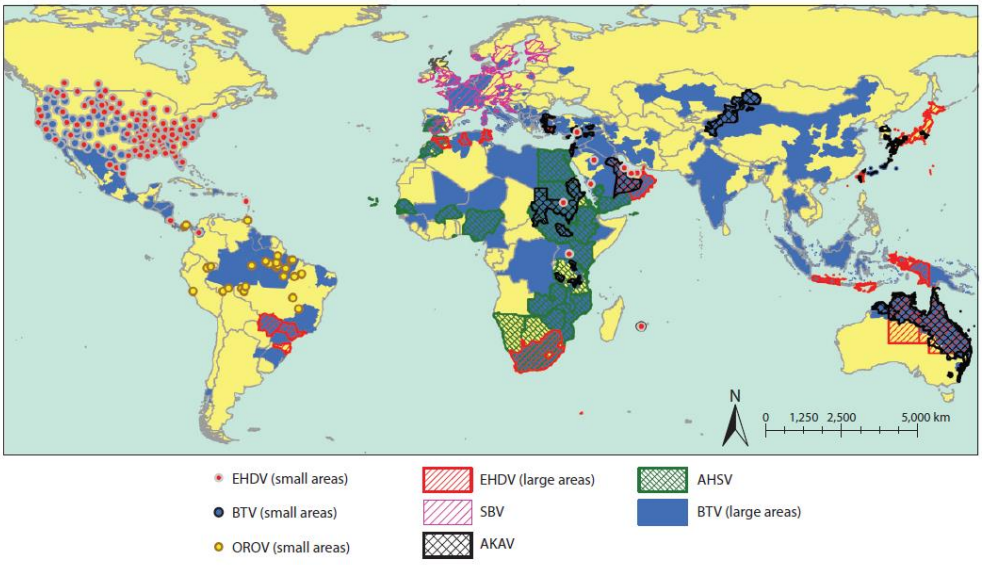


Figure 8. Distribution géographique des six virus principaux transmis par les *Culicoides* (Purse *et al.* 2015).

La peste équine est une maladie hémorragique des équidés pour laquelle le cheval présente une létalité très importante, pouvant atteindre 90 % des individus infectés. L'agent infectieux est un *Orbivirus* de la famille des Reoviridae, pour lequel neuf sérotypes ont été décrits. Le berceau d'origine de la maladie est le continent africain, où elle est restée cantonnée pendant longtemps, avec récemment des sorties de sa zone d'enzootie et a provoquant des flambées épizootiques meurtrières. En 1965, des foyers de peste équine apparaissent au Maroc (Diaz Montilla & Panos Marti 1967, 1968; Laaberki 1969) puis la maladie s'étend à l'Algérie et à la Tunisie, et traverse le détroit de Gibraltar en 1966. L'épizootie, provoquée par le sérotype 9, a été rapidement jugulée grâce aux mesures de vaccination et de police sanitaire. L'Europe est restée indemne jusqu'en 1987, où un foyer causé par le sérotype 4 a été confirmé dans la province de Madrid, suite à l'importation de zèbres en provenance de Namibie et destinés au zoo de la ville (Lubroth 1988). Malgré les mesures d'abattage et de vaccination (38 000 équidés vaccinés), de nombreux cas de peste équine furent observés l'année suivante dans le sud de l'Espagne, en Andalousie. En 1989, la peste équine traversa la frontière portugaise (Mellor *et al.* 1990) et le détroit de Gibraltar vers le Maghreb. En région afrotropicale, des infections expérimentales de *C. imicola* et *C. bolitinos* ont confirmé le rôle majeur de ces espèces dans la transmission du virus, mais d'autres *Culicoides* pourraient également être vecteurs (Mellor *et al.* 2000).

La fièvre catarrhale ovine est une maladie des ruminants sauvages et domestiques (ovins, bovins et caprins). Les races améliorées sont beaucoup plus sensibles que les espèces rustiques. Le virus présente 26 sérotypes, dont le nombre évolue avec les capacités de détection et l'effort de surveillance. Chez les ovins, les principaux symptômes sont une forte hyperthermie, des lésions et des érosions buccales, un œdème de la face, des écoulements nasaux, une raideur des membres, des boiteries, une fonte musculaire importante, et des avortements (Sperlova & Zendulkova 2011) (Figure 9). La morbidité peut atteindre 80% du troupeau. La mortalité est de 5 à 10%, et jusqu'à 40% en cas de mauvaise condition des animaux ; elle survient 10 à 12 jours suivant le début de la maladie. Les pertes économiques sont liées à la mortalité dans les cheptels, au déclassement des carcasses, à la mauvaise qualité de la laine, à la baisse de production laitière, à la mise en place des mesures de contrôle (vaccination, tests diagnostiques et visites vétérinaires), aux restrictions des mouvements d'animaux et à l'arrêt des exportations. La fièvre catarrhale ovine est largement répandue dans le monde, avec des couples espèces vectrices/sérotypes circulants spécifiques selon les régions : on parle de

pathosystèmes. En région afrotropicale *C. imicola* est décrit comme le vecteur principal (Venter *et al.* 2011).

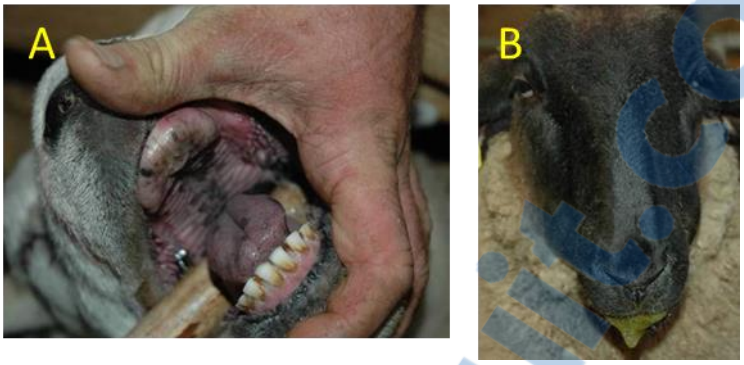


Figure 9. Illustration des symptômes de la fièvre catarrhale ovine. (A) Lésions et hémorragies buccales, congestion de la langue. (B) Jetage spumeux.

III.2. *Culicoides imicola* et le virus de la FCO dans le bassin méditerranéen

Culicoides imicola, est une espèce afrotropicale originellement distribuée à travers l'Afrique sub-saharienne, le Moyen-Orient et dans une partie de l'Asie (Inde, sud de la Chine) (Boorman 1986; Boorman & Wilkinson 1983; Braverman & Galun 1973; Dyce & Wirth 1983; Yü 2005). Il est à noter que la distribution orientale de l'espèce n'a pas été validée par des expertises récentes.

Avant 1998, la FCO était considérée comme enzootique dans la ceinture intertropicale avec quelques incursions épizootiques dans le bassin méditerranéen : à Chypre et en Israël (1951), en Espagne et au Portugal (1957-1960) et sur l'île grecque de Lesbos (Campano Lopez & Sanchez Botija 1958; Gambles 1949; Sellers 1975; Sellers *et al.* 1978; Taylor *et al.* 1985). Cependant, la présence de *C. imicola* n'est confirmée que plus tard, dans les années 70, à l'est du bassin méditerranéen, à savoir en Israël, en Turquie et sur les îles grecques Lesbos, Rhodes et Chios (Boorman 1986; Boorman & Wilkinson 1983; Braverman & Galun 1973). A la même époque, à l'ouest du bassin méditerranéen, *C. imicola* est décrit au Maroc et en Algérie (Bailly-Choumara & Kremer 1970). Une décennie plus tard, l'espèce vectrice est observée dans la péninsule ibérique où elle a d'abord été collectée en Espagne en 1983 (Mellor *et al.* 1983), puis au Portugal en 1985 (Mellor *et al.* 1985). A cette époque, la latitude 40 °N est admise comme la limite nord de

la distribution de *C. imicola*, étant donné que des enquêtes entomologiques ne l'avaient pas enregistré en dehors de cette limite (Mellor & Wittmann 2002).

A partir de 1998, l'Europe du sud subit une série inédite d'émergences de la FCO impliquant différents sérotypes (1, 2, 4, 8, 9, 16) majoritairement dans des zones où *C. imicola* était considéré comme absent (Mellor & Wittmann 2002). Les premiers foyers sont observés sur plusieurs îles grecques (Rhodes, Leros, Kos, Samos), avant d'atteindre progressivement et rapidement toutes les régions du bassin méditerranéen. En 1999, des foyers sont décrits au Maghreb, puis entre 2000 et 2004, en Espagne, en Grèce continentale, en Sardaigne et en France (Corse) (Mellor & Wittmann 2002) (Figure 10).

Suite aux foyers de FCO, des enquêtes entomologiques sont menées dans les régions touchées par la maladie afin de mettre en évidence la présence de *C. imicola*. Elles permettent rapidement de dresser la carte de distribution de l'espèce dans le bassin méditerranéen. En 2000, l'espèce vectrice est ainsi observée pour la première fois sur les îles grecques de Samos, Kos et Leros, et dans plusieurs localités en Grèce continentale (Patakakis 2004; Patakakis *et al.* 2009), mais semble absente au nord de la Grèce et en Bulgarie (Mellor & Wittmann 2002). A la même période, *C. imicola* est collecté en Italie (partie continentale, Sicile et Sardaigne). Les enquêtes entomologiques indiquent une limite septentrionale au niveau de la latitude 44 °N (au nord de la Toscane), avec une présence relativement rare au-delà de cette limite (Goffredo *et al.* 2004). En 2002, la présence de *C. imicola* est confirmée en Corse (Delécolle & De La Rocque 2002) où les populations se révèlent largement répandues et très abondantes (avec plus de 10,000 insectes collectés par nuit) (Venail *et al.* 2012). Suivent des enquêtes entomologiques qui confirment la présence de l'espèce dans les îles Baléares (Miranda *et al.* 2003), dans le nord de l'Espagne et en Catalogne (Monteys *et al.* 2005) (Figure 10).

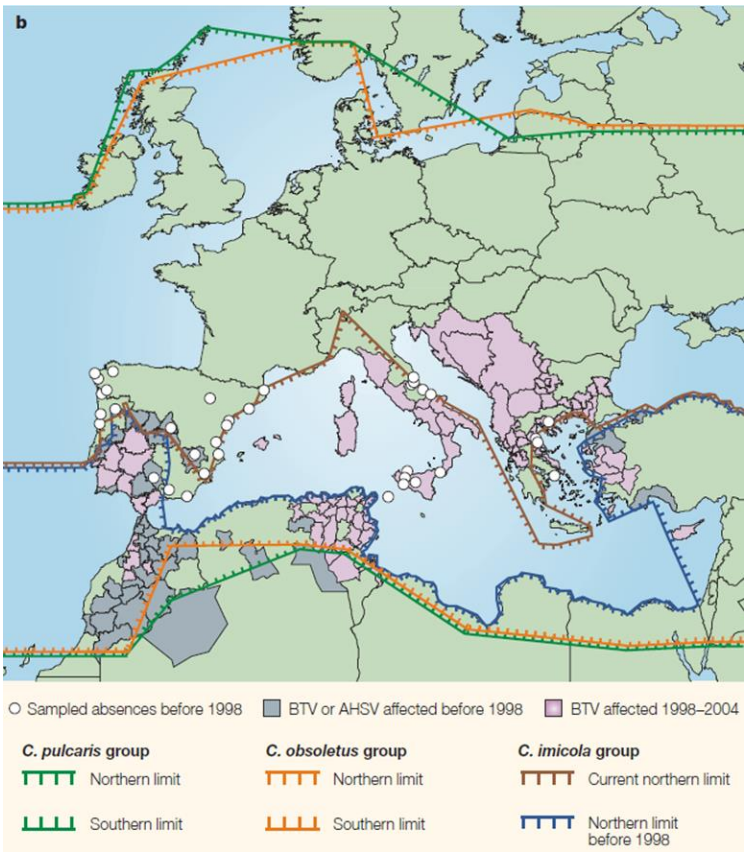


Figure 10. Distribution des sérotypes de la FCO entre 1998 et 2004 et répartition des populations de *Culicoides* vecteurs en Europe. La limite septentrionale de la distribution des populations de *C. imicola*, jusqu'en 2005, est représentée en marron.

En France continentale, l'espèce est collectée pour la première fois dans le département du Var en 2003 avec des abondances très faibles par rapport aux populations corses (Venail *et al.* 2012). La surveillance entomologique mise en place sur le littoral méditerranéen suite à cette introduction et jusqu'à aujourd'hui, montre une présence sur une zone limitée (Venail *et al.* 2012). La zone concernée est à cheval sur les départements du Var et des Alpes-Maritimes, plus précisément dans les plaines littorales des golfes de Fréjus et de Saint-Tropez. Une deuxième introduction est reportée en 2008 dans le département des Pyrénées-Orientales où les enquêtes entomologiques montrent comme dans le Var une distribution et des abondances limitées. Depuis 2005, la population présente dans le département du Var représente la population la plus au nord de la distribution de l'espèce.

III.2.1. Culicoides imicola *: une espèce envahissante ?*

La crise sanitaire suite à l'émergence de la FCO et la découverte des populations de *C. imicola* dans le bassin méditerranéen ont suscité un vif intérêt de la communauté scientifique pour comprendre les facteurs à l'origine du changement épidémiologique dans le système *C. imicola*/FCO. L'hypothèse la plus communément admise est une expansion géographique récente de *C. imicola* dans les écosystèmes méditerranéens depuis la zone historique de distribution. Cette remontée se serait accompagnée de la circulation du virus dans ces nouvelles zones grâce aux individus infectés. Plus précisément, les travaux de modélisation de Purse *et al.* (2005) suggèrent que cette expansion daterait de la fin du XX^{e} siècle et serait une conséquence du réchauffement climatique contemporain : l'augmentation des températures, les épizooties de FCO et les confirmations de la présence de *C. imicola* étant concomitants dans le temps et l'espace. L'augmentation globale des températures aurait permis l'installation et l'augmentation d'abondance des populations du vecteur dans de nouvelles zones (Purse *et al.* 2005). Parallèlement, Guis *et al.* (2012) ont examiné les effets du changement climatique sur la transmission du virus. Les auteurs confirment l'importance du climat pour expliquer l'émergence et la circulation du virus. L'augmentation des températures sur la période 1960-1999 aurait induit une augmentation de la densité des vecteurs en Europe méditerranéenne (Guis *et al.* 2012).

Pour tester cette hypothèse de remontée récente de *C. imicola* vers le Nord, trois études de phylogéographie, basées sur un fragment du gène mitochondrial cytochrome oxydase I (COI), se sont intéressées aux routes de colonisation (Calvo *et al.* 2009; Dallas *et al.* 2003; Nolan *et al.* 2008). Ces études ont toutes mis en évidence une différenciation génétique entre les populations de l'ouest (Portugal, Espagne, Corse, Italie, Maroc et Algérie) et de l'est (Grèce, Turquie et Israël) du bassin méditerranéen. Ces deux groupes coïncident avec les groupes génétiques décrits pour les sérotypes du virus de la FCO (Nomikou *et al.* 2009). Les auteurs ont donc conclu que *C. imicola* et le virus de la FCO auraient conjointement colonisé le bassin méditerranéen via deux routes de colonisation : (i) depuis l'Afrique du Nord vers l'Espagne, La France et l'Italie, et (ii) d'Israël à la Grèce en passant par la Turquie (Dallas *et al.* 2003; Nolan *et al.* 2008). Par ailleurs, une étude à partir de populations espagnoles suggère une expansion rapide des populations en Espagne déduite de la forme en étoile du réseau d'haplotypes du gène COI (Calvo *et al.* 2009). Cependant, ces conclusions ont été récemment remises en question par une étude de génétique de populations basée sur le polymorphisme nucléaire de marqueurs

microsatellites. A partir de populations de l'ouest du bassin méditerranéen (Afrique du Nord, Italie et France), cette étude révèle des niveaux de différenciations génétiques relativement faibles (Mardulyn *et al.* 2013). Les auteurs concluent une présence ancienne de *C. imicola* en Italie avec des migrations récurrentes entre les populations étudiées (Mardulyn *et al.* 2013).

L'hypothèse d'une présence ancienne de *C. imicola* dans le bassin méditerranéen, c'est-à-dire bien avant l'émergence de la FCO dans cette région, semble être aussi appuyée par certains travaux. Une étude rétrospective des enquêtes entomologiques réalisées en Italie montre que les populations de *C. imicola* sont restées stables dans le temps et dans l'espace entre 2001 et 2007 (Conte *et al.* 2009). De même, (Acevedo *et al.* 2010) suggèrent à partir de modèles de niches éco-climatiques (prenant en compte la température et l'humidité) et éco-géographiques (considérant les caractéristiques du sol et la disponibilité en hôtes), que l'aire de distribution de *C. imicola* ne s'étendrait probablement pas dans le futur en Espagne. De plus, les arguments basés sur les données historiques des enquêtes entomologiques pour expliquer l'absence de *C. imicola* ne sont pas des preuves irréfutables. Par exemple, les enquêtes entomologiques antérieures à 2000 en Italie (précédant les épidémies de FCO), comportaient des biais méthodologiques (sites de collectes, type de pièges, période de collecte), suggérant que ces collectes n'auraient pas permis de collecter des individus de *C. imicola*, même aujourd'hui (Goffredo & Meiswinkel 2004).

Dans ce contexte où les études génétiques se contredisent et où les données entomologiques historiques comportent des biais, la question de l'histoire de la colonisation du bassin méditerranéen par le moucheron vecteur *C. imicola*, reste non résolue et constitue un modèle de choix pour étudier les mécanismes impliqués dans son expansion géographique.

Problématique et objectifs de la thèse

Le succès invasif des espèces d'arthropodes vecteurs et la propagation de leurs pathogènes constituent un problème de santé publique et une menace pour la sécurité alimentaire à l'échelle de la planète. Dans un système hôtes-pathogènes-vecteurs, déterminer les facteurs sous-jacents de la dynamique des populations de vecteurs est crucial pour comprendre et prédire l'épidémiologie des maladies transmises. Dans un contexte d'invasion biologique, il est d'autant plus important de déterminer pourquoi et comment les espèces vectrices parviennent à coloniser avec succès de nouveaux habitats.

Dans cette thèse, nous nous intéresserons au modèle *C. imicola*/fièvre catarrhale ovine dans sa zone de distribution historique et dans le bassin méditerranéen. Ce modèle questionne sur les processus de colonisation d'un écosystème par une espèce vectrice non-native couplée à l'émergence d'une maladie d'importance en santé animale. L'objectif principal de ce travail est d'améliorer les connaissances sur l'histoire démographique et évolutive, la structuration génétique et la dynamique des populations de *C. imicola* dans le bassin méditerranéen. Ce travail permet également de décrire et comprendre les facteurs expliquant l'extension de son aire de répartition et ceux sous-jacents à sa répartition actuelle.

A ce jour, la caractérisation génétique des populations méditerranéennes de *C. imicola* suggère deux voies de colonisation dans le bassin méditerranéen. Les études posant ces hypothèses présentent des limites méthodologiques et la chronologie de la colonisation fait débats dans la littérature. Afin d'apporter de nouveaux éléments de compréhension, **le premier chapitre** présente une étude multi-loci combinant de manière intégrative des approches de phylogéographie, de génétique de populations et d'analyse basée sur les méthodes ABC. Les objectifs spécifiques sont de caractériser (i) la structuration génétique des populations de *C. imicola* sur l'ensemble de son aire de distribution, et de décrire (ii) les populations sources, les voies et la période de colonisation depuis son aire native.

La colonisation d'un nouveau milieu hors de l'aire native nécessite une stratégie de dispersion adaptée. La migration récurrente de *C. imicola* entre l'Afrique du Nord et l'Italie a déjà été suggérée comme un facteur clef influençant la dynamique des

populations italiennes de *C. imicola* (Mardulyn *et al.* 2013). Dans **le deuxième chapitre**, nous nous intéresserons à l'histoire de la colonisation du sud de l'Europe depuis les populations installées en Afrique du Nord. Nous décrirons la distribution de la variabilité génétique dans l'espace, et caractériserons le mode et l'intensité de la dispersion, et son rôle dans la colonisation et la dynamique des populations au sein du bassin méditerranéen.

Les modèles de niches éco-climatiques suggèrent que sous l'hypothèse d'une augmentation globale des températures, de nouvelles zones en région paléarctique seront favorables à *C. imicola* (Guichard *et al.* 2014; Wittmann & Baylis 2000). Comprendre les facteurs historiques, géographiques, démographiques et évolutifs expliquant la colonisation et l'installation de populations non-natives permet d'anticiper des phénomènes d'expansion des populations installées et de prédire de nouveaux phénomènes d'invasion (Guillemaud *et al.* 2011), et ainsi faciliter la gestion du risque sanitaire. Le **troisième chapitre** apporte une meilleure compréhension des processus d'expansions à la limite des distributions géographiques de *C. imicola.* Nous décrirons les routes de colonisation de la population observée dans le département des Pyrénées-Orientales et déterminerons le rôle de la dispersion par le vent dans cet évènement d'introduction. De plus, nous étudierons la dynamique spatio-temporelle des populations à l'échelle locale en Corse et dans le département du Var, et caractériserons la variabilité génétique des populations à la limite de l'aire de distribution.

L'ensemble des travaux conduits durant cette thèse seront enfin synthétisés et discutés en relation avec les questions écologiques, évolutives et démographiques soulevées dans chacun des chapitres. Nous terminerons en ouvrant notre travail vers de nouvelles perspectives de recherche.

Chapitre I : Histoire évolutive de la colonisation du bassin méditerranéen par *C. imicola*

Quelles sont les populations sources d'une population envahissante ? Quelles sont les routes empruntées par la population envahissante ? La population envahissante s'est-elle établie suite à une ou plusieurs introductions ?

Ces questions sont essentielles pour poser des hypothèses concernant les causes éco-évolutives du succès ou de l'échec des invasions biologiques, et mettre en place des stratégies de gestion et de contrôle dans les régions sources, le long des voies de colonisation ou au point d'entrée des régions envahies (Hulme 2009). Les données historiques relatives à la présence/absence des organismes envahissants sont souvent incomplètes voire inexistantes. En effet, la détection précoce de l'introduction d'un organisme non-natif et le suivi de son expansion sont toujours difficiles à mettre en œuvre. L'utilisation de méthodes indirectes basées sur l'information contenue dans les génomes se révèle de plus en plus indispensable pour répondre à ces questions (Wares et al. 2005, Estoup and Guillemaud 2010).

Dans le cas de *C. imicola*, les données entomologiques historiques statuant sur la présence de l'espèce sont biaisées. En effet, ce sont les foyers de la fièvre catarrhale ovine à la fin des années 1990 qui ont motivé la mise en place des collectes. Ainsi, il n'existe pas de preuve d'absence formelle de l'espèce sur la majorité des territoires méditerranéens avant les premières observations entomologiques. Les premières études génétiques n'ont pas permis de caractériser la chronologie et les routes de colonisation de l'espèce depuis son aire native (Dallas *et al.* 2003; Mardulyn *et al.* 2013; Nolan *et al.* 2008). De plus, jusqu'à présent, les connaissances de l'histoire évolutive et démographique de *C. imicola* restent incomplètes.

Ce premier chapitre présente une étude génétique avec une approche multi-marqueur qui permet d'apporter de nouveaux éléments à la compréhension du processus de colonisation de *C. imicola* entre la zone native et les zones supposées nouvellement colonisées. Pour la première fois, le jeu de données inclut des populations de plusieurs régions d'Afrique sub-saharienne (d'Afrique de l'ouest à l'océan indien), du Maghreb et d'Europe méditerranéenne. Neuf marqueurs microsatellites et deux gènes mitochondriaux, le cytochrome oxydase subunit I (COI) et le cytochrome b (CytBCytB) ont été utilisés conjointement. La diversité, la variabilité génétique intra- et inter-populationnelle et les liens généalogiques entre les populations de *C. imicola* ont été caractérisés en Afrique et dans les régions méditerranéennes avec des approches combinées de phylogéographie et de génétique de populations. Par ailleurs, les populations sources de l'aire native et les voies empruntées par les individus fondateurs de l'espèce ont été testées et déterminées à l'aide de méthodes d'inférence ABC.

Les résultats montrent que les populations de *C. imicola* sont génétiquement structurées en quatre groupes géographiques, avec une différenciation génétique fortement marquée ouest/est dans le bassin méditerranéen et un continuum ouest/sud en Afrique sub-saharienne. Les inférences démographiques basées sur les séquences mitochondriales indiquent que les populations de l'ouest du bassin méditerranéen et de l'Afrique sub-saharienne auraient subi une expansion démographique durant le Pléistocène (200,000-50,000), probablement en lien avec les variations climatiques caractéristiques de cette période. D'autre part, les analyses ABC suggèrent que des individus colonisateurs seraient remontés depuis la région sub-saharienne via une voie à la fin du Pléistocène ou au début de l'Holocène pour coloniser le bassin méditerranéen. Néanmoins, l'hypothèse d'une remontée via deux voies de colonisation n'est pas exclue comme le suggère le polymorphisme des gènes mitochondriaux. Les variations climatiques ou d'autres facteurs environnementaux durant ces périodes auraient ensuite contribué à moduler la distribution géographique de l'espèce.

Notre étude apporte ainsi pour la première fois, (i) une caractérisation génétique des populations de *C. imicola* dans son aire native comparativement à l'aire envahie et (ii) une description de l'histoire de colonisation du bassin méditerranéen par cette espèce. Notre travail confirme les résultats de (Mardulyn et al. 2013), bouleversant le dogme largement répandu dans la littérature d'une arrivée récente de *C. imicola* dans le bassin méditerranéen concomitante à l'épizootie de FCO dans cette région.

Ce chapitre a été accepté pour publication dans le journal *Molecular Ecology*, DOI: 10.1111/mec.13422.

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 DELÉCOLLE[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 (250 words)

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 through the identification of population sources and geographical pathways is critical for predicting future colonization and implementing effective 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 population sources 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) (Delécolle & 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 deduced 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 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 (Delécolle & 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 microsatellite 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 microsatellite 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 Réunion 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 sample sites 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 methods 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 artifact 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.

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.

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*. Havard university press, Cambridge, United Kingdom.

Bailly-Choumara H, Kremer M (1970) Epizooties de 1989, 1990 et 1991. Données épidemiologiques et stratégie 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. *Molecular Biology and Evolution* **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. *Molecular Ecology* **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 Evolutionary Biolology* **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. *Molecular Ecology* **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. *Scientific Technical Review* **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. *Molecular Ecology* **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. *Bulletin of Entomological Research* **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. *Medical and Veterinary Entomology* **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. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20159-20163.

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. *The American Journal of Human Genetics* **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. *Molecular Biology and Evolution* **30**, 1224-1228.

Corander J, Sirén J, Arjas E. (2008) Bayesian Spatial Modelling of Genetic Population Structure. *Computational Statistics* **23**, 111-129.

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* **30**, 1187-1189.

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 DIYABC: 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. *Medical and Veterinary Entomology* **17**, 379-387.

Delécolle 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. *Molecular Biology and Evolution* **29**, 1969-1973.

Dyce A, Wirth W (1983) Reappraisal of some Indian *Culicoides* species in the subgenus Avaritia (Diptera: Ceratopogonidae). *International Journal of Entomological Research* **25**, 221-225.

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology* **19**, 4113-4130.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **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**, 47-50.

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. *Molecular Ecology* **10**, 305-318.

Goffredo M, Conte A, Cocciolito R, Meiswinkel R (2003) The distribution and abundance of *Culicoides imicola* in Italy. *Vetirinaria Italiana* **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. *Vetirinaria Italiana* **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* **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. *Journal of the Royal Society Interface* **9**, 339-350.

Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology & Evolution* **16**, 700-706.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907-913.

Hewitt G (2001) Speciation, hybrid zones and phylogeography-or seeing genes in space and time. *Molecular Ecology* **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). *Molecular Ecology* **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. *Molecular Ecology* **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* **114**, 333-343.

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*. *Molecular Ecology* **23**, 5979-5997.

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*. *Molecular Ecology* **20**, 4654-4670.

Lorenzen ED, Heller R, Siegismund HR (2012) Comparative phylogeography of African savannah ungulates. *Molecular Ecology* **21**, 3656-3670.

Maan S, Samuel A, Maan N, *et al.* (2004) Molecular epidemiology of bluetongue viruses from disease outbreaks in the Mediterranean Basin. *Vetirinaria Italiana* **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. *Molecular Ecology* **22**, 2456-2466.

Mellor PS, Baylis M, Mertens PP (2009) Bluetongue Elsevier Academic Press, London, United Kingdom.

Mellor PS, Boorman J, Baylis M (2000) *Culicoides* biting midges: their role as arbovirus vectors. *Annual Review of Entomology* **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. *Preventive Veterinary Medicine* **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. *The Veterinary Journal* **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. *Medical and Veterinary Entomology* **17**, 52-54.

Monteys VSI, Saiz-Ardanaz M (2003) *Culicoides* midges in Catalonia (Spain), with special reference to likely bluetongue virus vectors. *Medical and Veterinary Entomology* **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**, 415-417.

Muirhead JR, Gray DK, Kelly DW, *et al.* (2008) Identifying the source of species invasions: sampling intensity vs. genetic diversity. *Molecular Ecology* **17**, 1020-1035.

Murrell B, Wertheim JO, Moola S, *et al.* (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics* **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. *Medical and Veterinary Entomology* **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. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16444-16447.

Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution* **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. *Molecular Ecology* **16**, 3069-3083.

Patakakis MJ (2004) *Culicoides imicola* in Greece. *Vetirinaria Italiana* **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. *Molecular Biology and Evolution* **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. *Nature Reviews Microbiology* **3**, 171-181.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **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. *Molecular Biology and Evolution* **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. *Molecular Ecology* **15**, 407-420.

Simberloff D, Martin JL, Genovesi P, *et al.* (2013) Impacts of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution* **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* **54**,163–182.

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 Research* **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. In: *Arthropods as vectors of emerging diseases* (Mehlhorn H). pp. 77-102. Springer-Verlag, Berlin, Germany.

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, Beijing, China.

Figures and Tables

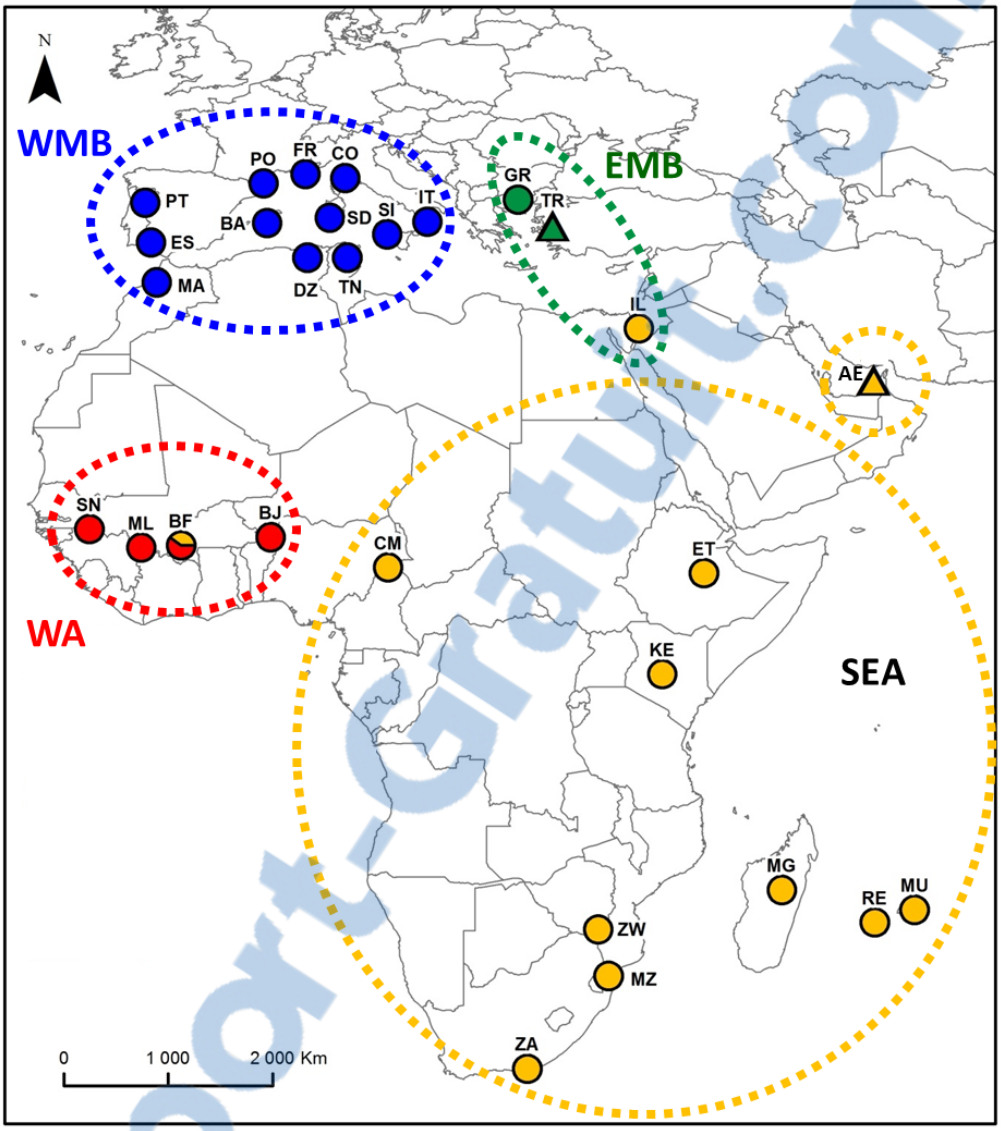


Fig. 1 Sampling locations and genetic clustering of *C. imicola* sampled populations based on COI sequences. Genetic groups were assessed by the spatial group clustering method of Corander et al. (2008) implemented in BAPS v.6.0. Sample sites with the same color belong to the same cluster. Details on sample codes are given in Table 1.

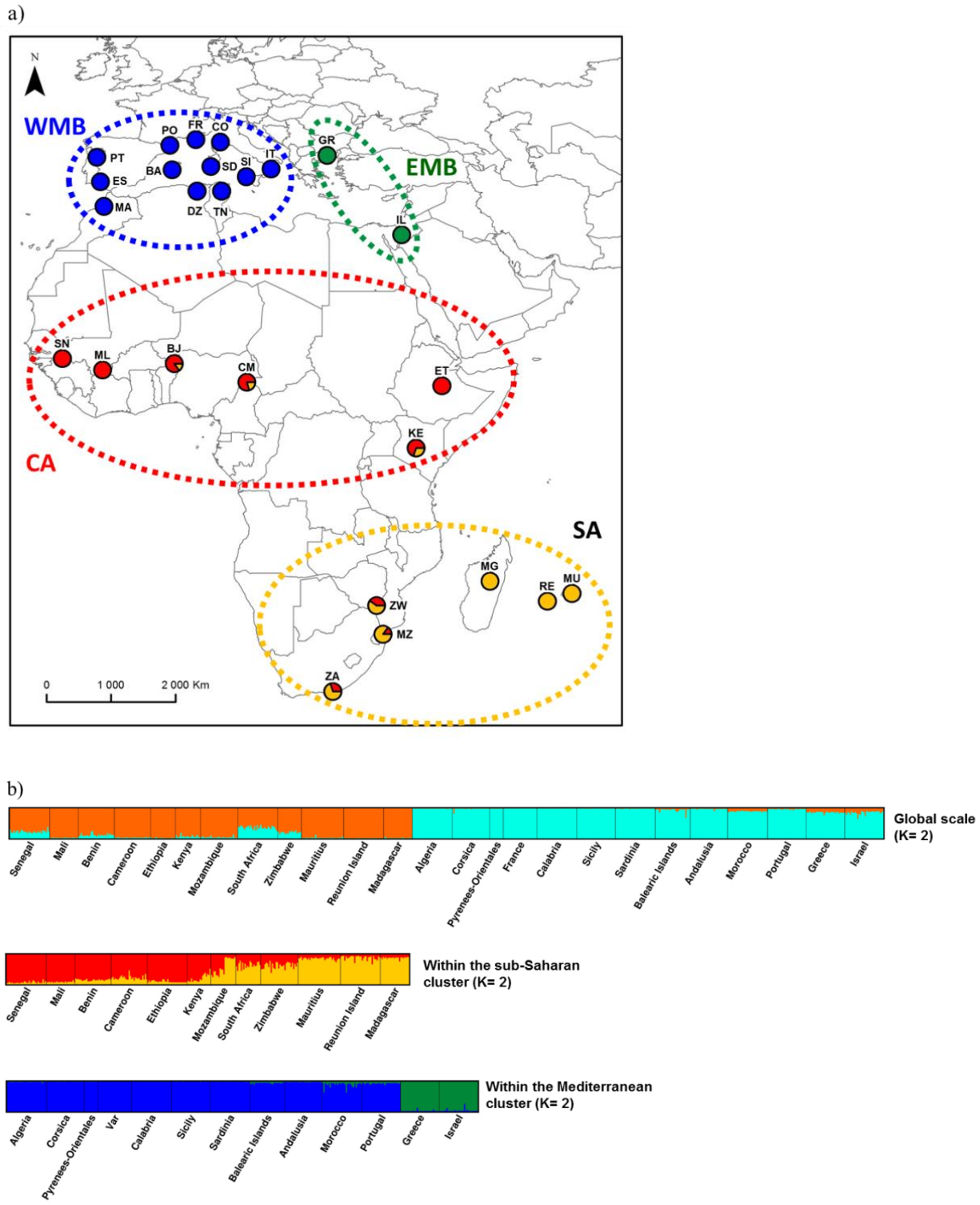


Fig. 2 Genetic clustering of *C. imicola* sampled populations. (a) Spatial Bayesian clustering based on microsatellite data. (b) Ancestry estimation assuming two population clusters at the global scale (upper part) and four population clusters at the genetic groups scale (K=2 within sub-Saharan Africa 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.

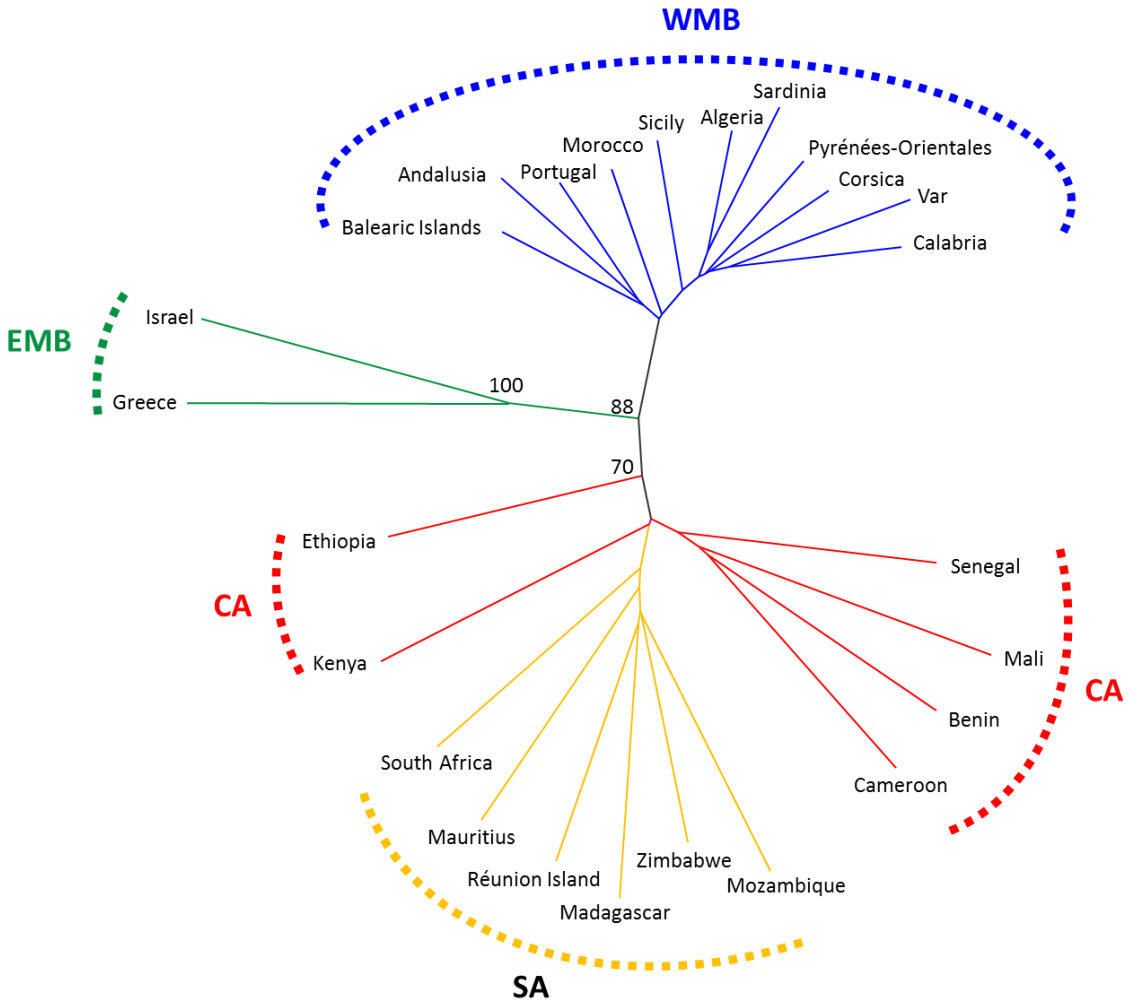


Fig. 3 Neighbor-joining tree of *C. imicola* population samples based on the chord distance of Cavalli-Sforza & Edwards (1967) computed on microsatellite polymorphism. Central African populations are shown in red, southern African populations in yellow, eastern Mediterranean populations in green and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

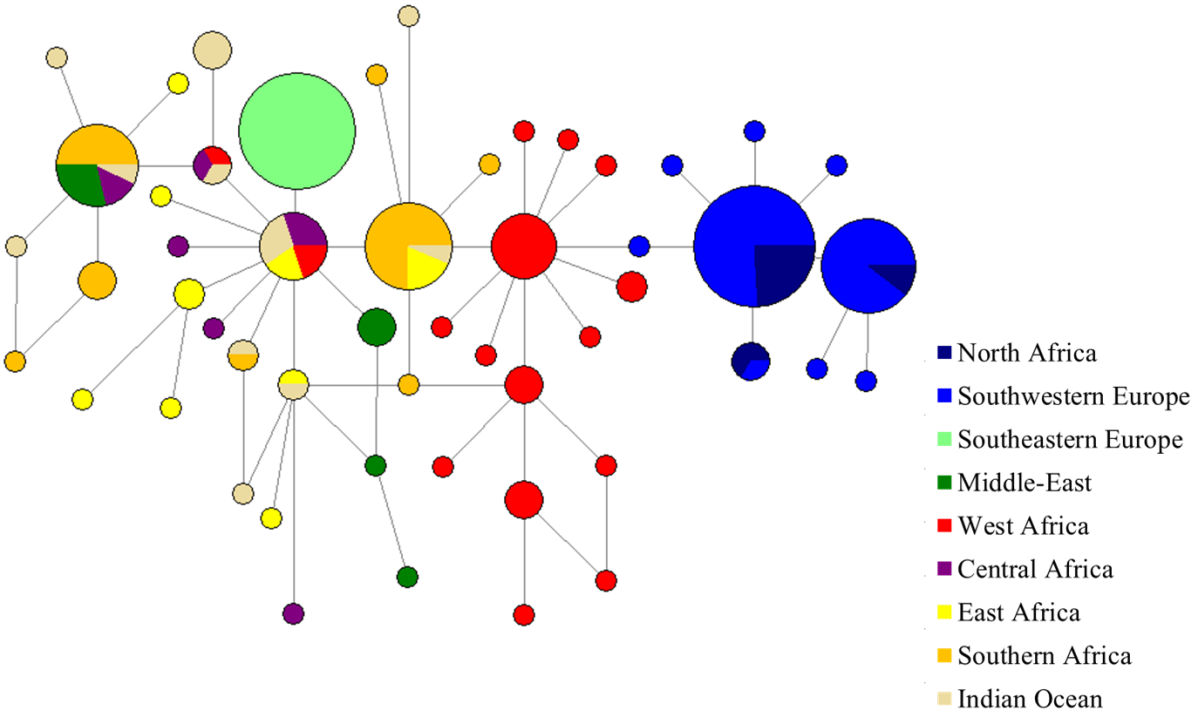


Fig. 4 Median-joining haplotype network of COI mitochondrial sequences of *C. imicola*. The size of the circles is proportional to the number of individuals with that haplotype. The length of the branches separating haplotypes is proportional to the number of mutational steps between them. Haplotype 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.

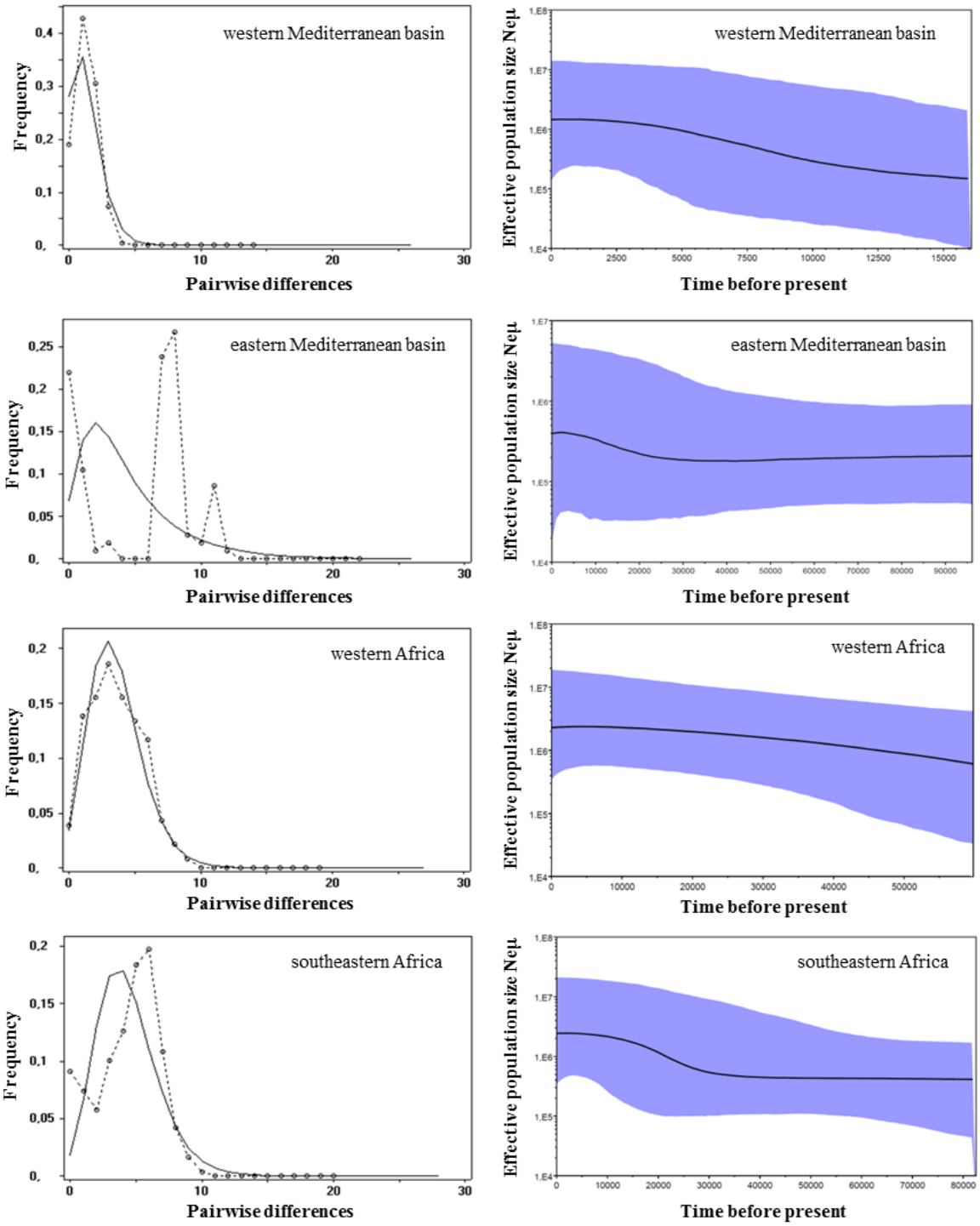


Fig. 5 Mismatch distribution among pairwise differences among haplotypes (a) and Bayesian skyline plot (b) based on COI and CytB combined dataset in *C. imicola* different geographical groups. Mismatch analyses were conducted according to a growth-decline model. Observed data and theoretical expected 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.

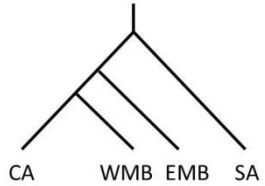
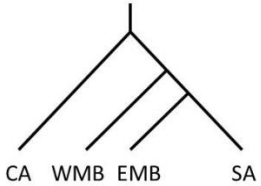
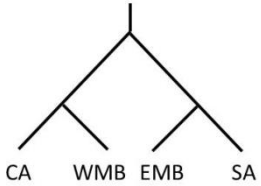
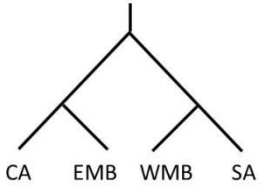
Table 1 Geographical location of *C. imicola* sampled sites, number of individuals typed for microsatellite analyses (Nmic), number mitochondrial sequences obtained (Nmit) data and sample grouping based on clustering analysis based on microsatellite (Mic) and mitochondrial (mtDNA) data.

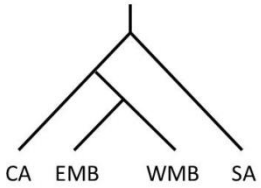
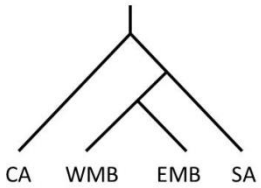
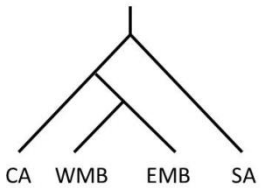
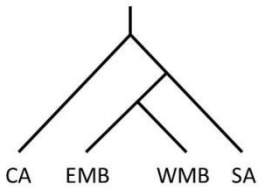
Sampling location	**Sampling Year**	**Latitude**	**Longitude**	**Nmic**	**Nmit**	**Sample Code**	**Cluster Code (Mic)**	**Cluster Code (mtDNA)**
Algeria	2003	36.8	8.5	32	9	DZ	WMB	WMB
France-Corsica	2008	42.8	9.4	30	8	CO	WMB	WMB
France-Pyrénées-Orientales	2012	42.4	2.8	11	8	PO	WMB	WMB
France-Var	2008	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.

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.

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 are highlighted in bold.

Table 3 Pairwise F_{ST} values across loci between the genetic clusters inferred by STRUCTURE v.2.3.3 of *C. imicola*.

	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**

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.

Supplementary Information

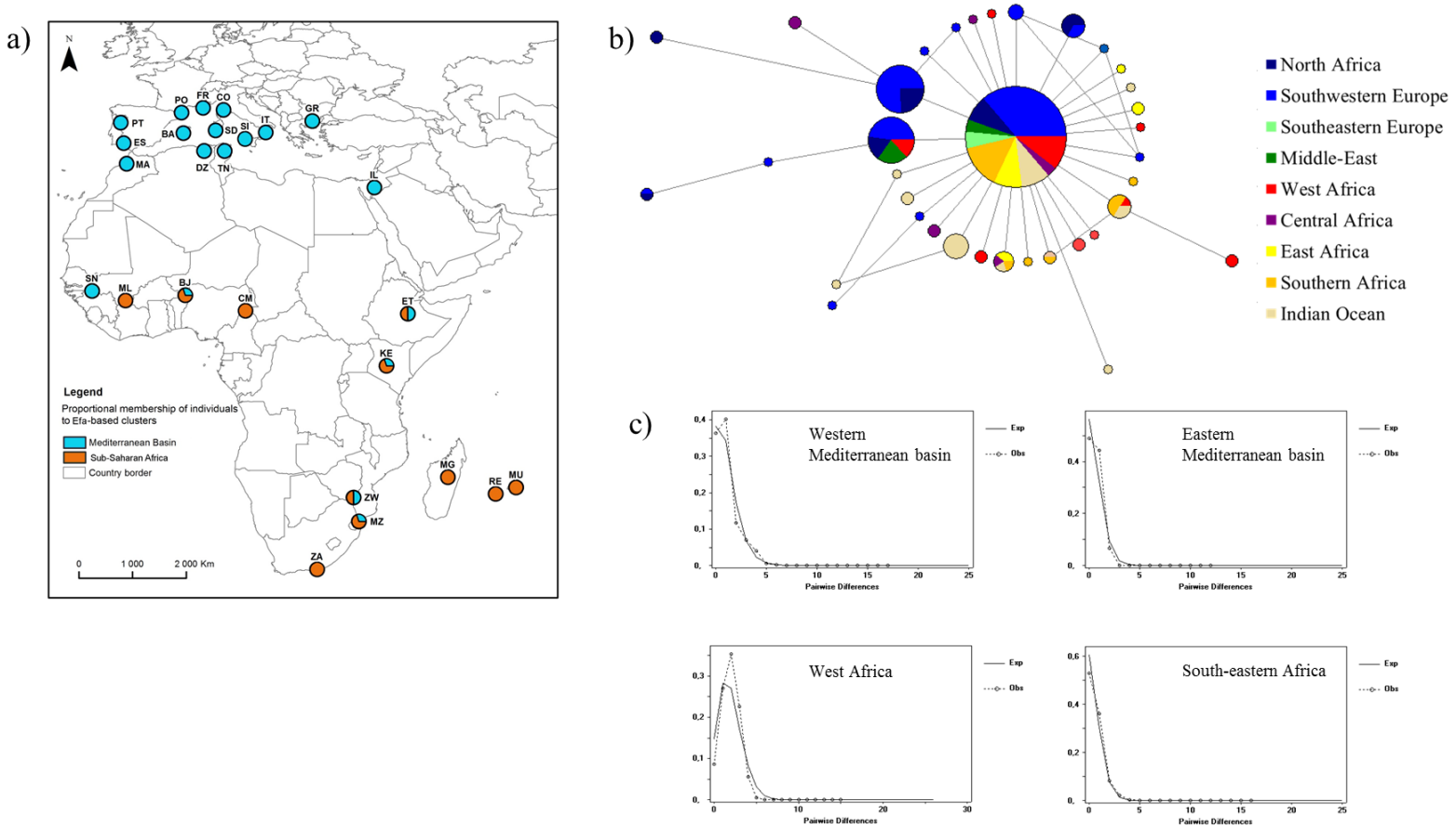


Fig. S1 Genetic clustering of *C. imicola* sampled populations (a), Efα median-joining haplotype network (b) and mismatch distribution among pairwise differences among haplotypes based on EFα sequences (c). (a) Genetic groups were assessed by the spatial group clustering method of Corander *et al.* (2008) implemented in BAPS v.6.0. Sample sites with the same color belong to the same cluster. (b) 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. Theb haplotype network was constructed using NETWORK v.4.6.1.2. (c) Mismatch analyses were conducted according to a growth-decline model. Observed data and theoretical expected distributions are represented by discontinuous and solid line, respectively.

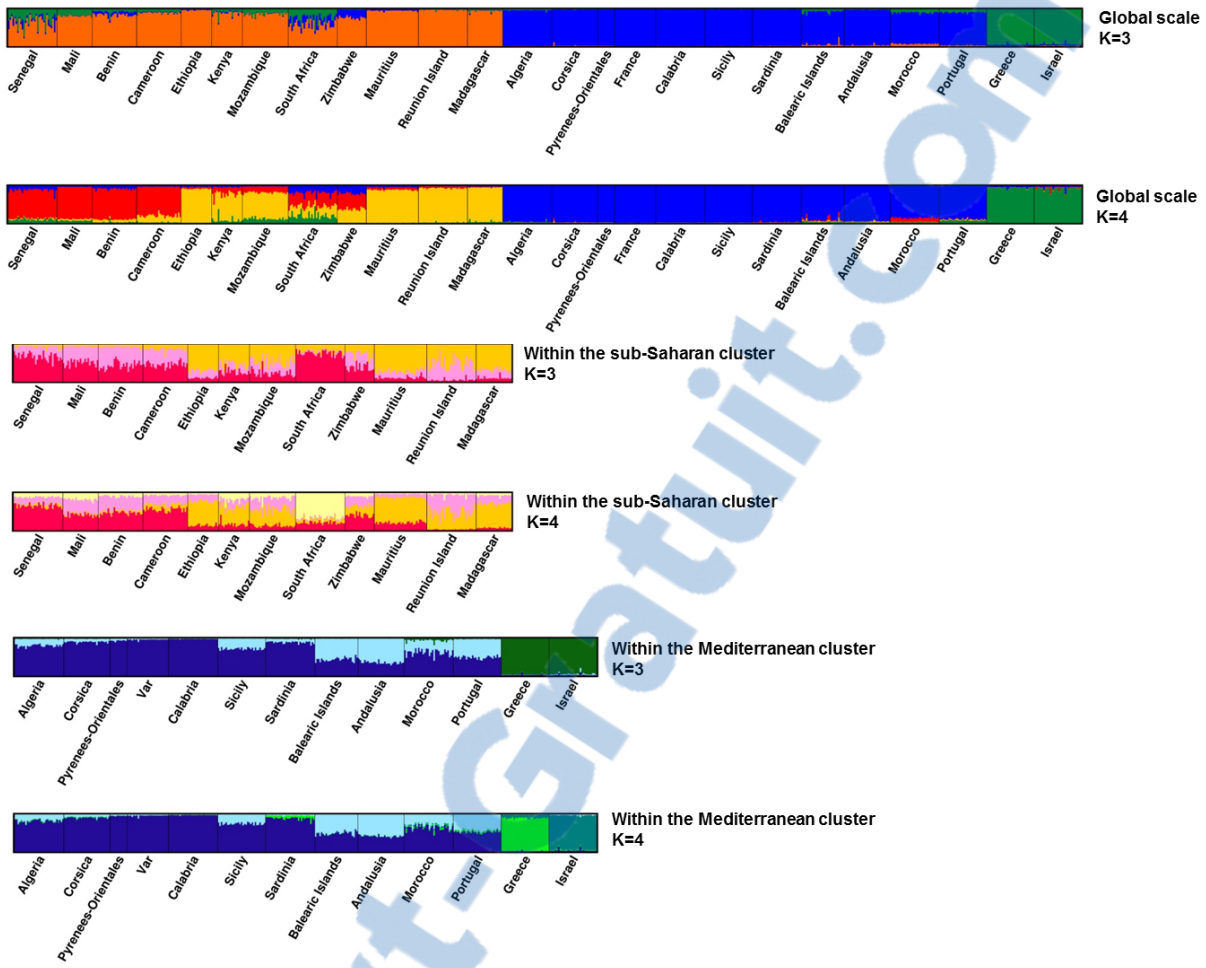


Fig. S2 Ancestry estimation assuming three and four population clusters (K=3 and K=4) based on the Bayesian clustering method implemented in STRUCTURE v.2.3.3. Each vertical line represents an individual, and each color represents a cluster. Individuals are grouped by sample site.

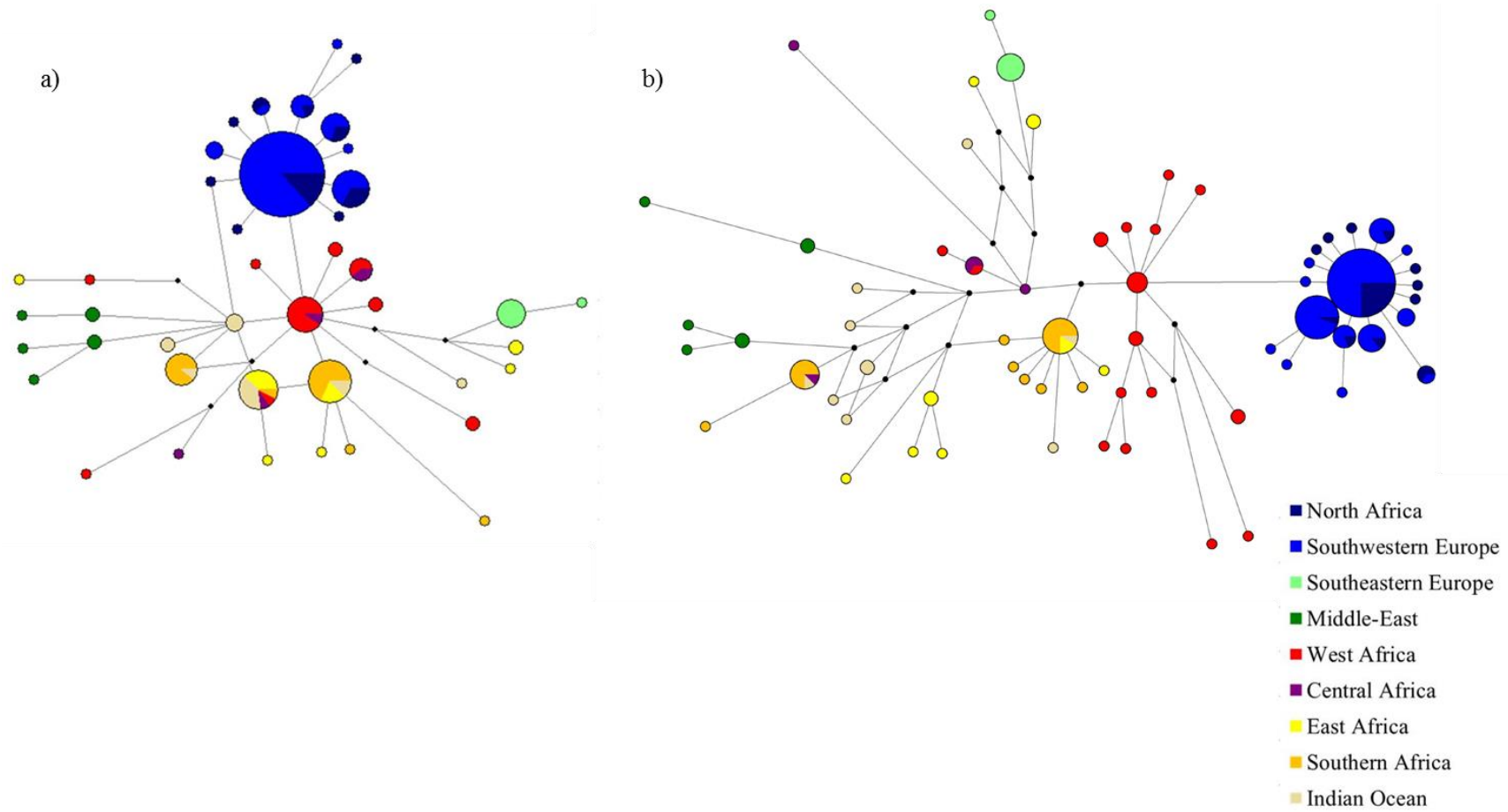


Fig. S3 Median-joining haplotype network of a) Cytochrome b and b) COI and Cytochrome b concatenated sequences of *C. imicola*. The size of the circles is proportional to the number of individuals with that haplotype. The length of the branches separating haplotypes is proportional to the number of mutational 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; Western 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.

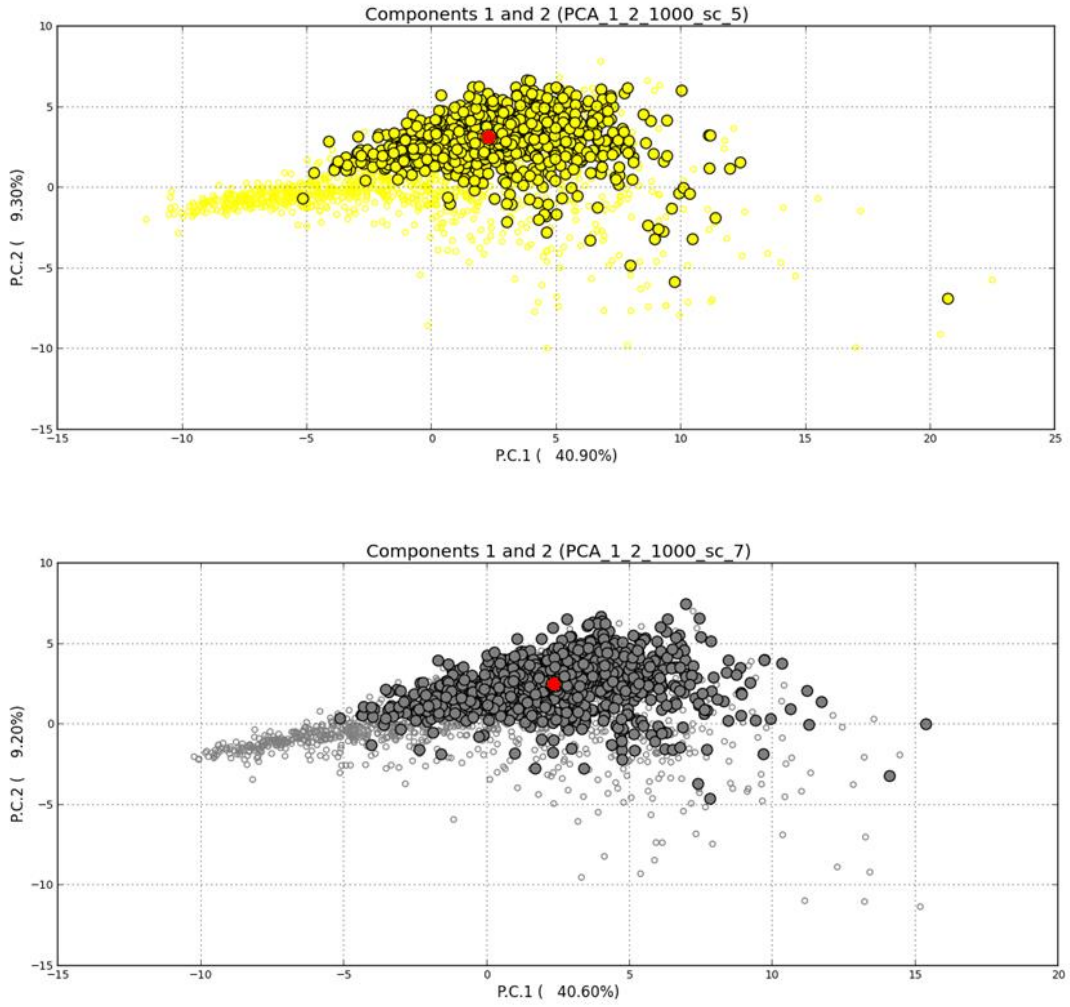


Fig. S4 PCA summary statistics simulated under the most probable scenarios: Scenario 5 (atop) and 7 (bottom). Small circles correspond to simulated statistics from the prior, the large dots correspond to the simulated posteriors and the red dot represents the real data set.

Table S1 Primers used for the amplification of the microsatellite loci in *C. imicola* (Mardulyn et al. 2013), allele size range observed in this study and annealing temperature (Ta) used for amplification.

Loci	Motif	Forward / Reverse	Allele size range (bp)	Ta (°C)
68	(GT)n	CTTTTCCGTTTCTTTTTATTTCTTT	95-105	60
		GTTTCTTTCTGGTCGCGTTGGTTGCTG		
12b	(CT)n	TTATGTGTGTATGTTAGCAAGGTCA	127-145	50
		GTTTCTTCTTCGGATCAAAGAAATTTTGCC		
3b	(AC)n	ATGCGGATGTTTGAAGTG	147-198	50
		GTTTCTTTTTTGTGTCTTATTGCCC		
31	(CAA)n	TTCTGTTCGGCTGTTGCGTT	151-171	60
		GTTTCTTCTTTTTACGTGGTGGTCATTC		
41b	(CT)n	GAGGAGGAGGTAGAA	157-168	50
		GTTTCTTTCTATTAGTCAATGGTG		
35t	(AC)n	TTTGTAAAAGCCAGTTCAACCG	164-214	60
		GTTTCTTATCGAACGAAGGAAATAACCAC		
88b	(AC)n	TTTGTTCGATTTGTAGTG	223-263	50
		GTTTCTTCCTCTCTTTCATTCGC		
16	(TG)n	TTGCCTTTGCTTGTTGAGGATG	270-307	60
		GTTTCTTTCCTCTTTAAAATCACTGACGTG		
88	(CAT)n	GTTGGTGCTTTGTTGTGTTGT	336-351	50
		GTTTCTTTTTCTTTTTCTCCTTTTTGTTTCTTTC		

Table S2 Prior distributions of demographic, historical and mutation parameters used for ABC inferences.

Parameter description	Parameter	Prior distribution
Stable effective population size in ancestral (a) and Mediterranean (m) populations	N_a, N_m	Loguniform [100; 200000]
Founding effective population size	N_{emb}, N_{wmb}	Loguniform [2; 100]
Duration of bottleneck	Db	Uniform [0; 50]
Time of colonization events in Eastern Mediterranean basin	t_{emb}	Uniform [430; 30000]
Time of colonization events in Western Mediterranean basin	t_{wmb}	Uniform [330; 30000]
Divergence time of native unsampled populations	t_{Uca1}, t_{Usa1}	Loguniform [430; 30000]
	t_{Uca2}, t_{Usa2}	Loguniform [330; 30000]
Ancestral divergence time	t_a	Uniform [100000; 200000]
Mean microsatellite mutation model	μ	Loguniform [6.10^{-6}; 10^{-4}]
Mean parameter of the geometric distribution of length	P	Uniform [10^{-1}; 3.10^{-1}]
Mean mutation rate for nucleotide instability	SNI	Uniform [10^{-8}; 10^{-4}]

Effective population sizes (N) are expressed in number of diploid individuals and times of events (t) in number of generations going back to the past. For each scenario, native unsampled populations were simulated so that each of them merges into the sampled native population at time t_{Uca} and t_{Usa}. Conditions among the parameters used during the simulations were t_{emb} and $t_{wmb} \leq t_a$, t_{Uca1}, t_{Usa1}, t_{Uca2} and $t_{Usa2} \leq t_a$.

Table S3 Genetic diversity based on mitochondrial data for each sampled site and genetic cluster assessed by the spatial group clustering implemented in BAPS v.6.0.

Locality	$\mathbf{H_{COI}}$	$\boldsymbol{Hd_{COI}}$	$\boldsymbol{\pi_{COI}}$	$\mathbf{H_{CytB}}$	$\boldsymbol{Hd_{CytB}}$	$\boldsymbol{\pi_{CytB}}$
WMB	**9**	**0.447 ± 0.056**	**0.0010 ± 0.0003**	**13**	**0.590 ± 0.057**	**0.0011 ± 0.0002**
Algeria	2	0.222 ± 0.166	0.0004 ± 0.0007	5	0.806 ± 0.120	0.0021 ± 0.0005
France-Corsica	2	0.429 ± 0.169	0.0009 ± 0.0001	3	0.464 ± 0.200	0.0008 ± 0.0004
France-Pyrénées-Orientales	4	0.750 ± 0.139	0.0022 ± 0.0002	2	0.250 ± 0.180	0.0003 ± 0.0002
France-Var	2	0.429 ± 0.169	0.0009 ± 0.0001	1	0	0
Italy-Calabria	2	0.571 ± 0.094	0.0012 ± 0.0002	1	0	0
Italy-Sardinia	1	0	0	3	0.679 ± 0.122	0.0013 ± 0.0003
Italy-Sicily	3	0.464 ± 0.200	0.0010 ± 0.0003	3	0.464 ± 0.200	0.0008 ± 0.0004
Morocco	4	0.750 ± 0.139	0.0019 ± 0.0004	4	0.750 ± 0.139	0.0019 ± 0.0004
Portugal	2	0.250 ± 0.180	0.0005 ± 0.0001	2	0.429 ± 0.169	0.0007 ± 0.0001
Spain-Andalusia	3	0.464 ± 0.200	0.0010 ± 0.0003	4	0.786 ± 0.113	0.0020 ± 0.0005
Spain-Balearic Islands	1	0	0	3	0.679 ± 0.122	0.0013 ± 0.0003
Tunisia	1	0	0	4	0.900 ± 0.161	0.0020 ± 0.0005
EMB	**4**	**0.500 ± 0.091**	**0.0032 ± 0.0006**	**7**	**0.781 ± 0.102**	**0.0055 ± 0.0008**
Greece	1	0	0	2	0.250 ± 0.180	0.0003 ± 0.0002
Turkey	2	0.095 ± 0.084	0.0006 ± 0.0001	-	-	-
Israel	3	0.679 ± 0.122	0.0046 ± 0.0005	5	0.905 ± 0.103	0.0049 ± 0.0009
WA	**16**	**0.899 ± 0.044**	**0.0041 ± 0.0005**	**9**	**0.798 ± 0.076**	**0.0030 ± 0.0006**
Benin	6	0.893 ± 0.111	0.0037 ± 0.0003	4	0.867 ± 0.129	0.0047 ± 0.0011
Burkina Faso	5	0.905 ± 0.103	0.0054 ± 0.0006	4	0.800 ± 0.172	0.0030 ± 0.0010
Mali	5	0.905 ± 0.103	0.0042 ± 0.0003	3	0.833 ± 0.222	0.0034 ± 0.0010
Senegal	4	0.750 ± 0.139	0.0019 ± 0.0002	1	0	0
SEA	**24**	**0.891 ± 0.026**	**0.0048 ± 0.0003**	**16**	**0.846 ± 0.030**	**0.0033 ± 0.0003**
Cameroon	5	0.905 ± 0.137	0.0040 ± 0.0005	5	0.933 ± 0.122	0.0041 ± 0.0008
Ethiopia	5	0.933 ± 0.122	0.0057 ± 0.0006	2	0.333 ± 0.215	0.0005 ± 0.0003
Kenya	4	0.810 ± 0.130	0.0024 ± 0.0002	5	0.857 ± 0.108	0.0048 ± 0.0008
Mozambique	5	0.786 ± 0.151	0.0035 ± 0.0002	3	0.524 ± 0.209	0.0016 ± 0.0008
South Africa	3	0.679 ± 0.122	0.0039 ± 0.0002	3	0.714 ± 0.127	0.0031 ± 0.0006
Zimbabwe	3	0.679 ± 0.122	0.0040 ± 0.0003	3	0.679 ± 0.122	0.0040 ± 0.0003
Madagascar	6	0.952 ± 0.096	0.0056 ± 0.0005	4	0.821 ± 0.101	0.0029 ± 0.0007
Mauritius	3	0.833 ± 0.222	0.0031 ± 0.0002	1	0	0
France- Réunion Island	4	0.800 ± 0.172	0.0042 ± 0.0003	4	0.800 ± 0.172	0.0042 ± 0.0003
U.A.E.	1	0	0	-	-	-

The number of haplotypes (H), haplotype diversity (*Hd*), and nucleotide diversity (π) and their standard deviations (*sd*) for COI and CytB were estimated using DNASP v.5.10.

Clusters are coded as: WMB, western Mediterranean Basin, EMB, eastern Mediterranean Basin, WA, western Africa, SEA, southeastern Africa

Table S4 Genetic diversity based on microsatellite data for each sampled site and genetic cluster assessed by STRUCTURE v.2.3.3.

Locality	Ar	Hs	Fis	Ho	He
WMB	**5.854 ± 2.059**	**0.624 ± 0.099**	**0.083***	**0.572 ± 0.105**	**0.624 ± 0.100**
Algeria	4.007 ± 1.003	0.608 ± 0.102	0.107	0.543 ± 0.090	0.607 ± 0.101
France-Corsica	3.953 ± 0.892	0.619 ± 0.089	0.112	0.550 ± 0.141	0.618 ± 0.090
France–Pyrénées-Orientales	4.111 ± 1.166	0.612 ± 0.096	-0.039	0.636 ± 0.136	0.614 ± 0.097
France-Var	3.427 ± 0.897	0.572 ± 0.121	0.024	0.558 ± 0.147	0.572 ± 0.122
Italy-Calabria	3.658 ± 0.762	0.578 ± 0.167	0.058	0.545 ± 0.183	0.578 ± 0.167
Italy-Sardinia	3.958 ± 1.400	0.591 ± 0.143	0.052	0.560 ± 0.156	0.591 ± 0.143
Italy-Sicily	4.229 ± 0.919	0.628 ± 0.071	0.077	0.580 ± 0.142	0.628 ± 0.072
Morocco	4.384 ± 1.073	0.617 ± 0.111	0.030	0.599 ± 0.118	0.617 ± 0.111
Portugal	4.347 ± 1.369	0.626 ± 0.158	0.096	0.566 ± 0.197	0.625 ± 0.159
Spain-Andalusia	4.062 ± 1.220	0.642 ± 0.111	0.025	0.626 ± 0.117	0.642 ± 0.111
Spain-Balearic Islands	4.430 ± 1.439	0.628 ± 0.110	0.097	0.567 ± 0.181	0.627 ± 0.110
EMB	**6.494 ± 3.883**	**0.529 ± 0.232**	**0.101***	**0.475 ± 0.221**	**0.530 ± 0.232**
Greece	3.804 ± 2.401	0.494 ± 0.315	0.051	0.528 ± 0.264	0.494 ± 0.314
Israel	3.938 ± 1.571	0.492 ± 0.175	-0.003	0.493 ± 0.202	0.492 ± 0.176
CA	**11.949 ± 6.66**	**0.729 ± 0.126**	**0.191***	**0.588 ± 0.092**	**0.728 ± 0.118**
Benin	6.535 ± 2.843	0.738 ± 0.160	0.254*	0.550 ± 0.120	0.734 ± 0.158
Cameroon	6.158 ± 2.661	0.701 ± 0.115	0.171*	0.567 ± 0.158	0.673 ± 0.198
Ethiopia	5.758 ± 1.701	0.693 ± 0.129	0.149*	0.590 ± 0.134	0.691 ± 0.129
Kenya	6.567 ± 3.176	0.742 ± 0.109	0.192*	0.599 ± 0.135	0.738 ± 0.108
Mali	6.808 ± 3.392	0.675 ± 0.200	0.159*	0.646 ± 0.190	0.735 ± 0.127
Senegal	6.664 ± 2.898	0.738 ± 0.126	0.124*	0.581 ± 0.161	0.699 ± 0.115
SA	**12.177 ± 5.991**	**0.745 ± 0.132**	**0.185***	**0.607 ± 0.106**	**0.744 ± 0.141**
Mozambique	6.235 ± 2.281	0.718 ± 0.103	0.188*	0.583 ± 0.113	0.714 ± 0.102
South Africa	6.550 ± 2.611	0.716 ± 0.152	0.189*	0.580 ± 0.159	0.713 ± 0.151
Zimbabwe	6.840 ± 2.777	0.768 ± 0.134	0.172*	0.636 ± 0.146	0.766 ± 0.132
Madagascar	6.797 ± 3.092	0.724 ± 0.159	0.081	0.665 ± 0.122	0.722 ± 0.157
Mauritius	6.189 ± 3.166	0.704 ± 0.186	0.134*	0.609 ± 0.170	0.702 ± 0.185
France- Réunion Island	7.111 ± 3.553	0.756 ±0.147	0.254*	0.564 ± 0.128	0.753 ± 0.145

The allelic richness (Ar), genetic diversity (Hs) and inbreeding coefficient (Fis) were computed with FSTAT v2.9.3.2. The observed and expected heterozygosity were computed with ARLEQUIN v3.5.2. Clusters are coded as: WMB, western Mediterranean Basin, EMB, eastern Mediterranean Basin, CA, central Africa, SA, southern Africa.

Significant *Fis* values, at the adjusted nominal level (5%) for multiple comparison of 0.0002, are indicated with an asterisk.

Table S5 Pairwise F_{ST} values between *C. imicola* population samples estimated with FSTAT v2.9.3.2.

	CA					**SA**						**WMB**											**EMB**	
Sites	ML	BJ	CM	ET	KE	MZ	ZA	ZW	MU	RE	MG	DZ	CO	PO	FR	CA	SI	SD	BA	ES	MA	PT	GR	IL
SN	0.01	0.02	0.02	**0.03**	0.02	**0.06**	0.06	**0.03**	**0.05**	**0.05**	**0.06**	**0.08**	**0.08**	**0.09**	**0.14**	**0.10**	**0.09**	**0.10**	**0.08**	**0.07**	**0.07**	**0.07**	**0.14**	**0.14**
ML		0.01	0.02	0.03	0.01	**0.06**	0.02	0.01	**0.03**	0.03	**0.04**	**0.09**	**0.10**	0.09	**0.14**	**0.10**	**0.10**	**0.11**	**0.09**	**0.08**	**0.08**	**0.08**	**0.15**	**0.16**
BJ			0.02	**0.04**	0.02	**0.05**	0.05	**0.02**	**0.04**	**0.04**	**0.05**	**0.09**	**0.09**	**0.08**	**0.14**	**0.11**	**0.08**	**0.09**	**0.08**	**0.08**	**0.07**	**0.08**	**0.17**	**0.20**
CM				**0.04**	0.02	**0.07**	0.08	0.03	0.03	**0.05**	**0.06**	**0.11**	**0.13**	**0.12**	**0.16**	**0.12**	**0.13**	**0.12**	**0.11**	**0.10**	**0.09**	**0.11**	**0.17**	**0.22**
ET					**0.03**	**0.06**	**0.03**	**0.03**	**0.05**	**0.03**	**0.04**	**0.08**	**0.08**	**0.08**	**0.10**	**0.10**	**0.08**	**0.11**	**0.06**	**0.05**	**0.05**	**0.06**	**0.14**	**0.17**
KE						0.03	0.04	0.02	**0.03**	0.02	0.03	**0.09**	**0.10**	**0.08**	**0.13**	**0.11**	**0.09**	**0.10**	**0.08**	**0.07**	**0.07**	**0.08**	**0.18**	**0.20**
MZ							0.04	0.02	**0.03**	0.02	0.01	**0.13**	**0.13**	**0.12**	**0.18**	**0.17**	**0.11**	**0.16**	**0.11**	**0.11**	**0.11**	**0.13**	**0.23**	**0.24**
ZA								0.03	**0.04**	0.01	0.02	**0.11**	**0.09**	**0.08**	**0.12**	**0.11**	**0.09**	**0.12**	**0.08**	**0.07**	**0.08**	**0.08**	**0.17**	**0.19**
ZW									0.02	0.00	0.01	**0.10**	**0.10**	**0.09**	**0.13**	**0.11**	**0.09**	**0.11**	**0.08**	**0.08**	**0.08**	**0.08**	**0.15**	**0.17**
MU										**0.01**	0.01	**0.11**	**0.12**	**0.11**	**0.18**	**0.13**	**0.11**	**0.14**	**0.11**	**0.10**	**0.10**	**0.11**	**0.20**	**0.24**
RE											0.01	**0.10**	**0.10**	**0.09**	**0.14**	**0.12**	**0.09**	**0.13**	**0.09**	**0.09**	**0.09**	**0.10**	**0.18**	**0.21**
MG												**0.11**	**0.12**	**0.11**	**0.16**	**0.14**	**0.10**	**0.14**	**0.10**	**0.10**	**0.10**	**0.11**	**0.20**	**0.23**
DZ													0.01	0.00	**0.03**	**0.02**	0.01	0.00	0.01	**0.02**	0.00	0.01	**0.15**	**0.20**
CO														0.00	0.02	0.01	0.01	0.02	0	**0.02**	0.01	**0.01**	**0.13**	**0.16**
PO															0.03	0.03	0.00	0.01	0.02	0.01	0.01	0.01	**0.19**	**0.20**
FR																**0.05**	**0.04**	**0.05**	**0.03**	**0.06**	0.02	**0.04**	**0.20**	**0.23**
CA																	**0.05**	**0.02**	**0.05**	**0.05**	**0.03**	**0.03**	**0.14**	**0.21**
SI																		**0.04**	0.01	0.02	0.02	0.03	**0.19**	**0.20**
SD																			**0.04**	**0.04**	**0.03**	0.02	**0.16**	**0.22**
BA																				0.01	-0.01	0.00	**0.13**	**0.16**
ES																					**0.01**	0.01	**0.13**	**0.15**
MA																						0.00	**0.13**	**0.18**
PT																							**0.10**	**0.16**
GR																								**0.14**

Details on sample codes are given in Table 1. F_{ST} values are grouped according to the genetic clusters inferred by STRUCTURE v.2.3.3.: (CA) central Africa, (SA) southern Africa, (WMB) western Mediterranean basin and (EMB) eastern Mediterranean basin. Significant values, at the adjusted nominal level (5%) for multiple comparison of 0.0002, are highlighted in bold.

Table S6 Results of demographic analysis based on COI and CytB concatenated data set for each cluster.

	Tajima'D	Fu's FS	Raggedness	R2	Thau	Expansion Time (BP)
WA	-1.450	**-11.109**	**0.0135**	**0.072**	2.853	61100-172000
SEA	-0.800	**-10.483**	0.0324	0.081	3.325	71200-200000
WMB	**-1.972**	**-16.644**	0.1301	**0.030**	1.273	27300-76700
EMB	0.198	1.015	0.1472	0.154	2.391	-

Results of demographic analysis based on COI and CytB concatenated data set for each cluster. Significant values (p-value < 0.05) are highlighted in bold. (WA) western Africa; (SEA) southeastern Africa, (WMB) western Mediterranean basin, (EMB) eastern Mediterranean basin.

Table S7 Pairwise F_{ST} values based on COI mitochondrial gene between the genetic clusters inferred by BAPS v.6.0.

	southeastern Africa	western Mediterranean Basin	eastern Mediterranean Basin
western Africa	**0.3754**	**0.6843**	**0.6036**
southeastern Africa	-	**0.6543**	**0.2962**
western Mediterranean Basin	-	-	**0.7895**

Population differentiation was assessed with the exact G test implemented in ARLEQUIN v.3.5.3.1 Significant values (p-value <0.05) are highlighted in bold.

Table S8 Description of the scenarios tested by approximate Bayesian computation analyses (ABC) on microsatellite data attempting to retrace the routes of colonization of *C. imicola* with different sets of sampled populations.

		Set 1				**Set 2**				**Set 3**			
Scenarios	Description of tested scenarios	Posterior probability	95% Credibility interval	Type I Error	Type II error	Posterior probability	95% Credibility interval	Type I Error	Type II error	Posterior probability	95% Credibility interval	Type I Error	Type II error
1	Introduction out from CA independently to WMB and EMB	0.0520	[0.0039, 0.1001]			0.0739	[0.0682, 0.0796]			0.0331	[0.0237, 0.0426]		
2	Introduction out from SA independently to WMB and EMB	0.0050	[0.0000, 0.0567]			0.0094	[0.0038, 0.0149]			0.0069	[0.0000, 0.0171]		
3	Introduction out from CA to WMB and from SA to EMB	0.0064	[0.0000, 0.0580]			0.0062	[0.0006, 0.0119]			0.0074	[0.0000, 0.0175]		
4	Introduction out from SA to WMB and from CA to EMB	0.0095	[0.0000, 0.0608]			0.0278	[0.0223, 0.0332]			0.0067	[0.0000, 0.0168]		
5	Introduction out from CA to WMB then introduction from WMB to EMB	0.2821	[0.2184, 0.3459]			**0.4076**	**[0.3958, 0.4193]**	**0.55**	**0.07**	**0.3217**	**[0.3146, 0.3288]**	**0.55**	**0.06**
6	Introduction out from SA to WMB then introduction from WMB to EMB	0.0516	[0.0034, 0.0998]			0.1043	[0.0974, 0.1112]			0.1218	[0.1138, 0.1298]		
7	Introduction out from CA to EMB then introduction from EMB to WMB	**0.5094**	**[0.4616, 0.5572]**	**0.60**	**0.07**	0.2965	[0.2850, 0.3079]			**0.3534**	**[0.3459, 0.3608]**	**0.52**	**0.06**
8	Introduction out from SA to EMB then introduction from EMB to WMB	0.0840	[0.0372, 0.1307]			0.0745	[0.0685, 0.0804]			0.1491	[0.1366, 0.1616]		

Confidence in scenario selection based on probabilities posterior, 95% confidence intervals and type I and II errors are given. 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.

Set 2: Senegal, Mauritius, Algeria, Greece.

Set 3: Mali, Mozambique, Italy-Calabria, Israel.

Set 4: Benin, France-Réunion Island, Spain-Andalusia Israel.

Table S9 ABC model checking of the most probable scenario (Scenario 5 and 7) using all the summary statistics (used and unused for model selection).

Scenario 5			**Scenario 7**		
summary statistics	observed value	proportion (simulated<observed)	summary statistics	observed value	proportion (simulated<observed)
NAL_1_1	8.5556	0.7610	NAL_1_1	8.5556	0.6900
NAL_1_2	8.7778	0.7360	NAL_1_2	8.7778	0.6960
NAL_1_3	6.0000	0.7165	NAL_1_3	6.0000	0.5960
NAL_1_4	5.0000	0.6550	NAL_1_4	5.0000	0.7160
HET_1_1	0.6990	0.5270	HET_1_1	0.6990	0.4690
HET_1_2	0.7028	0.5270	HET_1_2	0.7028	0.4900
HET_1_3	0.6174	0.6950	HET_1_3	0.6174	0.5585
HET_1_4	0.4922	0.3300	HET_1_4	0.4922	0.4055
VAR_1_1	2.9861	0.6420	VAR_1_1	2.9861	0.6060
VAR_1_2	2.7215	0.6095	VAR_1_2	2.7215	0.5555
VAR_1_3	1.3626	0.4430	VAR_1_3	1.3626	0.3610
VAR_1_4	4.0022	0.8640	VAR_1_4	4.0022	0.8620
MGW_1_1	1.0476	0.4980	MGW_1_1	1.0476	0.5085
MGW_1_2	1.1820	0.7125	MGW_1_2	1.1820	0.7195
MGW_1_3	0.9789	0.5600	MGW_1_3	0.9789	0.5410
MGW_1_4	0.8359	0.3235	MGW_1_4	0.8359	0.3270
N2P_1_1&2	11.5556	0.7235	N2P_1_1&2	11.5556	0.6770
N2P_1_1&3	9.6667	0.6985	N2P_1_1&3	9.6667	0.6250
N2P_1_1&4	10.0000	0.7885	N2P_1_1&4	10.0000	0.7445
N2P_1_2&3	10.2222	0.6890	N2P_1_2&3	10.2222	0.6325
N2P_1_2&4	10.5556	0.7710	N2P_1_2&4	10.5556	0.7635
N2P_1_3&4	7.4444	0.7030	N2P_1_3&4	7.4444	0.6615
H2P_1_1&2	0.7229	0.4680	H2P_1_1&2	0.7229	0.4270
H2P_1_1&3	0.6831	0.5475	H2P_1_1&3	0.6831	0.4475
H2P_1_1&4	0.6436	0.3825	H2P_1_1&4	0.6436	0.3720
H2P_1_2&3	0.6994	0.5235	H2P_1_2&3	0.6994	0.4430
H2P_1_2&4	0.6786	0.4135	H2P_1_2&4	0.6786	0.4170
H2P_1_3&4	0.6006	0.4985	H2P_1_3&4	0.6006	0.4600
V2P_1_1&2	2.9708	0.5950	V2P_1_1&2	2.9708	0.5340
V2P_1_1&3	2.1833	0.4835	V2P_1_1&3	2.1833	0.4420
V2P_1_1&4	3.7574	0.7550	V2P_1_1&4	3.7574	0.7425
V2P_1_2&3	2.1390	0.4380	V2P_1_2&3	2.1390	0.3720
V2P_1_2&4	3.8538	0.7425	V2P_1_2&4	3.8538	0.7225
V2P_1_3&4	2.9466	0.7290	V2P_1_3&4	2.9466	0.7070
FST_1_1&2	0.0555	0.0845	FST_1_1&2	0.0555	0.0970
FST_1_1&3	0.0718	0.1315	FST_1_1&3	0.0718	0.2445
FST_1_1&4	0.1371	0.5005	FST_1_1&4	0.1371	0.3845
FST_1_2&3	0.1034	0.1175	FST_1_2&3	0.1034	0.2040
FST_1_2&4	0.2021	0.6240	FST_1_2&4	0.2021	0.5230
FST_1_3&4	0.1337	0.4775	FST_1_3&4	0.1337	0.5580
LIK_1_1&2	1.5859	0.2685	LIK_1_1&2	1.5859	0.2450
LIK_1_1&3	1.6015	0.2190	LIK_1_1&3	1.6015	0.2675

LIK_1_1&4	1.9582	0.4505	LIK_1_1&4	1.9582	0.3375
LIK_1_2&1	1.7205	0.3080	LIK_1_2&1	1.7205	0.2760
LIK_1_2&3	1.9136	0.1900	LIK_1_2&3	1.9136	0.2165
LIK_1_2&4	2.4508	0.4675	LIK_1_2&4	2.4508	0.3755
LIK_1_3&1	1.3045	0.1950	LIK_1_3&1	1.3045	0.1795
LIK_1_3&2	1.6174	0.2305	LIK_1_3&2	1.6174	0.1690
LIK_1_3&4	1.6568	0.5805	LIK_1_3&4	1.6568	0.4790
LIK_1_4&1	1.4441	0.3175	LIK_1_4&1	1.4441	0.2765
LIK_1_4&2	1.8322	0.3630	LIK_1_4&2	1.8322	0.3165
LIK_1_4&3	1.3484	0.3780	LIK_1_4&3	1.3484	0.4875
DAS_1_1&2	0.2619	0.6190	DAS_1_1&2	0.2619	0.6635
DAS_1_1&3	0.2919	0.5995	DAS_1_1&3	0.2919	0.6375
DAS_1_1&4	0.3122	0.6665	DAS_1_1&4	0.3122	0.7040
DAS_1_2&3	0.2658	0.6375	DAS_1_2&3	0.2658	0.6785
DAS_1_2&4	0.2508	0.5785	DAS_1_2&4	0.2508	0.6355
DAS_1_3&4	0.3593	0.5450	DAS_1_3&4	0.3593	0.5760
DM2_1_1&2	0.5893	0.3700	DM2_1_1&2	0.5893	0.3450
DM2_1_1&3	0.1277	0.0565	DM2_1_1&3	0.1277	0.0565
DM2_1_1&4	1.1481	0.6745	DM2_1_1&4	1.1481	0.5715
DM2_1_2&3	0.3939	0.1205	DM2_1_2&3	0.3939	0.1315
DM2_1_2&4	2.2678	0.7230	DM2_1_2&4	2.2678	0.6515
DM2_1_3&4	1.2269	0.7330	DM2_1_3&4	1.2269	0.7380

For each statistic, the observed value and probability (simulated < observed) are given.

Chapitre II : Colonisation du sud de l'Europe par *C. imicola* : rôle de la dispersion

Toutes les espèces ont le potentiel d'élargir leur aire de distribution (Wilson *et al.* 2009), un trait qui se révèle important compte tenu du contexte actuel de changement climatique. Mais la colonisation d'un nouveau milieu nécessite souvent le franchissement de barrières géographiques ; les stratégies de dispersion sont alors un facteur déterminant du succès de cette colonisation. La dispersion est un trait d'histoire de vie fondamental des processus d'invasions qui influence à la fois la dynamique d'expansion et la structure spatiale de la diversité génétique des populations (Bohrer *et al.* 2005). L'étude des modes et l'intensité de dispersion est donc essentiel pour prédire la capacité d'une espèce à coloniser un habitat hors de son aire d'origine et à s'étendre géographiquement.

Comme nous l'avons vu précédemment, la dispersion par le vent est décrite comme le principal moyen de transport des espèces de *Culicoides*. Cette dispersion passive ou semi-passive leur permet d'être transportées au-dessus des mers sur des distances de plusieurs centaines de kilomètres (Braverman 1991; Sellers *et al.* 1977). Ainsi le risque d'introduction d'espèces de *Culicoides* dans des environnements éloignés de leur aire originelle est très élevé. Les travaux de Mardulyn *et al.* (2013) suggèrent que *C. imicola* aurait colonisé l'Italie depuis au moins une trentaine d'années et que depuis des migrations récurrentes depuis l'Afrique du Nord contribuent à la structure spatiale des populations.

Ce deuxième chapitre s'intéresse dans une première partie à la divergence et la structure génétique des populations de l'ouest et l'est du bassin méditerranéen – groupes génétiques précédemment décrits dans le premier chapitre. De plus, en considérant les populations de l'ouest du bassin méditerranéen, nous nous sommes intéressés plus spécifiquement à décrire et comprendre l'histoire évolutive et démographique de la colonisation du sud-ouest de l'Europe par *C. imicola*. En utilisant les marqueurs génétiques présentés dans le premier chapitre, nous avons déterminé les populations sources, les routes et le scénario de colonisation à l'aide d'analyses basées sur les méthodes ABC. La dynamique de dispersion de l'espèce a été caractérisée en estimant les taux de migrations, le nombre de migrants par génération et la distance parcourue par ces migrants. Des patrons d'isolement par la distance ont été également testés pour l'ensemble des populations et sur des transects intra- et inter-clusters génétiques.

Nos résultats suggèrent que les populations ouest et est du bassin méditerranéen auraient divergées d'une population ancestrale il y a au moins 500 ans. Parallèlement, nos inférences démographiques dans l'ouest du bassin méditerranéen indiquent que *C. imicola* aurait

colonisé le sud-ouest de l'Europe via deux routes de colonisation depuis l'Afrique du Nord : (i) du Maroc vers l'Espagne et le Portugal, et (ii) de l'Algérie vers la péninsule italienne et la France (Corse et département du Var). La faible différenciation génétique obtenue est vraisemblablement la résultante d'une colonisation ancienne, datée d'au moins 200 ans, accompagnée de migration récurrente entre dèmes depuis l'évènement d'introduction. La dynamique de dispersion de l'espèce se révèle relativement importante à l'échelle intra-continentale mais aussi inter-continentale. Il est intéressant de noter que la présence de la mer n'agit pas comme une barrière absolue mais attenue cette dispersion.

Notre étude confirme la capacité de *C. imicola* à se disperser sur de longues distances, et estime pour la première fois l'intensité de ce phénomène. Une caractérisation de la contribution de la dispersion active et par rapport à la dispersion passive, ainsi qu'une estimation de la vitesse de dispersion permettraient de mieux cerner et de comprendre ces processus. Ceci est essentiel à la construction de modèles prédictifs d'expansion.

Ce chapitre fait l'objet d'un article qui sera soumis à *Molecular Ecology* après la conduite d'analyses supplémentaires.

Dispersal dynamics of the biting midge vector species Culicoides imicola (Diptera: Ceratopogonidae) in the Mediterranean Basin: role of airborne dispersal

S. JACQUET[1, 2, 3*], C. GARROS[1], H. GUIS[1], A. CHASKOPOULOU[4, §], M. DJERBAL[5, §], M. GOFFREDO[6, §], Y. GOTTLIEB[7, §], Y. LHOR[8, §], J. LUCIENTES[9, §], M. MIRANDA[10, §], N. PAGES[11, 12, §], I. PEREIRA DA FONSECA[13, §], D. W. RAMILO[13, §], ML SETIER-RIO[14, §], A. TABBABI[15, §], T. BALENGHIEN[1], T BALDET[1], C. CHEVILLON[2, 3], J. BOUYER[1], K. HUBER[1, 16]

[1] CIRAD UMR15 CMAEE 34398 Montpellier France, [2] CNRS, Université de Montpellier, UMR 5290 Maladies Infectieuses & Vecteurs-Ecologie, Génétique, Ecologie, Contrôle (MIVEGEC), [3] IRD, UR 224 MIVEGEC, BP 64501, Agropolis, 34394 Montpellier Cedex 5, [4] USDA-ARS European Biological Control Laboratory, 54623 Thessaloniki, Greece, [5] Regional Veterinary Laboratory of Draa-Ben-Kheda, Tizi-Ouzou, Algeria, [6] Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise 'G. Caporale', 64100 Teramo, Italy, [7] Koret School of Veterinary Medicine, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, [8] Office National de Sécurité Sanitaire des produits Alimentaires (ONSSA), Rabat, Morocco, [9] Departamento de Patología Animal, Facultad de Veterinaria, Universidad de Zaragoza, Zaragoza, Spain, [10] Laboratory of Zoology, University of Balearics (UIB), Palma de Mallorca, Spain, [11] Centre de Recerca en Sanitat Animal (CReSA), UAB-IRTA, Campus de la Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Spain, [12] CIRAD, UMR CMAEE, 97170 Petit Bourg, Guadeloupe, France INRA, UMR1309 CMAEE, F-34398 Montpellier, France, [13] CIISA, Faculdade de Medicina Veterinária, ULisboa, 1300-477 Lisboa, Portugal, [14] EID Méditerranée, 34184 Montpellier, France, [15] Pasteur Institute of Tunis, Lab of Medical Parasitology, Biotechnologies & Biomolecules (LR 11 IPT 06), [16] INRA UMR1309 CMAEE 34398 Montpellier France

[*] Corresponding author, [§] by alphabetical order

Keywords: approximate Bayesian computation, *Culicoides imicola*, dispersal, population genetics, microsatellites, mitochondrial genes, vector species

Running title: Dispersion of *C. imicola*

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

Abstract (184 words)

Colonizing species provide opportunities to investigate evolutionary and demographic aspects of range expansion processes. Particularly, dispersal strategies are essential mechanisms underlying spatial distribution and colonizing ability of species. In the current study, we used a combination of nuclear microsatellite markers and mitochondrial sequences to evaluate colonization dynamics and dispersal patterns of the biting midge *Culicoides imicola,* a vector species which has expanded its range in southern Europe from Northern Africa. We found that western and eastern Mediterranean populations diverged from an ancestral ghost population during the Holocene. In the western Mediterranean region, our results indicate that *C. imicola* has colonized south-western Europe at least 200 years ago using two routes of colonization. This colonization has been facilitated by windborne dispersal which also allows ongoing gene flow between spatially distant populations. However, the contribution of other factors such as climate change is not excluded. At the continental level, a combination of passive and active dispersal may contribute to the species range expansion. However, the extent to which each mode of dispersal occurs should be further investigated for a better understanding of *C. imicola* dispersal dynamics.

Introduction

Identifying factors influencing the geographical distribution of species is a fundamental question at the interface of ecology and evolution (Gaston 2003). Dispersal plays a key role in species distribution and population connectivity. However, knowledge on means and rate of dispersal is still lacking for the majority of mobile organisms. From an ecological perspective, dispersal is considered as a key life history trait for colonization success. From an evolutionary perspective, it influences the genetic and demographic structure of expanding populations and affects rates of local adaptation, speciation and extinction (Kokko & López-Sepulcre 2006). Therefore, information on species' dispersal ability and strategy is essential for understanding range expansion processes (Kokko & López-Sepulcre 2006), and population functioning and dynamics (Puth & Post 2005).

Dispersal into new habitats can occur by (i) local or regional diffusion, involving a wave of advance at the fringe of the distribution (Fisher 1937; Shigesada *et al.* 1995), (ii) by long-distance dispersal, where individuals will colonize habitats far from their native range (Kareiva *et al.* 1990) or (iii) by the combination of both short and long distance dispersal termed as stratified dispersal (Ciosi *et al.* 2011; Muirhead *et al.* 2006). Within arthropod vector species, long distance migrations followed by introductions are well documented through human-mediated activities (Abbott *et al.* 2003; Fonzi *et al.* 2015). The best documented example is probably aedine mosquitoes that have colonized the world using good trade and transportations (e.g. Anderson & Stebbins 1954; Ayres *et al.* 1999; Baumel *et al.* 2001; Beaumont & Rannala 2004). Windborne dispersal records are usually seen as accidental, but examples of long range movements of vector species using winds are known in black flies, mosquitoes and biting midges (Jacquet submitted; Reynolds *et al.* 2006). Information on the relative role of natural active (i.e. flight) and passive dispersal is crucial for predicting future spread of species and designing practical recommendations for vector control strategies, species monitoring or managing commercial trade movements.

The biting midge, *Culicoides imicola* Kieffer (Diptera: Ceratopogonidae) is a major vector species of economically important livestock viruses such as bluetongue virus affecting domestic and wild ruminants (Mellor *et al.* 2000). Originally native from sub-Saharan Africa, *C. imicola* is present today in the Middle-East (Boorman 1986; Boorman & Wilkinson 1983; Braverman & Galun 1973) with eastwards records in Asia (i.e. India (Dyce & Wirth 1983) and southern China (Yü 2005)). The species is also widely distributed in the Mediterranean

region where it has been described as an invasive species recently expanding its distribution range (Calvo *et al.* 2009; Dallas *et al.* 2003; Nolan *et al.* 2008). However, this assumption is challenged by genetic analyses that supported a relatively old presence of *C. imicola* in the Mediterranean basin (Jacquet *et al.* in press; Mardulyn *et al.* 2013)

Pattern of genetic variations of *C. imicola* populations within the Mediterranean basin have previously been investigated. All the studies indicated a clear genetic differentiation between western and eastern populations with no shared haplotypes or apparent admixture events (Calvo *et al.* 2009; Dallas *et al.* 2003; Jacquet *et al.* in press; Nolan *et al.* 2008). The authors concluded that environmental and/or geographical constraints likely limited gene flow between the two geographical groups. In contrast, despite the large geographical distances between North African and south-western Mediterranean regions, low levels of genetic variation were observed among the populations (Calvo *et al.* 2009; Dallas *et al.* 2003; Jacquet *et al.* in press; Nolan *et al.* 2008). These results allowed hypothesising that North Africa is the most likely source of south-western European populations (Calvo *et al.* 2009; Dallas *et al.* 2003; Jacquet *et al.* in press; Nolan *et al.* 2008), suggesting potential long-distance dispersal of *C. imicola* between these two areas. Long-distance dispersal over water bodies through winds is well described for *Culicoides* (Braverman & Chechik 1996; Burgin *et al.* 2013), and would explain the spread of bluetongue virus throughout southern Europe with infected midges coming from North Africa (Mellor *et al.* 2008). Classically, *Culicoides* individuals are described to actively disperse on limited distances because of their small size. However, research on *Culicoides* active dispersal indicated omnidirectional flight at a range of up to 2.21 kilometres per day (Kirkeby *et al.* 2013; Kluiters *et al.* 2015). This would enable the vector species to track new habitats with suitable conditions and maintain gene flow among settled populations. This was supported by a recent genetic study by Mardulyn *et al.* (2013) that showed that recurrent gene flow through *C. imicola* exchanged migrants accounts for the observed pattern of genetic variations in Italy. However, this study focused on a restricted area. A complete description of *C. imicola* dispersal dynamics within the Mediterranean basin underlying its geographical expansion and the current distribution of genetic variation is therefore lacking.

In this study, we investigated further the range expansion process of *C. imicola* from the North African coast to southern Europe. More specifically, we aimed to (i) extensively describe the spatial structure of neutral genetic variation of *C. imicola* populations across the Mediterranean basin, (ii) infer the historical pathways of colonization, and (iii) characterized

the dispersal dynamics associated with the range expansion success of *C. imicola*. We carried out phylogeographical and traditional population genetics analyses using a combination of microsatellite and mitochondrial sequence dataset. Moreover, we compared different scenarios of colonization aiming to test whether the colonization dynamics of *C. imicola* imply recurrent migration among populations. To that purpose, we used approximate Bayesian computation (ABC) methods (Beaumont *et al.* 2002; Bertorelle *et al.* 2010) which have been shown to successfully infer complex evolutionary scenarios in a number of studies (e.g. Ascunce *et al.* 2011; Brouat *et al.* 2014; Ciosi *et al.* 2008; Mardulyn *et al.* 2013; Miller *et al.* 2005).

Materials and Methods

Sampling, DNA extraction, microsatellite genotyping and mitochondrial sequencing

Individuals of *C. imicola* species were collected between 2003 and 2013 from 35 locations in the Mediterranean basin (see details in Table 1 and Fig. 1). Insects were sampled using black light suction traps placed near livestock or horses and were preserved in 70% ethanol. Morphological identification of *C. imicola* was then carried out under a binocular microscope using the description references (Delécolle & De La Rocque 2002).

Genomic DNA from each single midge was extracted 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. Isolated DNA from each individual was genotyped at nine microsatellite markers previously developed for *C. imicola* (Mardulyn *et al.* 2013) (Table S1) following the protocol described in (Jacquet *et al.* in press). The length of microsatellite alleles was scored manually with GeneMapper version 4.0 (AppliedBiosystems).

A subset of successful genotyped individuals (4 to 10 individuals per population) was amplified and sequenced for a portion of the mitochondrial genes cytochrome oxydase subunit I (COI, ~ 474bp) and cytochrome b (CytB, ~ 633bp) using the primers C1J1718/C1N2191 and CytB_12329F/CytB_13038R, respectively, as described in (Jacquet *et al.* in press). Sequences of COI and CytB genes were edited and aligned using a Clustal W algorithm (Thompson *et al.* 1994) available in the software Geneious v.6.0.5 (Biomatters, http://www.geneious.com).

Genetic diversity within populations (microsatellite and mtDNA)

Linkage disequilibrium between pairs of microsatellite loci and Hardy-Weinberg equilibrium for each locus, each site and across loci and sites were tested using 10,000 permutations in FSTAT v2.9.3.2 (Goudet 1995). For each site sampled with more than 15 individuals, the same software was used to compute basic measures of genetic diversity within populations, including allelic richness (A_R) and the genetic diversity (H_S).

For mtDNA analyses, the population samples were grouped in 5 geographical groups: North Africa (Algeria, Morocco, Tunisia), Iberian peninsula (Portugal, continental Spain, Balearic Islands), France-Italy (continental France, Corsica, Sardinia, Sicily, continental Italy) and Greece-Israel. Israel and Greece were grouped together owing to the limited number of samples. Patterns of genetic diversity within sampled populations were characterized for both mitochondrial COI and CytB data set, by calculating the haplotype number (H), haplotype diversity (H_d) and nucleotide diversity (π) for each site using DnaSP v.5.10 (Librado & Rozas 2009).

Population structure and genealogical relationships (microsatellite and mtDNA)

Several statistical methods were used to infer genetic variation and population structure among populations from microsatellite data. First, the spatial genetic structure was investigated using the Bayesian approach implemented in STRUCTURE v.2.3.3 (Pritchard *et al.* 2000). This method detects statistically, a number of genetic clusters and assigns individuals to a cluster according to their genotypes. The analysis was run using the admixture model and correlated alleles' frequencies (Falush *et al.* 2003), with sampling locations as priors' information (Locprior model). Ten independent runs, for each value of K (1 to 12), were performed. Each run consisted of Markov chain Monte Carlo (MCMC) of 10 million iterations after a burn-in of 10,000 steps. The most probable number of clusters was selected using the ΔK method (Evanno *et al.* 2005). The Bayesian clustering analysis was then repeated for each genetic group previously inferred to assess sub-genetic structure within group. The results were edited and displayed using DISTRUCT v.1.1 (https://web.stanford.edu/group/rosenberglab/distruct.html) (Rosenberg 2004). Second, 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 carrying out 1,000 bootstrap replicates.

The relative importance of the genetic clusters previously inferred by STRUCTURE and the populations in differentiation was assessed with the multilocus hierarchical F-statistics Fpopulations-clusters and Fclusters-total, respectively. This analysis was performed with the Hierfstat package (Goudet 2005). These tests were based on 10,000 permutations of either culicoides genotypes among populations and within clusters (H0: 'Fpopulations-cluster = 0'), or populations among clusters (H0: 'Fclusters-total = 0'). Genetic differentiation between all pairs of samples and the inferred clusters was further assessed through the Weir and Cockerham (1984)'s unbiased estimates F_{ST} and the significance was tested using the exact G test over 10,000 permutations of genotypes among samples as implemented in FSTAT v2.9.3.2 (Goudet 1995). Sites sampled with less than 15 individuals were excluded for the NJ tree reconstruction and for the population differentiation test between pairs of sites.

Spatial population structuring and genealogical relationships were further investigated using mitochondrial polymorphism. This was done by computing a median-joining network (Bandelt *et al.* 1999) based on COI and CytB concatenated haplotypes, with the program NETWORK v.4.6.1.2 (www.fluxus-engineering.com). Genetic differentiation between sites was tested with ARLEQUIN v.3.5.2, based on F_{ST} statistic (Excoffier *et al.* 2005)

Divergence of western and eastern Mediterranean populations

To assess the demographic and evolutionary history which has led to the current spatial genetic patterns of *C. imicola* between western and eastern Mediterranean regions, we compared three competing models using ABC approach (Beaumont *et al.* 2002) implemented in the software package ABCTOOLBOX v.2.0 (Wegmann *et al.* 2010). This method allows drawing inferences of complex evolutionary scenarios by incorporating prior biological and historical knowledge and using summary statistics to compare simulated and real datasets (Beaumont *et al.* 2002). Our first model supposed that western and eastern populations diverged from an ancestral unsampled population. The second model assumed an eastern origin of western populations, while the third model assumed a western origin of eastern populations (Fig. 2a).

Model selection procedure

ABC analyses were conducted on microsatellite data using the North African and Israeli populations assuming that these regions were first colonized by *C. imicola*. Analyses were performed with demographic, historical and mutational parameters drawn from the prior distributions described in Table S2. Historical parameters were scaled by assuming 10

generations per year (Braverman & Linley 1988) and a divergence time within the last 30,000 generations. The average microsatellite mutation rate prior was set to range from 6.10^{-6} to 1.10^{-4} per generation.

We used a conventional rejection sampling approach (Beaumont *et al.* 2002) in order to generate simulated data sets under a given model with parameters drawn from the prior distributions. Summary statistics are used to compare simulated and observed data sets. We defined a set of summary statistics calculated within and between populations for both simulated and real data. In all ABC runs, we computed in each population the mean number of alleles, the mean expected heterozygosity (Nei 1987), and the mean allelic size variance. The same statistics were computed over loci and populations. We additionally calculated pairwise F_{ST} values (Weir & Cockerham 1984) and the genetic distance $(\delta\mu)^2$ between all pairs of populations (Berkvens *et al.* 2010). The program SIMCOAL 2.0 and ARLSUMSTAT (Laval & Excoffier 2004) were used to simulated genetic data and calculate summary statistics, respectively. A total of 5,000 simulations closest to the observed data were then retained for a postsampling regression adjustment (ABC-GLM; Leuenberger *et al.* 2009) as implemented in the software package ABCTOOLBOX v.2.0. Models were compared and discriminated by calculating the Bayes factor and probabilities for each model.

Validation of the estimation procedure

Model selection and parameter estimation under an ABC framework can suffer from the loss of information when the observed and simulated data are limited to summary statistics (Robert *et al.* 2011). Therefore, an evaluation of ABC performances in model choice and parameter estimation is essential. To that purpose, we validated our ABC analysis with three different procedures.

We first carried out a principal component analysis to check the position of the observed data compared to the 5,000 simulations closest to it under the selected model. We further evaluated the robustness of our model selection by investigating the rate of true and false positives for each pairwise model comparison using pseudo-observed data sets (*pods*), whereby parameters are randomly drawn from the prior distributions under a given model. We generated 1,000 *pods* under each model, using the same prior distributions. Each *pods* was then analysed as if it was real observed data, estimating the posterior probability of each model using the ABC-GLM regression method. This allows validating our model selection and parameter estimation as these *pods* were generated under a known model and corresponding parameters. For each model, we assumed the rate of true positives as the fraction of pods correctly assigned to the

model that generated them, and the false positives as the number of pods assigned to a model other than the one that generated them.

Finally, we generated by simulation, with parameter value extracted from their posterior distribution under the best model, the posterior distribution of four summary statistics not used in the model choice analysis and performed principal component analysis to graphically show the position of the real data set.

Colonisation history and dispersal dynamics in the western Mediterranean region

Colonisation history of south-western Europe

The demography history and routes of colonisation followed by *C. imicola* in south-western Europe (i.e. regions from Portugal to Italy) were tested using ABCTOOLBOX v.2.0. The same model selection and validation procedures as well as the set of statistics described previously were used.

According to the standard genetic analyses outputs, i.e. the microsatellite neighbor-joining tree and the genetic structure inferred by STRUCTURE v.2.3.3 (see results), we investigated the possible colonisation scenarios which could generate the observed genetic variability. Three models (Fig. 2b) were compared assuming the same pattern of population divergence, but exhibiting different migration histories. Looking forward, at T_a, the North African populations (Morocco and Algeria) diverged from an ancestral unsampled population. Then at T_w and T_c, the Iberian Peninsula and France-Italy regions were colonized by a few number of founding individuals originated from Morocco and Algeria, respectively. The effective size of the two newly created populations increased to their current size effective size at T_w - D_{Bw} and T_c - D_{Bc} with D_B the duration of bottleneck. The first model (M1B) assumed a recent colonisation within the last 600 generations, the second one (M2B) supposed an ancient colonisation within the last 30,000 generations and a recent occurrence of migration between populations which started 600 generations ago, and the last model (M3B) allowed for constant migration between populations since the colonisation event which arisen within the last 30,000 generations. Model M2B and M3B assumed two rates of migration, whereby the rate of migration within North Africa and between Morocco and the Iberian Peninsula is greater than the rate of long-distance migration over water bodies. The migration rate prior was set to range from 10^{-6} to 1.10^{-4}.

ABC analyses were conducted on microsatellite data. Sampled locations belonging to the same country were pooled as a unique population. One representative pooled population per regional group (i.e. Morocco, Iberian Peninsula, France-Italy and Algeria) was used to test

our models. These analyses were replicated by using other pooled populations belonging to the same regional group as replicates of the same evolutionary history.

Dispersal dynamics

Patterns and rate of dispersal of *C. imicola* populations in the western Mediterranean region were investigated using different approaches.

As migration is inclined to occur more usually between neighbouring populations, a correlation between genetic and geographic distance, so called isolation by distance (IBD), is expected in many systems (Wright 1943, 1946). The IBD method uses a regression of the genetic distance between populations as a function of geographic distance. We performed this method for (i) the whole set of samples throughout the western Mediterranean region, (ii) each genetic sub-cluster inferred by the software STRUCTURE and (iii) the sampled populations from Portugal, Spain, France and Italy. All correlations were tested with partial Mantel tests as implemented in the online program GENEPOP v.4.2 (Raymond & Rousset 1995; Rousset 2008), and were based on the logarithmic geographic distances and the pairwise F_{ST} values. The significance of the regression was assessed with 100,000 permutations.

Historical gene flow (i.e. based on genetic structure) between pairs of populations and effective population sizes (Ne) were jointly estimated using the coalescence approach implemented in the software MIGRATE-N v.3.6.10 (Beerli & Felsenstein 2001). This software estimates the mutation-scaled effective population size $\Theta = 4N_e\mu$, where μ is the mutation rate per locus per generation, as well as mutation-scaled migration rates $M = m/\mu$. This measure quantifies the number of new variants introduced into the population by immigration (m) relative to mutation assuming migration-drift equilibrium in populations. We ran two replicates of maximum likelihood Monte Carlo Markov chain (MCMC) using one representative sample site per country, a Brownian motion mutation model with constant mutation rates for all loci and starting parameters based on F_{ST} calculations. Each run consisted of (i) ten short chains of 2,000 recorded steps, 200,000 visited genealogies including 10,000 burn-in trees and a sampling increment of 100 and (ii) three long chains of 20,000 recorded steps, a sampling increment of 100 and 2,000,000 visited genealogies including a burn-in of 10,000 trees. A static scheme was employed with four chains at temperatures of 1, 4, 8 and 12. Convergence of the MCMC runs was assessed by comparing the results of each run. As a quality control, we performed two replications of the analysis by using other sample sites belonging to the same country.

The magnitude and directionality of contemporary gene flow between population pairs and genetic cluster pairs was estimated with the program BAYESASS v.3.0.3 (Wilson & Rannala 2003). BAYESASS uses assignment tests in a Bayesian framework and a Monte Carlo Markov chain to estimate migration rates between populations over the last several generations (Wilson & Rannala 2003).

In contrast to MIGRATE-N, BAYESASS does not assume migration-drift equilibrium in populations. Pairwise comparisons were carried out for the same sample sites and genetic groups that were used in the MIGRATE-N analysis. A total of 50,000,000 MCMC iterations were run including a burn-in of 10,000,000 steps and a sampling frequency of 2,000 iterations. The default delta value of 0.15 was used for allele frequency, migration rate and inbreeding. Five replicates runs at a different starting point were performed and the program TRACER v.1.6 (Rambaut A 2014) was used to qualitatively assess MCMC convergence.

We further investigated the dispersal dynamics of *C. imicola* populations within the western Mediterranean region by calculating the number of immigrants per generation and the dispersal distances between reproducing midges and their parents as suggested by Bouyer *et al.* (2009). To that purpose, we first estimated the densities of *C. imicola* through the N_e values obtained from MIGRATE-N analysis. We used the equation $D_e = N_e / D_u$, where D_u is the surface area of *C. imicola* distribution. We used different values of D_u ranging from 1 km to 5 km. Dispersal distances σ (in kilometres) between reproducing midges and their parents were then estimated from the regression slopes (b) obtained from the isolation-by-distance analysis and the equation $\sigma = \sqrt{D\sigma^2 / D_e}$ with $D\sigma^2 = 1/4\pi b$, D being the effective population density (Rousset 1997). We finally computed the number of immigrants per generation for each population using the regression slope with $Nm = 1/2\pi b$.

Results

Population structure and genetic diversity (microsatellite and mtDNA)

A total of 835 *C. imicola* individuals sampled from 35 sites were genotyped at nine microsatellite loci (Table 1). None of the pairwise allelic tests rejected the null hypothesis of linkage equilibrium and all the sampled populations were in Hardy-Weinberg equilibrium.

A clear geographical structure was observed among *C. imicola* Mediterranean populations from both microsatellite and mtDNA dataset. At a global scale, the Bayesian clustering performed with STRUCTURE on microsatellite data confirmed the genetic differentiation between western and eastern Mediterranean regions (for K = 2; Fig. 1b), previously described

(Dallas *et al.* 2003; Jacquet *et al.* in press; Nolan *et al.* 2008). Within the western Mediterranean region, ΔK was clearly maximum for K = 2 and supported a sharp genetic differentiation between west and central Mediterranean populations, defining the WMB (Morocco, Spain and Portugal) and CMB (Algeria, Italy and France) sub-clusters (Fig. 1c). Within eastern Mediterranean region, ΔK was greater for K = 2 suggesting two geographical groups: Greece *versus* Israel populations (Fig. 1d). All these genetic clusters were consistently retrieved in the neighbour-joining tree based on microsatellite markers (Fig. 3a).

Based on a total of 257 sequences, the mtDNA haplotype network also confirmed the west/east geographical structure of *C. imicola* populations (Fig. 3b). Of the 51 haplotypes obtained from the COI and CytB concatenated dataset, none were shared between western and eastern Mediterranean regions. However, the genetic differentiation observed within in each region from the microsatellite markers was less clear than from mitochondrial data. A predominant haplotype was found to be distributed throughout the western Mediterranean region with strong genetic relationships between North Africa, Portugal, Spain, France and Italy. Nevertheless, we could notice (i) the occurrence of two haplotypes predominantly found in Algeria, France, and Italy (corresponding to the CMB sub-cluster inferred from microsatellite data) and (ii) no shared haplotypes between Greece and Israel (Fig. 4), thus reminding the sub-structure observed from microsatellite clustering analyses.

In the western Mediterranean region, microsatellite-based pairwise F_{ST} values ranged from -0.0001 to 0.0660 within and between the western and central sub-clusters (Table S3). F_{ST} values were relatively higher within eastern Mediterranean region and extended to 0.2385 (Table S3). Likewise, pairwise F_{ST} values based on mitochondrial concatenated data were relatively lower within the western Mediterranean region ($F_{ST} \leq 0.37$), compared with the eastern region ($0.61 \leq F_{ST} \leq 0.95$) (Table S4). Genetic differentiation tests were significant for all pairwise comparisons between clusters inferred by STRUCTURE. Considering the hierarchy of sampling, significant differentiation was detected between the two genetic clusters (Fcluster-total = 0.065; P = 0.0001) but also within clusters (Fpopulations-clusters = 0.016; P = 0.0001).

The allelic richness (A_R) varied from 3.232 to 4.919 alleles per locus and genetic diversity (H_S) ranged from 0.492 to 0.652. Both of these measures were comparable among populations (Table S5). Likewise, levels of genetic diversity computed from COI and CytB gene were equivalent among populations, with a lower value for the Greek population (Table S6).

Divergence history of western and eastern Mediterranean populations (microsatellite)

The ABC analyses clearly supported scenario #1 (Fig. 2, scenario M1A) which allowed a divergence of western and eastern populations from an ancestral unsampled population (p = 0.9999 and Bayes factor = 1939). The two other competing scenarios were not supported at all with probabilities of p = 0.0005, Bayes factor = 0.0005 and p = $2.8\ 10^{-6}$, Bayes factor = $2.8\ 10^{-6}$, for scenario M2A and M3A respectively (Fig. 2).

Checking the adequacy of the selected model to generate data sets similar to the observed data using a PCA showed that points simulated from the posterior distribution were relatively well centered on the target corresponding to the real dataset (Fig. S1a). Type I and II errors rate of the selected scenario were moderately high with a mean value of 0.23 and 0.33, respectively.

Colonization history and dispersal dynamics in the western Mediterranean region

Colonization history in south-western Europe

Testing the scenarios of *C. imicola* colonization in south-western Europe provided higher posterior probabilities for scenario #3 (Fig. 2, scenario M3B) which involved recurrent migration occurring between all the sampled populations (p = 1, Bayes factor = $1.2\ 10^{+19}$). The two other models (Fig. 2, scenario M1B and M2B), were clearly rejected with low probabilities (p = $1.8\ 10^{-21}$, Bayes factor = $1.8\ 10^{-21}$ for M1B and p = $8.5\ 10^{-20}$, Bayes factor = $8.5\ 10^{-20}$ for scenario M2B). The same results were obtained for the replicate run using different sampled site (p = $1.5\ 10^{-28}$, Bayes factor = $1.5\ 10^{-28}$ for M1B and p = $3.1\ 10^{-25}$, Bayes factor = $3.1\ 10^{-25}$ for scenario M2B, p = 1, Bayes factor = $3.3\ 10^{+24}$ for M3B)

As previously, the PCA analysis indicated that the simulated data fitted well the observed data (Fig. S1b). The analysis of true and false positives based on *pods*, suggested that the ABC approach and the set of summary statistics that we used have reasonable power to discriminate model M1 and M3. The rate of type I and II errors of the selected scenario #3 (Fig. 2, scenario M3B) were estimated with a mean value of 0.45 and 0.06, respectively. The high rate of false positive was mostly associated with the scenario M2B, which involved an ancient colonization with recent migration.

Dispersal dynamics

According to the most probable scenario of *C. imicola* colonization in south-western Europe, which involves recurrent migration between populations, we assessed the dispersal dynamics of *C. imicola* in the western Mediterranean region. A significant correlation between the genetic and geographical distances was found when considering distant sample sites (≥ 2000

km) separated by large bodies of water (the Mediterranean sea). Indeed, an isolation by distance pattern was found for the whole dataset (r = 0.463, P = 0.0001), the central Mediterranean sub-cluster (CMB, r = 0.467, P = 0.0005) and between the Iberian Peninsula and Italy-France populations (r = 0.637, P = 0.0027). However, the partial Mantel test was not significant for the western sub-cluster (WMB, r = 0.108, P = 0.3159), which included intra-continental sites distant from less than 1200 km and separated by a narrow body of water (Strait of Gibraltar).

Using the coalescent approach implemented in MIGRATE, the long term migration rates between populations were estimated to range from 4.10^{-5} to 4.10^{-4} assuming a mutation rate of 6.10^{-6} per generation (similar to the mutation rates estimated for Drosophila, Schlötterer *et al.* 1998) (Table 2). No asymmetric gene flow was found between populations (Table 2). The BAYESASS analysis indicated ongoing gene flow among populations (Table 3).

Mutation-scaled effective population sizes of *C. imicola* were comparable among sampled sites, with a lower value in continental France and higher estimate in Sardinia, and extended from 0.090 to 0.618 (Table 4). The number of immigrants (Nm) inferred from the Mantel regression slopes was 25 immigrants per generation into each population when taking into account the whole dataset. Using the effective population sizes and the $D\sigma^2$ values computed from the isolation-by-distance approach, dispersal distances between reproducing individuals and their parents were estimated to range from 22 m to 529 m when taking into consideration all the samples.

Discussion

Colonization events provide unique opportunities to examine the dynamic and evolutionary processes associated with range expansion (Bock *et al.* 2015; Lee 2002; Sax *et al.* 2007). This study investigated the processes behind the current genetic distribution of *C. imicola* populations within the Mediterranean basin. Combining a phylogeographical approach with population genetics and approximate Bayesian computation methods allowed us to infer the divergence history of *C. imicola* western and eastern Mediterranean populations and characterize the colonization dynamics of the species in south-western Europe.

Divergence history of western and eastern population within the Mediterranean basin

Previously, Jacquet *et al.* (in press) showed that *C. imicola* has colonized the Mediterranean basin through a single colonization event from sub-Saharan Africa. However, the subsequent processes underlying the observed spatial genetic structure in the Mediterranean region had

not been fully characterized. The current study provides further information on the demographic history of *C. imicola* within the Mediterranean basin. We confirmed the western and eastern geographical groups previously discovered (Calvo *et al.* 2009; Dallas *et al.* 2003; Jacquet *et al.* in press; Nolan *et al.* 2008). These geographical groups have most likely diverged from an ancestral population during the Holocene (at least 200 years ago), as suggested by the ABC analyses. This scenario is consistent with *C. imicola* colonization history described by (Jacquet submitted): once established, the population originating from sub-Saharan Africa could have splitted into two genetic differentiated groups. The factors that have driven this divergence are not fully understood. However, given the observed genetic patterns, one could hypothesize that environmental factors including climatic variations during the Holocene period, habitat fragmentation and physical barriers could have contributed to their genetic differentiation. No shared haplotypes or admixture events were detected suggesting that gene flow between western and eastern populations is restricted. Nevertheless, it remains unclear if these populations are totally disconnected or jointed in a not prospected contact zone (such as Egypt or Libya). A broad description and understanding of western and eastern population structuring would thus require a complete sampling in the eastern part of the Mediterranean region. This would also help to define the colonization history of south-eastern Europe by *C. imicola*, and characterize the process responsible of the observed high levels of genetic differentiation between south-eastern and Middle-East populations.

A long time presence of C. imicola *in south-western Europe*

Genetic evidences confirm a North African origin of south-western European populations (Calvo *et al.* 2009; Dallas *et al.* 2003; Nolan *et al.* 2008), and support two routes of colonization: starting from Morocco to the Iberian Peninsula and from Algeria to France and Italy. Several studies have shown that considerable amounts of genetic diversity can be retained in newly colonized areas owing to high gene flow between populations (Blackburn *et al.* 2015; Blair & Hufbauer 2010; Lockwood *et al.* 2009; Roman & Darling 2007). In our study case, genetic diversity was comparable among *C. imicola* populations and low levels of genetic differentiation were obtained between the source and colonizing populations. Whilst these patterns could emanate from a recent colonization event as suggested by the entomological survey data, our ABC analysis supported an ancient colonization (i.e. at least 200 years ago) with ongoing gene flow between populations.

Nevertheless, a high rate of false positive (0.45) was obtained for the selected scenario. This high value of type I error was mostly associated to the model admitting the same timing of colonization but assuming recent gene flow between populations. This limitation of model validation may due to the fact that these models are relatively close, and there may not be enough power in the data to discriminate between both scenarios using our ABC approach. Nevertheless, type I and II errors were almost never (0.02 and 0.04) associated with the scenario involving a recent colonization event without migration. This provides confidence in our ABC inferences although the migration patterns and chronology required further investigations.

If the observed genetic pattern result from an ancient colonization, what could explain the contrast between genetic and entomological survey data?

First, the available entomological survey data may be biased as no extensive surveys were carried out prior to bluetongue outbreaks. Even though punctual entomological surveys for *C. imicola* had been undertaken since the 70's, the sampling strategies (i.e. type of traps, period of the year, collection sites) could have impacted the outcome of the trappings as shown by (Goffredo *et al.* 2003). Thus, one could not state with confidence that *C. imicola* was only present at the time of the first records in south-western Europe. The species could have been present for a longer time before the emergence of bluetongue in these territories without being detected by entomological surveys.

Second, although our genetic inferences suggest an ancient colonization, we cannot dismiss the potential role of contemporary global warming in the range expansion of *C. imicola*, as initially admitted by Purse *et al.* (2005). Indeed, the global increase of temperature influences insects' development, reproduction, survival, and population dynamics (Bossdorf *et al.* 2005). One could hypothesize that the elevation of temperature has allowed an increase of population abundance and/or the regional expansion of *C. imicola* established populations. Complementarily, the increase in temperature could have impacted the dispersal dynamics of *C. imicola*, by inducing new migration events (as suggested by the scenario M2B) or by increasing their intensity owing to more suitable climatic conditions for windborne transport.

High ability of C. imicola *to disperse over long-distance*

Long-distance windborne dispersal is considered as the primary means of transport of *Culicoides* (Mellor *et al.* 2000), when anthropogenic transportation have been shown to have a limited role in *Culicoides* movement (de Vos *et al.* 2012; Napp *et al.* 2013; Nie *et al.* 2005;

Reye 1964). The extent to which *Culicoides* windborne dispersal occurs has previously been investigated using modeling (e.g. Ducheyne *et al.* 2007; Hendrickx *et al.* 2008); however this has rarely been done with molecular markers. To our knowledge, this current study is the first assessing migration patterns of *C. imicola* using polymorphic markers. Gene flow was detected amongst all the sampled populations, with slightly higher emigration rates in North Africa which could be due to prevalent winds in this direction. These results confirm the ability of *C. imicola* to passively disperse through winds over hundreds of kilometers and highlight the significant role of windborne dispersal in colonization of remote habitats and population dynamics, as it has been shown for other invasive species (e.g. Lander *et al.* 2014). As certain prevailing winds occur in many parts of the world, wind directions should be taken into account to predict the spread of invasive species using this mode of dispersal (Kellogg & Griffin 2006) and/or to relate patterns of genetic differentiation to aerial dispersal pathways.
Significant isolation-by-distance patterns were detected in populations disconnected by large water bodies (i.e. Algeria, France and Italy). This supports a stepping-stone model of migrations between distant locations, and thus confirms that sea bodies separating the populations does not act as an absolute barrier to gene flow. Nonetheless, it is noteworthy that successful long-distance dispersal may not be frequent enough to homogenize the genetic composition of geographically distant populations. On the other hand, other demographic and evolutionary process, such as demographic bottlenecks are involved in range expansions. Therefore, the intensity of isolation-by-distance patterns will depend on the balance between gene flow and genetic drift (Cassey *et al.* 2004).

Inland windborne dispersal of *Culicoides* has been understudied owing to both the confounding movements of ruminants and to intrinsic difficulties in modeling wind movements over topography (Braverman & Chechik 1996; Ducheyne *et al.* 2007). Thus, most of the studies focused on active dispersal patterns of *Culicoides* on land masses (Caswell *et al.* 2003; Kirkeby *et al.* 2013; Kluiters *et al.* 2015). The Mantel test failed to detect a pattern of isolation by geographic distance in the continental populations of *C. imicola* (i.e. Iberian Peninsula). Moreover, these populations were dimly differentiated whilst they were 900 km apart. These genetics patterns support the existence of gene flow amongst populations, and thus suggest that the species highly disperses on land masses, probably through a combination of short and long-distance movement. It appears that *C. imicola* continental populations may use a more complex strategy of dispersal than a simple diffusion model. However, the relative

contribution of each mode of dispersal is still unclear. On the other hand, we cannot exclude potential fine-scale genetic structure that our data could not detect.

Taken together, our results suggest that *C. imicola* likely disperses through a combination of long-distance wind-borne dispersal and active flight, of which the amount active flight needs to be quantified. It is worth noting that these results should be further explored as potential effects of environmental factors on dispersal patterns and population structure, such as landscape (e.g. rivers, relief and habitat fragmentation) were not taken into account. It would be interesting to address these points using combined genetics and geospatial approaches such as landscape genetics (Manel *et al.* 2003). Furthermore, our genetic inferences could be completed by studying dispersal patterns of *C. imicola* using direct methods such as capture-mark-release-recapture. Indeed, although genetic tools have been shown to be powerful to infer dispersal pattern (Bouyer *et al.* 2009), genetic-based estimations of distance dispersal inform about the effective dispersal, that is, gene flow resulting from dispersal followed by successful reproduction. Yet, all the migrants will not necessarily reproduce. Thus, the inferred dispersal may not reflect the total extent of dispersal dynamics. A combination of indirect and direct methods has been shown to provide valuable information on species dispersal patterns (Bouyer *et al.* 2009).

Conclusion

Genetic data confirmed the existence of a geographical subdivision between western and eastern Mediterranean populations which have likely diverged from an ancestral population. In the western Mediterranean region, our findings show that *C. imicola* populations have colonized south-western Europe several hundred years ago thanks to long-distance windborne dispersal. However, the role of contemporary climate warming is still unclear. This study also highlights that established populations maintain high gene flow through exchanged migrants. A combination of local continuous and long-distance discontinuous dispersal may have contributed to *C. imicola* range expansion. Now, the questions are in what extent does each mode of dispersal (i.e. active short-distance and passive long-distance dispersal) contribute to *C. imicola* geographical expansion, and how do environmental factors influence dispersal dynamics.

Acknowledgements

We wish to express our sincere thanks to R. Rivallan for genotyping assistance. This work was supported by the high performance cluster of the UMR AGAP Cirad. The authors would like to thank the people that helped with *Culicoides* trappings in France, Spain, Portugal, Northern Africa, Israel and Greece (EID-Med agents, farmers, Catherine Cêtre-Sossah (Cirad), Jean-Claude Delécolle (University of Strasbourg)), and *Culicoides* identification (Xavier Allène (Cirad), Ignace Rakotoarivony (Cirad), Jean-Claude Delécolle (University of Strasbourg)). 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 VmergeXX (http://www.vmerge.eu).

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.

Author contributions

SJ, CG, and KH designed the study. SJ genotyped the samples. SJ, CG and KH analyzed the data. AC, MD, MG, YG, YL, JL, MM, NP, P-D-FI, DR, M-LS-R, TA collected and/or identified the *C. imicola* samples. HG, TB, ThB, CC and JB contributed to the manuscript firstly written by SJ, CG and KH. All authors read and commented the final manuscript version.

References

Abbott RJ, James JK, Milne RI, Gillies ACM (2003) Plant introductions, hybridization and gene flow. *Philosophical Transactions of the Royal Society B: Biological Sciences* **358**, 1123-1132.

Anderson E, Stebbins G (1954) Hybridization as an evolutionary stimulus. *Evolution* **8**, 378-388.

Ascunce MS, Yang CC, Oakey J, *et al.* (2011) Global invasion history of the fire ant *Solenopsis invicta*. *Science* **331**, 1066-1068.

Ayres DR, Garcia-Rossi D, Davis HG, Strong DR (1999) Extent and degree of hybridization between exotic (*Spartina alterniflora*) and native (*S. foliosa*) cordgrass (Poaceae) in California, USA determined by random amplified polymorphic DNA (RAPDs). *Molecular Ecology* **8**, 1179-1186.

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Baumel A, Ainouche M, Levasseur J (2001) Molecular investigations in populations of *Spartina anglica* CE Hubbard (Poaceae) invading coastal Brittany (France). *Molecular Ecology* **10**, 1689-1701.

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, 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 of the United States of America* **98**, 4563-4568.

Berkvens N, Moens J, Berkvens D, *et al.* (2010) *Dinocampus coccinellae* as a parasitoid of the invasive ladybird Harmonia axyridis in Europe. *Biological control* **53**, 92-99.

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.

Blackburn TM, Lockwood JL, Cassey P (2015) The influence of numbers on invasion success. *Molecular Ecology* **24**, 1942-1953.

Blair AC, Hufbauer RA (2010) Hybridization and invasion: one of North America's most devastating invasive plants shows evidence for a history of interspecific hybridization. *Evolutionary Applications* **3**, 40-51.

Bock DG, Caseys C, Cousens RD, *et al.* (2015) What we still don't know about invasion genetics. *Molecular Ecology* **24**, 2277-2297.

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.

Bossdorf O, Auge H, Lafuma L, *et al.* (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia* **144**, 1-11.

Bouyer J, Balenghien T, Ravel S, *et al.* (2009) Population sizes and dispersal pattern of tsetse flies: rolling on the river? *Molecular Ecology* **18**, 2787-2797.

Braverman Y, Chechik F (1996) Air streams and the introduction of animal diseases borne on *Culicoides* (Diptera, Ceratopogonidae) into Israel. *Scientific Technical Review* **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. *Molecular Ecology* **23**, 4153-4167.

Burgin LE, Gloster J, Sanders C, *et al.* (2013) Investigating incursions of bluetongue virus using a model of long-distance *Culicoides* biting midge dispersal. *Transboundary and Emerging Diseases* **60**, 263-272.

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.

Cassey P, Blackburn TM, Sol D, Duncan RP, Lockwood JL (2004) Global patterns of introduction effort and establishment success in birds. *Proceedings of the Royal Society of London B: Biological Sciences* **271**, S405-S408.

Caswell H, Lensink R, Neubert MG (2003) Demography and dispersal: life table response experiments for invasion speed. *Ecology* **84**, 1968-1978.

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. *The American Journal of Human Genetics* **19**, 233-257.

Ciosi M, Miller N, Toepfer S, Estoup A, Guillemaud T (2011) Stratified dispersal and increasing genetic variation during the invasion of Central Europe by the western corn rootworm, *Diabrotica virgifera virgifera*. *Evolutionary Applications* **4**, 54-70.

Ciosi M, Miller NJ, Kim KS, *et al.* (2008) Invasion of Europe by the western corn rootworm, *Diabrotica virgifera virgifera*: multiple transatlantic introductions with various reductions of genetic diversity. *Molecular Ecology* **17**, 3614-3627.

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. *Medical and Veterinary Entomology* **17**, 379-387.

de Vos CJ, Hoek CA, Nodelijk G (2012) Risk of introducing African horse sickness virus into the Netherlands by international equine movements. *Preventive Veterinary Medicine* **106**, 108-122.

Delécolle 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.

Ducheyne E, De Deken R, Becu S, *et al.* (2007) Quantifying the wind dispersal of *Culicoides* species in Greece and Bulgaria. *Geospatial Health* **1**, 177-189.

Dyce A, Wirth W (1983) Reappraisal of some Indian *Culicoides* species in the subgenus Avaritia (Diptera: Ceratopogonidae). *International Journal of Entomological Research* **25**, 221-225.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **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**, 47-50.

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.

Fisher RA (1937) The wave of advance of advantageous genes. *Annals of Eugenics* **7**, 355-369.

Fonzi E, Higa Y, Bertuso AG, Futami K, Minakawa N (2015) Human-Mediated Marine Dispersal Influences the Population Structure of *Aedes aegypti* in the Philippine Archipelago. *PLoS Neglected Tropical Diseases* **9**, e0003829.

Gaston KJ (2003) *The structure and dynamics of geographic ranges* Oxford University Press, New York, United States of America.

Goffredo M, Conte A, Cocciolito R, Meiswinkel R (2003) The distribution and abundance of *Culicoides imicola* in Italy. *Vetirinaria Italiana* **47**, 22-32.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.

Hendrickx G, Gilbert M, Staubach C, *et al.* (2008) A wind density model to quantify the airborne spread of *Culicoides* species during north-western Europe bluetongue epidemic, 2006. *Preventive Veterinary Medicine* **87**, 162-181.

Jacquet S (submitted).

Jacquet S, Garros C, Lombaert E, *et al.* (in press) Colonization of the Mediterranean Basin by the vector biting midge species *Culicoides imicola*: an old story. *Molecular Ecology*.

Kareiva P, Mullen A, Southwood R (1990) Population dynamics in spatially complex environments: theory and data [and discussion]. *Philosophical Transactions of the Royal Society B: Biological Sciences* **330**, 175-190.

Kellogg CA, Griffin DW (2006) Aerobiology and the global transport of desert dust. *Trends in Ecology & Evolution* **21**, 638-644.

Kirkeby C, Bodker R, Stockmarr A, Lind P, Heegaard PM (2013) Quantifying dispersal of european culicoides (Diptera: Ceratopogonidae) vectors between farms using a novel mark-release-recapture technique. *PLoS One* **8**, e61269.

Kluiters G, Swales H, Baylis M (2015) Local dispersal of Palaearctic *Culicoides* biting midges estimated by mark-release-recapture. *Parasites & Vectors* **8**, 54.

Kokko H, López-Sepulcre A (2006) From individual dispersal to species ranges: perspectives for a changing world. *Science* **313**, 789-791.

Lander TA, Klein EK, Oddou-Muratorio S, *et al.* (2014) Reconstruction of a windborne insect invasion using a particle dispersal model, historical wind data, and Bayesian analysis of genetic data. *Ecology and Evolution* **4**, 4609-4625.

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.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution* **17**, 386-391.

Leuenberger C, Wegmann D, Excoffier L (2009) Bayesian computation and model selection in population genetics. *Genetics* **184**, 243-252.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Lockwood JL, Cassey P, Blackburn TM (2009) The more you introduce the more you get: the role of colonization pressure and propagule pressure in invasion ecology. *Diversity and Distributions* **15**, 904-910.

Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-197.

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. *Molecular Ecology* **22**, 2456-2466.

Mellor PS, Boorman J, Baylis M (2000) *Culicoides* biting midges: their role as arbovirus vectors. *Annual Review of Entomology* **45**, 307-340.

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.

Miller N, Estoup A, Toepfer S, *et al.* (2005) Multiple transatlantic introductions of the western corn rootworm. *Science* **310**, 992.

Muirhead JR, Leung B, Overdijk C, *et al.* (2006) Modelling local and long-distance dispersal of invasive emerald ash borer *Agrilus planipennis* (Coleoptera) in North America. *Diversity and Distributions* **12**, 71-79.

Napp S, Garcia-Bocanegra I, Pagès N, *et al.* (2013) Assessment of the risk of a bluetongue outbreak in Europe caused by *Culicoides* midges introduced through intracontinental transport and trade networks. *Medical and Veterinary Entomology* **27**, 19-28.

Nei M (1987) *Molecular evolutionary genetics* Columbia University Press, New York, United States of America.

Nie WZ, Li JC, Li DX, Gratz NG (2005) A preliminary report of introduced living biting midges (Diptera: Ceratopogonidae) found on entry ships arriving at Qinhuangdao Port, China. *Medical Entomology and Zoology* **56**, 359-361.

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. *Medical and Veterinary Entomology* **22**, 340-351.

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. *Nature Reviews Microbiology* **3**, 171-181.

Puth LM, Post DM (2005) Studying invasion: have we missed the boat? *Ecology Letters* **8**, 715-721.

Rambaut A SM, Xie D & Drummond AJ (2014) Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer.

Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution* **49**, 1280-1283.

Reye EJ (1964) The problems of biting midges (Diptera: Ceratopogonidae) in Queensland. *Australian Journal of Entomology* **3**, 1-6.

Reynolds DR, Chapman JW, Harrington R (2006) The migration of insect vectors of plant and animal viruses. *Advances in Virus Research* **67**, 453-517.

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 of the United States of America* **108**, 15112-15117.

Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology & Evolution* **22**, 454-464.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**, 137-138.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219-1228.

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

Sax DF, Stachowicz JJ, Brown JH, *et al.* (2007) Ecological and evolutionary insights from species invasions. *Trends in Ecology & Evolution* **22**, 465-471.

Schlötterer C, Ritter R, Harr B, Brem G (1998) High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Molecular Biology and Evolution* **15**, 1269-1274.

Shigesada N, Kawasaki K, Takeda Y (1995) Modeling stratified diffusion in biological invasions. *American Naturalist* **2**, 229-251.

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 Research* **22**, 4673-4680.

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

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population-Structure. *Evolution* **38**, 1358-1370.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177-1191.

Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.

Wright S (1946) Isolation by distance under diverse systems of mating. *Genetics* **31**, 39-59.

Yü I (2005) Ceratopogonidae of China:(Insecta: Diptera). Military Medical Science Press, Beijing, China.

.

Figures and Tables

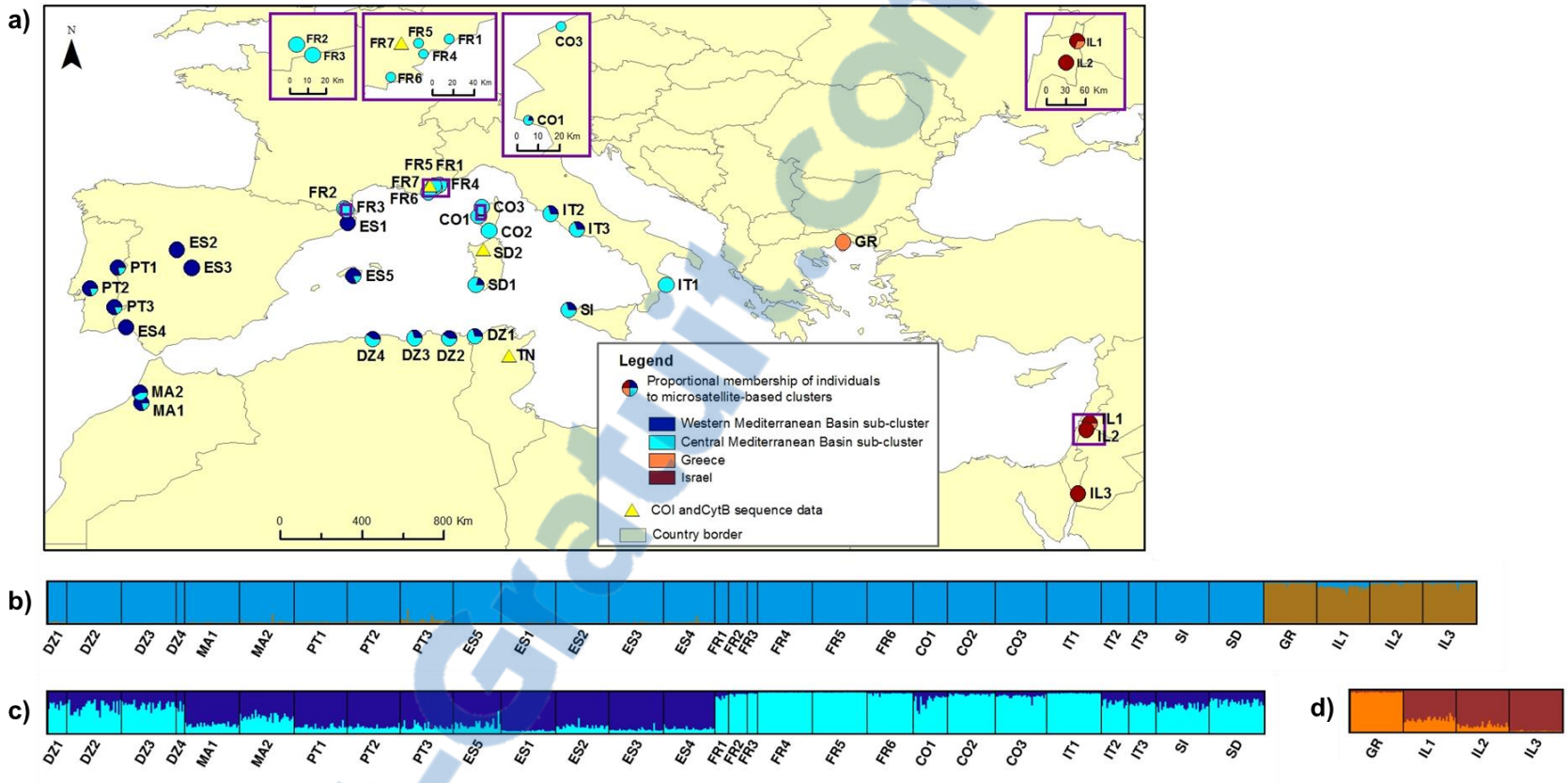


Fig. 1 Genetic clustering of *C. imicola* sampled populations. Genetic clustering of *C. imicola* sampled populations. Sampled locations and spatial Bayesian clustering based on microsatellite data (a). Ancestry estimation assuming two population clusters at the global scale (b), two population clusters within the western genetic group (b) and two population clusters within the eastern genetic group (c) 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.

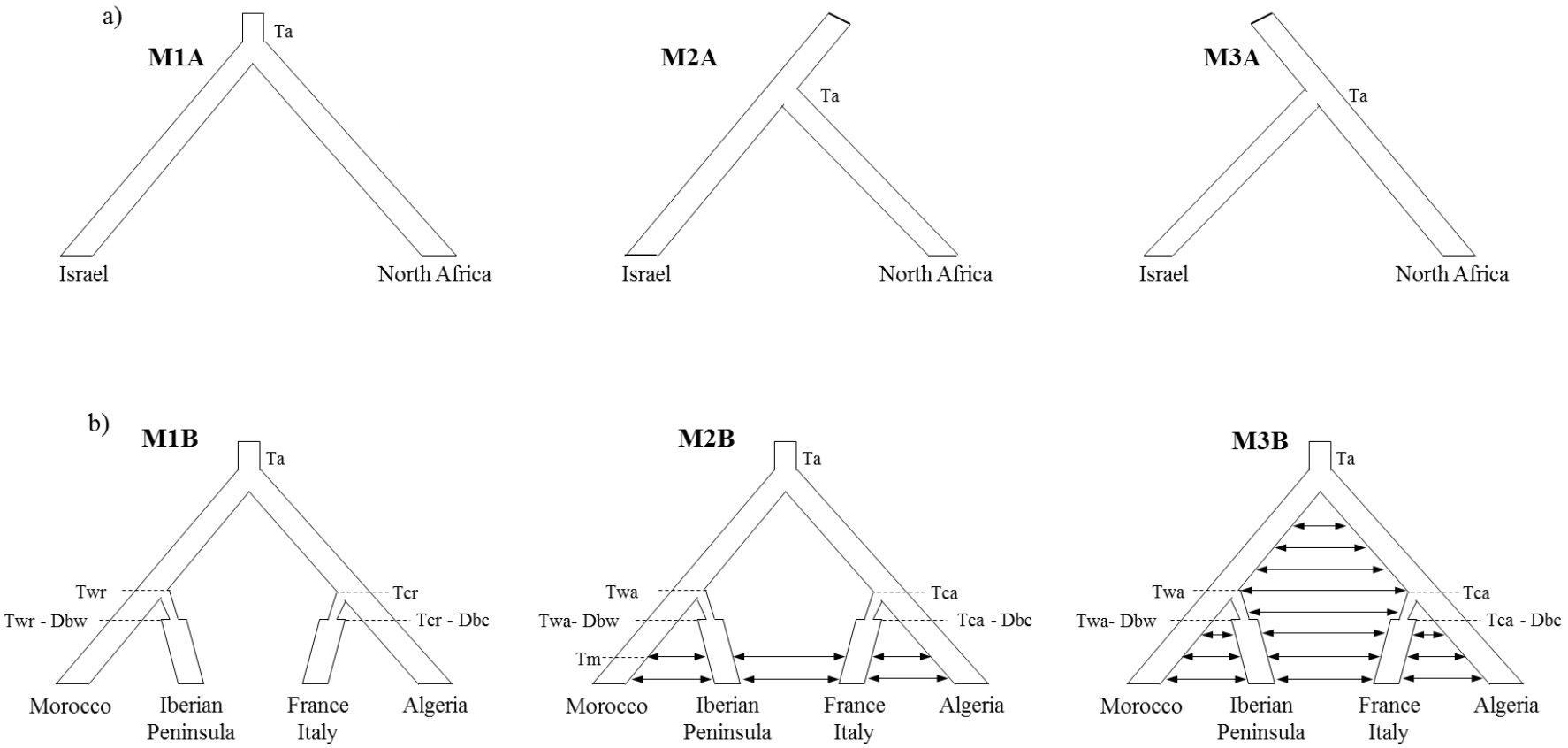


Fig. 2 Graphical representation of the tested scenarios regarding demographic histories of *C. imicola* (a) within the Mediterranean basin (western and eastern regions) and (b) within the western Mediterranean region. Microsatellite data were used and data were simulated using an approximate Bayesian computation (ABC) approach. Ta, Twr, Twa, Tc and Tca refer to the time of events, Dbw and Dbc, to the duration of bottleneck events. Details on the parameter priors are given in Table S2.

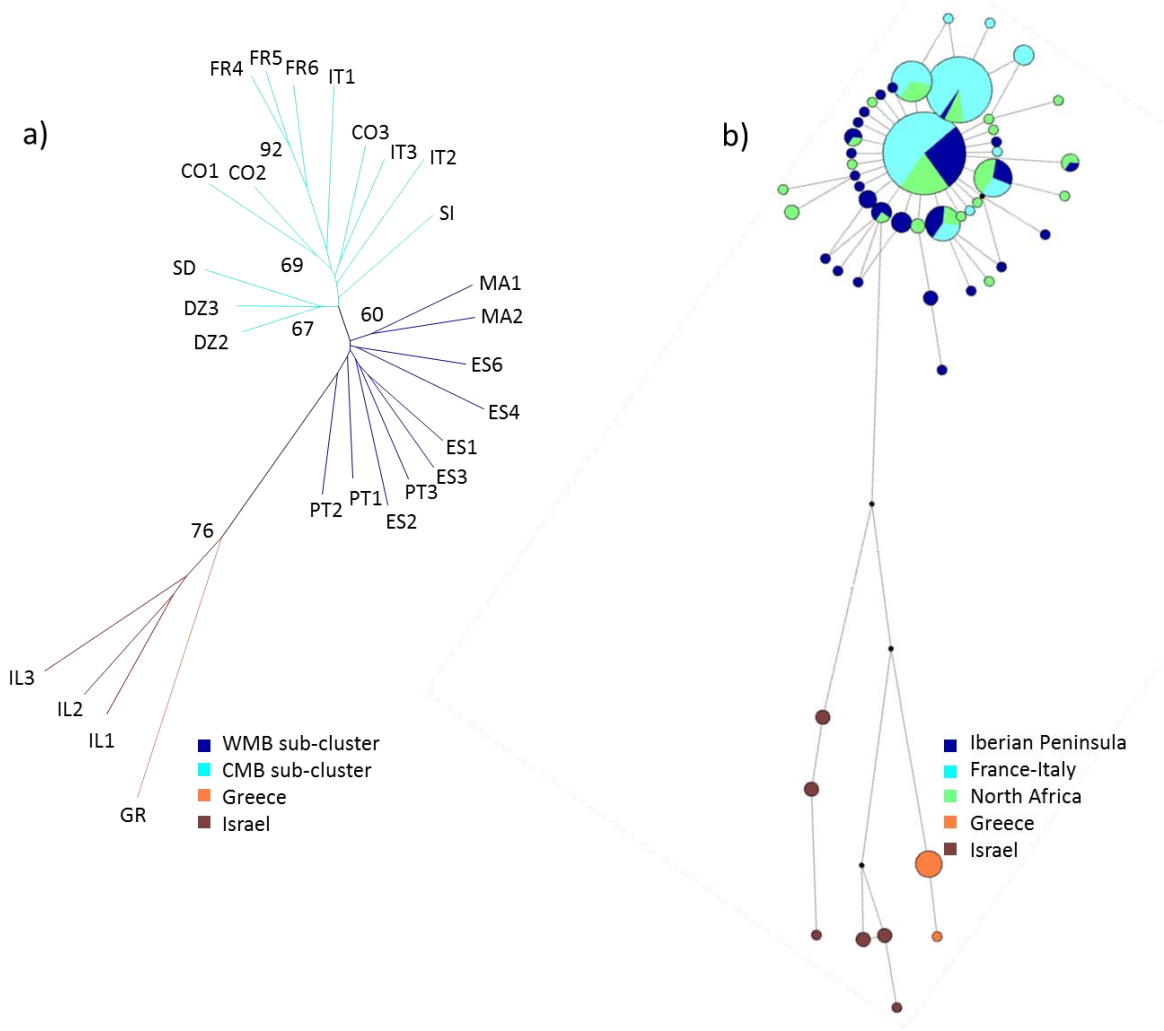


Fig. 3 Neighbor-joining tree of *C. imicola* population samples based on the chord distance of Cavalli-Sforza & Edwards (1967) computed on microsatellite polymorphism (a), and median-joining haplotype network of COI and CytB concatenated mitochondrial sequences (b). Colours represent the genetic clusters inferred by STRUCTURE in the neighbor-joining tree and the geographical regions of sampled specimens in the haplotype network.

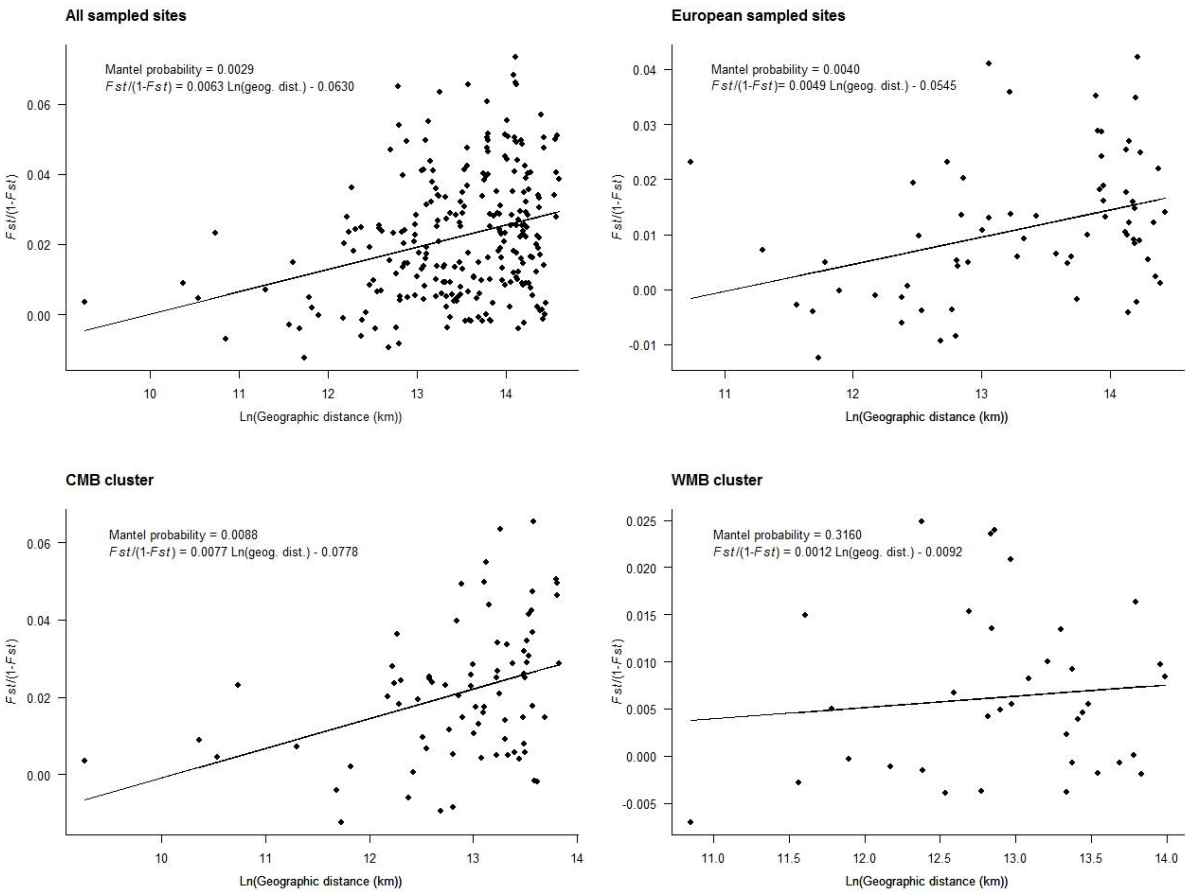


Fig. 4 Results of the Mantel tests for isolation by distance (IBD) between ln-transformed geographic distances and F_{ST} pairwise estimates based on microsatellite data. The tests were carried out for whole data set, Portugal/Spain/France/Italy, and the sub-clusters CMB (Algeria, France and Italy) and WMB (Portugal, continental Spain and Balearic Islands).

Table 1 Geographical location of sampled sites of *C. imicola*, number of individuals typed for microsatellite analyses (N_{mic}), number mitochondrial sequences obtained (N_{mit}) data and sample grouping based on clustering analysis based on microsatellite.

Country	Location	Collection date	Latitude	Longitude	N_{mic}	N_{mit}	Sample codes	Cluster codes	Region codes
Continental Spain	Girona	October 2012	41.84	2.85	32	7	ES1	WMB	Iberian Peninsula
Continental Spain	Avila	October 2012	40.66	-4.69	31	8	ES2	WMB	Iberian Peninsula
Continental Spain	Toledo	August 2012	39.86	-4.04	32	8	ES3	WMB	Iberian Peninsula
Continental Spain	Huelva	October 2012	37.24	-6.94	30	8	ES4	WMB	Iberian Peninsula
Portugal	Castelo Branco	April 2010	39.87	-7.30	31	8	PT1	WMB	Iberian Peninsula
Portugal	Coruche	August 2010	38.96	-8.53	31	8	PT2	WMB	Iberian Peninsula
Portugal	Beja	August 2010	38.14	-7.45	31	7	PT3	WMB	Iberian Peninsula
Spain Majorca	Felanitx	October 2012	39.50	3.10	28	8	ES5	WMB	Iberian Peninsula
Morocco	Khemisset	August 2002	33.91	-6.27	32	7	MA1	WMB	North Africa
Morocco	Sidi Yahia El Gharb	October 2004	34.37	-6.31	32	7	MA2	WMB	North Africa
Algeria	El Kala	June 2003	36.84	8.46	11	9	DZ1	CMB	North Africa
Algeria	Bekkouche Lakhdar	June 2003	36.74	7.31	32	10	DZ2	CMB	North Africa
Algeria	Kaous	June 2003	36.77	5.79	32	8	DZ3	CMB	North Africa
Algeria	Draa Ben Keda	June 2003	36.73	3.95	5	5	DZ4	CMB	North Africa
Tunisia	Sebikha	May 2013	36.0	10.0	-	5	TN	-	North Africa
Continental France	Alpes-Maritimes	October 2009	43.53	6.90	6	4	FR1	CMB	France-Italy
Continental France	Pyrénées-Orientales	September 2012	42.47	2.70	8	6	FR2	CMB	France-Italy
Continental France	Pyrénées-Orientales	September 2012	42.42	2.78	11	8	FR3	CMB	France-Italy
Continental France	Roquebrune-sur-Argens	August 2008	43.40	6.68	32	8	FR4	CMB	France-Italy
Continental France	Roquebrune-sur-Argens	August 2008	43.49	6.64	32	8	FR5	CMB	France-Italy
Continental France	Bormes-les-Mimosas	August 2008	43.20	6.40	27	8	FR6	CMB	France-Italy
Continental France	Trans-en-Provence	October 2008	43.49	6.50	-	8	FR7	-	France-Italy
Continental Italy	Roccabernarda	September 2012	39.10	16.90	32	8	IT1	CMB	France-Italy
Continental Italy	Tarquinia	September 2011	42.25	11.80	16	4	IT2	CMB	France-Italy
Continental Italy	Carrara	October 2012	41.57	12.95	16	4	IT3	CMB	France-Italy
France Corsica	Cargese	September 2008	42.14	8.62	20	6	CO1	CMB	France-Italy
France Corsica	Figari	September 2008	41.50	9.08	28	8	CO2	CMB	France-Italy
France Corsica	Calvi	September 2008	42.54	8.76	30	8	CO3	CMB	France-Italy
Italy Sardinia	San Giovanni Suergiu	November 2012	39.10	8.50	32	8	SD1	CMB	France-Italy
Italy Sicily	Paceco	October 2012	38.00	12.60	31	8	SI	CMB	France-Italy
Italy Sardinia	Mores	November 2012	40.66	8.85	-	8	SD2	-	France-Italy
Greece	Kavala	July 2013	41.00	24.70	31	8	GR	Greece	Greece
Israel	Ayelet Hashahar	September 2010	33.02	35.57	31	8	IL1	Israel	Israel
Israel	Ramat Tzvi	September 2011	32.72	35.42	31	8	IL2	Israel	Israel
Israel	Yotvata	November 2010	29.90	35.06	31	8	IL3	Israel	Israel

Table 2 Historical mutation-scaled migration rates (M) and migration rates (m) between each pair of *C. imicola* populations.

	Spain Majorca	Continental Spain	Morocco	Portugal	Algeria	France Corsica	Continental France	Continental Italy	Italy Sicily	Italy Sardinia
Spain Majorca	-	**36.15** 0.000216	**23.98** 0.000143	**13.23** 0.000079	**18.44** 0.000110	**18.88** 0.000113	**14.34** 0.000086	**23.95** 0.000143	**37.83** 0.000227	**18.09** 0.000108
Continental Spain	**36.17** 0.000217	-	**13.28** 0.000079	**23.83** 0.000143	**30.90** 0.000185	**29.02** 0.000174	**26.77** 0.000160	**23.81** 0.000142	**26.83** 0.000160	**37.66** 0.000225
Morocco	**37.70** 0.000226	**16.62** 0.000099	-	**22.80** 0.000136	**22.26** 0.000133	**15.35** 0.000092	**25.25** 0.000151	**46.81** 0.000280	**24.85** 0.000149	**21.18** 0.000127
Portugal	**30.37** 0.000182	**26.76** 0.000160	**14.62** 0.000087	-	**42.81** 0.000256	**37.79** 0.000226	**28.55** 0.000171	**39.34** 0.000236	**19.52** 0.000117	**33.06** 0.000198
Algeria	**38.23** 0.000229	**40.33** 0.000242	**14.55** 0.000087	**45.02** 0.000270	-	**19.12** 0.000114	**24.99** 0.000149	**53.93** 0.000323	**36.10** 0.000216	**50.95** 0.000305
France Corsica	**22.45** 0.000134	**33.04** 0.000198	**7.65** 0.000045	**36.77** 0.000220	**15.80** 0.000094	-	**32.61** 0.000195	**69.32** 0.000415	**33.36** 0.000200	**35.86** 0.000215
Continental France	**24.03** 0.000144	**19.44** 0.000116	**16.38** 0.000098	**25.15** 0.000150	**27.78** 0.000166	**24.87** 0.000149	-	**36.53** 0.000219	**18.66** 0.000111	**39.29** 0.000235
Continental Italy	**23.24** 0.000139	**27.03** 0.000162	**6.79** 0.000040	**21.59** 0.000129	**16.51** 0.000099	**23.95** 0.000143	**39.21** 0.000235	-	**26.83** 0.000160	**23.28** 0.000139
Italy Sicily	**63.28** 0.000379	**27.50** 0.000165	**12.66** 0.000075	**10.69** 0.000064	**16.87** 0.000101	**14.59** 0.000087	**26.52** 0.000159	**19.75** 0.000118	-	**25.48** 0.000152
Italy Sardinia	**29.90** 0.000179	**26.63** 0.000159	**21.22** 0.000127	**23.91** 0.000143	**29.91** 0.000179	**12.67** 0.000076	**46.19** 0.000277	**34.37** 0.000206	**31.16** 0.000186	-

The mutation-scaled migration rates (M) size inferred from MIGRATE-N, were kept to compare relative mutation-scaled immigration and emigration between populations (values atop in bold). Migration rates (m) were estimated from the mutation-scaled migration rates using the following the equation: $m = M\mu$, where and μ is the mutation rate per locus per generation assumed as 6.10^{-6} (microsatellite mutation rate of drosophila) (values below). This was done to compare historical migration rates and recent migration estimated with BAYESASS v.3.0.3.
Values are migration rates from the populations in the horizontal row into the populations in the vertical column.

Table 3 Recent migration rates estimated between each pair of populations.

	Spain Majorca	Continental Spain	Morocco	Portugal	Algeria	France Corsica	Continental France	Continental Italy	Italy Sicily	Italy Sardinia
Spain Majorca	**0.6759 (0.0091)**	0.0080 (0.0078)	0.0082 (0.0080)	0.0106 (0.0102)	0.0084 (0.0082)	0.0094 (0.0091)	0.0078 (0.0076)	0.0080 (0.0078)	0.0082 (0.0081)	0.0081 (0.0079)
Continental Spain	0.0104 (0.0101)	**0.6760 (0.0093)**	0.0118 (0.0116)	0.0462 (0.0380)	0.0115 (0.0112)	0.0105 (0.0101)	0.0080 (0.0078)	0.0100 (0.0100)	0.0096 (0.0094)	0.0099 (0.0096)
Morocco	0.0149 (0.0206)	0.0087 (0.0088)	**0.6830 (0.0226)**	0.0122 (0.0119)	0.0174 (0.0264)	0.0096 (0.0094)	0.0081 (0.0080)	0.0082 (0.0081)	0.0087 (0.0086)	0.0089 (0.0089)
Portugal	0.1670 (0.0390)	0.2348 (0.0303)	0.1728 (0.0432)	**0.7842 (0.0478)**	0.0348 (0.0259)	0.0162 (0.0156)	0.0112 (0.0107)	0.0130 (0.0123)	0.0337 (0.0339)	0.0229 (0.0196)
Algeria	0.0523 (0.0314)	0.0102 (0.0101)	0.0388 (0.0306)	0.0162 (0.0150)	**0.7484 (0.0392)**	0.0129 (0.0121)	0.0100 (0.0096)	0.0104 (0.0102)	0.0139 (0.0137)	0.0168 (0.0159)
France Corsica	0.0089 (0.0087)	0.0080 (0.0078)	0.0082 (0.0080)	0.0105 (0.0101)	0.0087 (0.0086)	**0.6785 (0.0108)**	0.0078 (0.0076)	0.0080 (0.0079)	0.0083 (0.0081)	0.0081 (0.0080)
Continental France	0.0088 (0.0086)	0.0079 (0.0077)	0.0080 (0.0078)	0.0101 (0.0097)	0.0081 (0.0079)	0.0089 (0.0086)	**0.6746 (0.0078)**	0.0080 (0.0078)	0.0081 (0.0079)	0.0080 (0.0078)
Continental Italy	0.0310 (0.0232)	0.0210 (0.0190)	0.0375 (0.0282)	0.0626 (0.0327)	0.1319 (0.0368)	0.2200 (0.0290)	0.2561 (0.0226)	**0.9127 (0.0255)**	0.2123 (0.0437)	0.2259 (0.0322)
Italy Sicily	0.0189 (0.0206)	0.0110 (0.0110)	0.0122 (0.0133)	0.0281 (0.0217)	0.0188 (0.0192)	0.0185 (0.0144)	0.0081 (0.0079)	0.0109 (0.0106)	**0.6863 (0.0211)**	0.0118 (0.0121)
Italy Sardinia	0.0119 (0.0115)	0.0144 (0.0136)	0.0195 (0.0161)	0.0191 (0.0167)	0.0120 (0.0125)	0.0155 (0.0142)	0.0082 (0.0080)	0.0108 (0.0106)	0.0109 (0.0108)	**0.6797 (0.0133)**

Values are migration rates from the populations in the horizontal row into the populations in the vertical column, inferred with BAYESASS v.3.0.3. The diagonal values in bold are the percentages of resident individuals in each population per generation, and the values in the brackets are the standard deviation of the marginal posterior distribution for each estimate.

Table 4 Estimates of dispersal distances between reproducing adults and their parents using isolation by distance estimates of $D\sigma^2$.

	Θ	Du (km)	σ (km)	Du (km)	σ (km)	Du (km)	σ (km)
Spain Majorca	0.265	1	0.034 – 0.138	3	0.059 – 0.239	5	0.076 – 0.309
Continental Spain	0.296	1	0.032 – 0.130	3	0.055 – 0.226	5	0.071 – 0.292
Morocco	0.420	1	0.027 – 0.110	3	0.046 – 0.190	5	0.060 – 0.245
Portugal	0.360	1	0.029 – 0.118	3	0.050 – 0.205	5	0.065 – 0.265
Algeria	0.266	1	0.034 – 0.138	3	0.058 – 0.239	5	0.075 – 0.308
France Corsica	0.335	1	0.030 – 0.123	3	0.052 – 0.213	5	0.067 – 0.274
Continental France	0.090	1	0.058 – 0.237	3	0.100 – 0.410	5	0.130 – 0.530
Continental Italy	0.190	1	0.040 – 0.163	3	0.069 – 0.282	5	0.089 – 0.365
Italy Sicily	0.154	1	0.044 – 0.181	3	0.077 – 0.314	5	0.099 – 0.405
Italy Sardinia	0.618	1	0.022 – 0.090	3	0.038 – 0.157	5	0.049 – 0.202

Dispersal distances were obtained by the effective population size ($Ne = \Theta / 4\mu$, where μ is the mutation rate $6.10^{-6} – 1.10^{-4}$) and using the regression slope of the Mantel tests for the whole data set ($b = 0.0063$); b is the value of the regression slope. Du is the surface area of *C. imicola* distribution, Θ is the mutation-scaled effective population size (inferred with MIGRATE-N software).

Supplementary Information

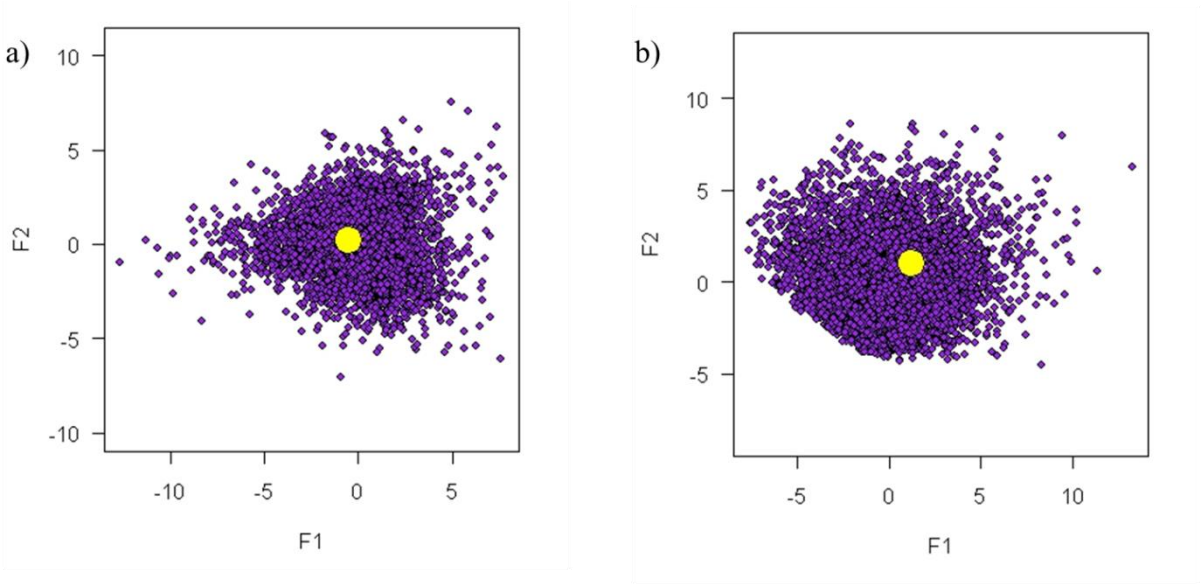


Fig. S1 PCA analysis for the best fitting model (a) of western and eastern Mediterranean divergence and (b) demographic history in western Mediterranean region. The yellow dot represents the real data set and the purple dots represent the 5,000 simulation closest to the real data.

Table S1 Primers used for the amplification of the microsatellite loci in *C. imicola* (Mardulyn et al. 2013), allele size range observed in this study and annealing temperature (Ta) used for amplification.

Loci	Motif	Forward / Reverse	Allele size range (bp)	Ta (°C)
68	(GT)n	CTTTTCCGTTTCTTTTTATTTCTTT	95-105	60
		GTTTCTTTCTGGTCGCGTTGGTTGCTG		
12b	(CT)n	TTATGTGTGTATGTTAGCAAGGTCA	129-139	50
		GTTTCTTCTTCGGATCAAAGAAATTTTGCC		
3b	(AC)n	ATGCGGATGTTTGAAGTG	154-175	50
		GTTTCTTTTTTGTGTCTTATTGCCC		
31	(CAA)n	TTCTGTTCGGCTGTTGCGTT	152-168	60
		GTTTCTTCTTTTTACGTGGTGGTCATTC		
41b	(CT)n	GAGGAGGAGGTAGAA	155-166	50
		GTTTCTTTCTATTAGTCAATGGTG		
35t	(AC)n	TTTGTAAAAGCCAGTTCAACCG	171-194	60
		GTTTCTTATCGAACGAAGGAAATAACCAC		
88b	(AC)n	TTTGTTCGATTTGTAGTG	243-256	50
		GTTTCTTCCTCTCTTTCATTCGC		
16	(TG)n	TTGCCTTTGCTTGTTGAGGATG	292-299	60
		GTTTCTTTCCTCTTTAAAATCACTGACGTG		
88	(CAT)n	GTTGGTGCTTTGTTGTGTTGT	344-348	50
		GTTTCTTTTTCTTTTTCTCCTTTTTGTTTCTTTC		

Table S2 Prior distributions of demographic, historical and mutation parameters used for ABC inferences.

Parameter description	Parameter	Prior distribution
Stable effective population size	N_E	Loguniform [1000; 100,000]
Founding effective population size	N_S	Loguniform [2; 100]
Duration of bottleneck	D_B	Uniform [1; 50]
Recent time of colonization events in Iberian Peninsula	T_{WR}	Uniform [320; 620]
Recent time of colonization events in France-Italy	T_{CR}	Uniform [220; 620]
Ancient time of colonization events in Iberian Peninsula	T_{WA}	Uniform [2,000; 20,000]
Ancient time of colonization events in France-Italy	T_{CA}	Uniform [2,000; 20,000]
Ancestral divergence time of Morocco and Algeria	T_A	Uniform [5,000; 30,000]
Ancestral divergence time of Israel and North Africa	T_A	Uniform [5,000; 30,000]
Mean microsatellite mutation model	μ	Loguniform [6.10^{-6}; 10^{-4}]
Migration rate	m	Loguniform [10^{-6}; 10^{-3}]

Effective population sizes (N) are expressed in number of diploid individuals and times of events (t) in number of generations going back to the past. Conditions among the parameters used during the simulations were T_{WR}, T_{CR}, T_{WA} and $T_{CA} \leq T_A$.

Table S3 Microsatellite pairwise F_{ST} values between *C. imicola* population samples estimated with FSTAT v2.9.3.2.

		WMB									CMB														Greece	Israel		
	Sites	ES2	ES3	ES4	ES5	MA1	MA2	PT1	PT2	PT3	DZ2	DZ3	CO1	CO2	CO3	FR4	FR5	FR6	IT1	IT2	SI	SD	IT3	GR	IL1	IL2	IL3
WMB	ES1	-0.0081	-0.0029	-0.0054	0.0031	0.0057	0.0084	-0.0016	-0.0061	-0.0009	-0.0061	**0.0287**	**0.0228**	**0.0238**	**0.0308**	**0.0177**	**0.0606**	**0.0454**	**0.0409**	**0.0444**	**0.0288**	**0.0254**	0.0107	**0.1020**	**0.0887**	**0.0800**	0.0505
	ES2		0.0012	0.0040	0.0022	0.0008	0.0066	-0.0071	0.0017	0.0022	-0.0011	0.0254	**0.0143**	**0.0237**	0.0195	**0.0166**	**0.0496**	**0.0358**	**0.0328**	**0.0469**	0.0227	0.0243	-0.0001	**0.1155**	**0.0889**	**0.1021**	**0.1392**
	ES3			-0.0006	0.0076	0.0084	0.0036	-0.0019	-0.0059	0.0071	-0.0047	**0.0298**	**0.0347**	**0.0286**	**0.0235**	**0.0174**	**0.0567**	**0.0416**	**0.0378**	**0.0466**	**0.0344**	**0.0312**	**0.0150**	**0.1178**	**0.0902**	**0.1030**	**0.1446**
	ES4				0.0039	0.0007	0.0032	-0.0001	-0.0037	0.0126	-0.0023	0.0239	**0.0386**	**0.0252**	**0.0265**	**0.0164**	**0.0578**	**0.0471**	**0.0447**	**0.0428**	0.0231	0.0195	0.0188	**0.1200**	**0.0911**	**0.1040**	**0.1453**
	ES5					-0.0040	0.0010	0.0020	-0.0026	0.0051	-0.0060	0.0174	0.0143	**0.0151**	**0.0160**	**0.0061**	**0.0379**	**0.0356**	**0.0279**	**0.0411**	0.0055	0.0208	0.0151	**0.0917**	**0.0497**	0.0066	0.0164
	MA1						-0.0007	0.0027	-0.0007	0.0104	0.0037	0.0128	**0.0163**	0.0210	**0.0111**	**0.0120**	**0.0434**	**0.0375**	**0.0352**	**0.0414**	0.0058	0.0124	**0.0105**	**0.1229**	**0.0934**	**0.1059**	**0.1493**
	MA2							0.0038	0.0004	0.0136	0.0013	0.0106	0.0271	0.0226	0.0171	**0.0107**	**0.0277**	**0.0160**	**0.0206**	**0.0362**	0.0137	0.0194	0.0117	**0.1302**	**0.0938**	**0.1094**	**0.1571**
	PT1								-0.0036	-0.0093	-0.0034	0.0121	0.0281	**0.0231**	0.0183	**0.0110**	**0.0570**	**0.0425**	**0.0413**	**0.0286**	0.0265	0.0201	0.0073	**0.1324**	**0.1030**	**0.1111**	**0.1587**
	PT2									-0.0071	-0.0078	0.0180	0.0243	**0.0127**	0.0163	**0.0049**	**0.0428**	**0.0316**	**0.0299**	**0.0360**	0.0196	0.0169	0.0120	**0.1329**	**0.1044**	**0.1128**	**0.1633**
	PT3										-0.0019	0.0230	0.0284	**0.0225**	**0.0241**	**0.0065**	**0.0660**	**0.0453**	**0.0447**	**0.0344**	0.0366	**0.0345**	0.0155	**0.1348**	**0.1087**	**0.1199**	**0.1639**
CMB	DZ2											0.0160	0.0146	0.0071	0.0132	**0.0041**	**0.0426**	**0.0322**	0.0243	**0.0270**	0.0191	0.0169	0.0038	**0.1383**	**0.1088**	**0.1223**	**0.1668**
	DZ3												0.0188	0.0141	0.0118	**0.0094**	**0.0388**	**0.0395**	**0.0287**	0.0173	0.0118	0.0039	0.0051	**0.1417**	**0.1090**	**0.1271**	**0.1689**
	CO1													0.0232	0.0238	**0.0157**	**0.0413**	**0.0494**	**0.0190**	**0.0388**	0.0078	0.0334	0.0015	**0.1487**	**0.1120**	**0.1274**	**0.1720**
	CO2														0.0083	0.0062	**0.0225**	0.0269	**0.0229**	**0.0265**	0.0219	0.0121	0.0055	**0.1505**	**0.1175**	**0.1347**	**0.1736**
	CO3															**0.0046**	**0.0280**	**0.0261**	**0.0196**	**0.0156**	0.0134	0.0079	0.0045	**0.1514**	**0.1250**	**0.1414**	**0.1880**
	FR4																0.0232	0.0256	**0.0222**	**0.0164**	0.0057	0.0087	0.0076	**0.1674**	**0.1283**	**0.1479**	**0.1886**
	FR5																	0.0009	**0.0129**	**0.0392**	**0.0314**	**0.0377**	**0.0374**	**0.1684**	**0.1286**	**0.1533**	**0.1934**
	FR6																		0.0051	**0.0390**	**0.0474**	**0.0417**	**0.0312**	**0.1699**	**0.1297**	**0.1534**	**0.1966**
	IT1																			**0.0368**	**0.0342**	**0.0435**	**0.0204**	**0.1734**	**0.1380**	**0.1540**	**0.2021**
	IT2																				**0.0382**	**0.0191**	**0.0227**	**0.1774**	**0.1385**	**0.1544**	**0.2069**
	SI																					0.0088	0.0189	**0.1831**	**0.1416**	**0.1605**	**0.2079**
	SD																						0.0101	**0.1977**	**0.1526**	**0.1639**	**0.2120**
	IT3																							**0.1994**	**0.1624**	**0.1668**	**0.2182**
Greece	GR																								**0.1672**	**0.1795**	**0.2305**
Israel	IL1																									**0.1824**	**0.2311**
	IL2																										**0.2385**

F_{ST} values are grouped according to the genetic clusters inferred by STRUCTURE v.2.3.3.: (WMB) western Mediterranean basin, (CMB) central Mediterranean basin, (SEE) south-eastern Europe and (ME) Middle-East. Significant values, at the adjusted nominal level (5%) for multiple comparison of 0.0002, are highlighted in bold.
Details on sample codes are given in Table 1.

Table S4 Pairwise F_{ST} values based on COI and CytB concatenated sequences between the population samples and the geographical regions.

		North Africa		Iberian Peninsula			France-Italy					Greece	Israel
	Locations	MA	TN	BA	ES	PT	CO	FR	IT	SD	SI	GR	IL
North Africa	DZ	**0.0549**	0.0086	**0.0504**	**0.0504**	**0.0592**	0.0349	**0.1897**	0.0896	0.0229	0.0304	**0.8799**	**0.7524**
	MA		-0.0130	0.0229	0.0229	**0.0466**	**0.0730**	**0.2026**	**0.1021**	**0.0652**	-0.0034	**0.8273**	**0.6569**
	TN			**0.1932**	0.0022	0.0308	0.0226	**0.2828**	0.1412	-0.0057	-0.0405	**0.9310**	**0.6059**
Iberian Peninsula	BA				**0.1277**	**0.1588**	**0.1576**	**0.3698**	**0.3141**	**0.3034**	0.1904	**0.9308**	**0.6634**
	ES					0.0180	**0.0929**	**0.2090**	**0.1135**	**0.0497**	-0.0127	**0.8611**	**0.7231**
	PT						0.0226	**0.2828**	0.1412	-0.0057	-0.0405	**0.9310**	**0.6059**
France-Italy	CO							**0.0747**	-0.0116	0.0598	-0.0088	**0.8737**	**0.7128**
	FR								-0.0037	**0.3101**	**0.1597**	**0.9203**	**0.8207**
	IT									**0.1854**	0.0324	**0.9229**	**0.7172**
	SD										-0.0181	**0.9468**	**0.7316**
	SI											**0.9275**	**0.6482**
Greece	GR												**0.6345**

Iberian Peninsula group: continental Spain, Portugal, Balearic Islands
France-Italy group: Corsica, continental France, Sardinia, Sicily, continental Italy
Population differentiation was assessed with the exact G test implemented in ARLEQUIN v.3.5.3.1. Significant values (p-value < 0.05) are highlighted in bold.

Table S5 Genetic diversity based on microsatellite data for each sampled site and genetic cluster, assessed by STRUCTURE v.2.3.3.

Locations / clusters	H_S	A_R	F_{IS}
WMB sub-cluster	**0.621 ± 0.118**	**6.861 ± 2.075**	**0.059***
ES1	0.608 ± 0.102	4.165 ± 1.063	0.059
ES2	0.609 ± 0.054	4.345 ± 0.918	0.002
ES3	0.628 ± 0.122	4.497 ± 1.151	0.037
ES4	0.642 ± 0.121	4.141 ± 1.125	0.103
ES5	0.628 ± 0.090	4.544 ± 0.924	0.094
MA1	0.605 ± 0.094	4.568 ± 0.651	-0.044
MA2	0.617 ± 0.127	4.505 ± 0.527	0.075
PT1	0.627 ± 0.122	4.443 ± 0.924	0.024
PT2	0.581 ± 0.167	4.341 ± 0.764	0.058
PT3	0.652 ± 0.116	4.857 ± 1.552	0.090
CMB sub-cluster	**0.614 ± 0.088**	**5.974 ± 1.453**	**0.078***
DZ2	0.605 ± 0.072	4.087 ± 0.969	0.077
DZ3	0.614 ± 0.143	3.895 ± 1.464	0.052
CO1	0.585 ± 0.104	3.869 ± 0.958	0.035
CO2	0.595 ± 0.110	3.985 ± 1.504	0.097
CO3	0.620 ± 0.136	4.018 ± 1.049	0.064
FR4	0.583 ± 0.112	3.232 ± 1.202	0.126
FR5	0.557 ± 0.139	3.272 ± 1.390	0.010
FR6	0.572 ± 0.111	3.477 ± 1.281	0.025
IT1	0.579 ± 0.132	3.711 ± 1.413	0.111
IT2	0.625 ± 0.111	4.202 ± 1.134	0.030
IT3	0.611 ± 0.158	4.209 ± 1.428	0.112
SI	0.629 ± 0.156	4.317 ± 1.223	0.096
SD	0.591 ± 0.123	4.031 ± 1.775	-0.017
Greece (GR)	**0.495 ± 0.315**	**5.000 ± 2.483**	**0.051**
Israel	**0.567 ± 0.132**	**5.840 ± 2.293**	**0.079***
IL1	0.610 ± 0.155	4.919 ± 1.910	0.125
IL2	0.575 ± 0.115	4.266 ± 1.331	0.061
IL3	0.492 ± 0.175	4.055 ± 1.661	-0.003

The allelic richness (A_r), genetic diversity (H_s) and inbreeding coefficient (F_{IS}) were computed with FSTAT v2.9.3.2.

A_R is based on the minimum sample size of 16 diploid individuals.

WMB sub-clustr: continental Spain, Portugal, Balearic Islands, Morocco

CMB sub-cluster: Corsica, continental France, Sardinia, Sicily, continental Italy, Algeria

Significant F_{IS} values, at the adjusted nominal level (5%) for multiple comparison of 0.0002, are indicated with an asterisk**.**

Table S6 Genetic diversity based on mitochondrial data for each sampled site and geographical regions.

Locations / clusters	N_{COI}	H_{COI}	Hd_{COI}	π_{COI}	N_{CTB}	H_{CTB}	Hd_{CTB}	π_{CTB}	N_{CON}	H_{CON}	Hd_{CON}	π_{CON}
North Africa	**51**	**8**	**0.383 ± 0.086**	**0.0011 ± 0.0003**	**52**	**13**	**0.713 ± 0.064**	**0.0015 ± 0.0002**	**50**	**19**	**0.860± 0.039**	**0.0014 ± 0.0002**
Algeria	32	3	0.232 ± 0.094	0.0006 ± 0.0003	31	6	0.690 ± 0.072	0.0015 ± 0.0002	31	8	0.800 ± 0.052	0.0011 ± 0.0001
Morocco	14	7	0.758 ± 0.116	0.0026 ± 0.0007	16	6	0.683 ± 0.120	0.0015 ± 0.0004	14	11	0.956 ± 0.033	0.0020 ± 0.0002
Tunisia	5	1	0	0	5	4	0.900 ± 0.161	0.0020 ± 0.0005	5	4	0.900 ± 0.161	0.0011 ± 0.0003
Iberian Peninsula	**62**	**8**	**0.348 ± 0.077**	**0.0008 ± 0.0002**	**62**	**15**	**0.719 ± 0.059**	**0.0018 ± 0.0003**	**62**	**22**	**0.840 ± 0.044**	**0.0014 ± 0.0002**
Balearic Islands	8	1	0	0	8	3	0.679 ± 0.122	0.0013 ± 0.0003	8	3	0.679 ± 0.122	0.0007 ± 0.0002
Mainland Spain	31	8	0.454 ± 0.111	0.0010 ± 0.0003	31	11	0.662 ± 0.095	0.0015 ± 0.0003	31	17	0.841 ± 0.080	0.0014 ± 0.0002
Portugal	23	4	0.320 ± 0.121	0.0007 ± 0.0003	23	2	0.735 ± 0.080	0.0020 ± 0.0004	23	9	0.840 ± 0.061	0.0014 ± 0.0002
France-Italy	**115**	**8**	**0.544 ± 0.033**	**0.0013 ± 0.0001**	**114**	**5**	**0.324 ± 0.054**	**0.0006 ± 0.0001**	**114**	**13**	**0.717 ± 0.028**	**0.0009 ± 0.0001**
Corsica	22	5	0.615 ± 0.097	0.0017 ± 0.0004	22	4	0.542 ± 0.101	0.0010 ± 0.0002	22	8	0.861 ± 0.040	0.0013 ± 0.0002
Mainland France	53	5	0.557 ± 0.034	0.0013 ± 0.0001	53	3	0.112 ± 0.035	0.0002 ± 0.0001	53	7	0.615± 0.044	0.0006 ± 0.0001
Mainland Italy	16	2	0.5251 ± 0.055	0.0011 ± 0.0001	16	3	0.342 ± 0.140	0.0006 ± 0.0002	16	5	0.700 ± 0.080	0.0008 ± 0.0002
Sardinia	16	1	0	0	15	3	0.514 ± 0.116	0.0009 ± 0.0002	15	3	0.514 ± 0.116	0.0005 ± 0.0001
Sicily	8	3	0.464 ± 0.200	0.0010 ± 0.0003	8	3	0.464 ± 0.200	0.0008 ± 0.0004	8	5	0.786 ± 0.151	0.0009 ± 0.0003
Greece	**8**	**1**	**0**	**0**	**8**	**2**	**0.250 ± 0.180**	**0.0003 ± 0.0002**	**8**	**2**	**0.250 ± 0.180**	**0.0023 ± 0.0002**
Israel	**24**	**5**	**0.717 ± 0.053**	**0.0042 ± 0.0004**	**10**	**5**	**0.822 ± 0.097**	**0.0047 ± 0.0006**	**10**	**6**	**0.911 ± 0.062**	**0.0048 ± 0.0006**

The number of haplotypes (H), haplotype diversity (Hd), and nucleotide diversity (π) and their standard deviations for COI, CytB and COI+CytB concatenated dataset were estimated using DNASP v.5.10.

Chapitre III : Expansion géographique et colonisation de nouveaux milieux à la limite nord de la répartition de *C. imicola*

L'expansion géographique des populations est gouvernée par de nombreux facteurs. Parmi ces facteurs, la capacité de dispersion et la taille des populations sont déterminants au pour le succès d'expansion géographique (Caswell *et al.* 2003; Sakai *et al.* 2001). Les phénomènes d'expansion géographique ont des conséquences sur la généalogie des populations, et ces signatures génétiques persistent dans le temps. Il est alors possible d'étudier les facteurs démographiques et génétiques impliqués dans ces processus. Etudier la variabilité spatio-temporelle des populations en expansion peut fournir des informations sur les modes et les taux de dispersion, ainsi que sur les processus sous-jacents à leurs établissements dans de nouveaux milieux (Peterson & Vieglais 2001).

Depuis son installation dans le bassin méditerranéen, les données entomologiques suggèrent que *C. imicola* pourrait avoir étendu son aire de distribution vers le nord, atteignant la France continentale. En 2003, des individus sont collectés dans le département du Var, et en 2008, l'espèce est signalée dans le département des Pyrénées-Orientales, à proximité de la frontière espagnole. Ces découvertes soulèvent plusieurs questions. Quels sont les facteurs principaux à l'origine de cette expansion ? Quelle est la structuration des populations dans un environnement marqué par les barrières physiques et géographiques ? Quels sont les modes et l'intensité de la dispersion ? Quelle est l'évolution spatio-temporelle de la dynamique des populations en expansion ? Ce troisième chapitre présente deux études adressant ces questions à une échelle locale.

La première étude s'intéresse à l'expansion de *C. imicola* dans le département des Pyrénées-Orientales et à la source des populations présentes dans ce département. Etant donné la proximité géographique des populations de Catalogne (Espagne), l'hypothèse de départ était que la population source figurait parmi ces populations installées. Une méthodologie multi-approche a été suivie combinant des données de suivis entomologiques, des analyses de génétique de populations et de phylogéographie basées sur neuf marqueurs microsatellites et deux gènes mitochondriaux, des approches ABC et des analyses de modélisation de la dispersion passive à l'aide d'un modèle atmosphérique lagrangien, le Numerical Atmospheric-dispersion Modelling Environment (NAME). Ce modèle permet de simuler le transport, le relâchement et le mélange de particules transportées par le vent, ainsi que leur retrait de l'atmosphère (Jones *et al.* 2007).

Nos analyses génétiques combinées à la modélisation avec le modèle NAME suggèrent que *C. imicola* aurait colonisé les Pyrénées-Orientales depuis des populations abondantes de

Corse via des vents balayant la zone et pouvant transporter des particules jusqu'à la France continentale. Cette étude révèle, en outre, l'importance de l'abondance des populations sources couplées à la présence de vents aux conditions favorables permettant la dispersion de l'espèce sur de longues distances. Ce résultat confirme une nouvelle fois la forte capacité de dispersion de l'espèce par le vent, et le rôle de celui-ci dans les phénomènes d'expansion géographique. Il serait intéressant d'évaluer pour le gestionnaire en santé animale la fréquence de telles introductions pour estimer le risque épidémiologique de transport passif de *Culicoides* infectés lors des épizooties importantes dans la zone.

La deuxième étude s'intéresse à la dynamique des populations en Corse et dans le département du Var. Comme précédemment, une stratégie multi-approche combinant la génétique de populations et les enquêtes entomologiques a été utilisée. Plus précisément, une étude génétique spatio-temporelle sur quatre ans (2002-2012) basée sur le polymorphisme de neuf marqueurs microsatellites a permis de caractériser la diversité et la variabilité génétique des populations.

L'étude de la dynamique des populations corses et du département du Var révèle que les fortes abondances des populations en Corse couplées à une dispersion par le vent ont permis la colonisation du Var depuis l'île. Nos données génétiques indiquent un patron génétique caractéristique des expansions géographiques avec une faible diversité génétique dans le nouvel environnement et une différenciation génétique significative entre les populations corses et varoises, qui aurait significativement augmenté au cours du temps. Ces résultats pourraient être la résultante d'évènements de goulot d'étranglement suivie des effets de la dérive génétique suite à l'introduction. Parallèlement, la différentiation génétique spatio-temporelle à l'échelle locale dans les deux zones est faible, voire absente, renforçant l'hypothèse d'une forte capacité de dispersion de l'espèce. Une estimation du taux et de la vitesse de dispersion permettrait d'évaluer la dynamique et les patrons de dispersion active et passive intra-continentale.

Ce chapitre a fait l'objet de deux articles le premier soumis à Scientific Reports et le second soumis à Parasites & Vectors.

Range expansion of the Bluetongue vector *Culicoides imicola* in continental France thanks to meteorological events

Stéphanie Jacquet[1, 16, 17*], Karine Huber[2], Nonito Pagès[3, 4], Sandra Talavera[4], Laura E. Burgin[5], Simon Carpenter[6], Christopher Sanders[6], Ahmadou H. Dicko[7, §], Mouloud Djerbal [8, §], Maria Goffredo[9, §], Youssef Lhor[10, §], Javier Lucientes[11, §], Miguel A. Miranda-Chueca[12, §], Isabel Pereira Da Fonseca[13, §], David W. Ramilo[13, §], Marie-Laure Setier-Rio[14, §], Jérémy Bouyer[1, 15], Christine Chevillon[16, 17], Thomas Balenghien[1], Hélène Guis[1], Claire Garros[1*]

[1] Cirad, UMR15 CMAEE, 34398 Montpellier, France; INRA, UMR1309 CMAEE, 34398 Montpellier, France.[2] INRA, UMR1309 CMAEE, 34398 Montpellier, France; Cirad, UMR15 CMAEE, 34398 Montpellier, France. [3] Cirad, UMR15 CMAEE, 97170 Petit-Bourg, France; INRA, UMR1309 CMAEE 34398 Montpellier, France. [4] Centre de Recerca en Sanitat Animal (CReSA), UAB-IRTA, Campus de la Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Spain. [5] Met Office, Exeter, UK. [6] Vector-borne Viral Diseases Programme, The Pirbright Institute, Pirbright, UK. [7] West African Science Service on Climate Change and Adapted Land Use, Climate Change Economics Research Program, Cheikh Anta Diop University, Dakar-Fann, Sénégal. [8] Institut National de la Médecine Vétérinaire (IMV), Laboratoire vétérinaire régional, Tizi Ouzou, Algeria. [9] Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise 'G. Caporale', 64100 Teramo, Italy. [10]Office National de Sécurité Sanitaire des produits Alimentaires (ONSSA), Rabat, Morocco. [11] Faculdad de Veterinaria, University of Zaragoza (UZ), Zaragoza, Spain.[12] Laboratory of Zoology, University of Balearics (UIB), Palma de Mallorca, Spain. [13] CIISA, Faculdade de Medecina Veterinaria, Universidade de Lisboa (FMV-ULisboa), Lisboa, Portugal. [14] Entente interdépartementale pour la démoustication-Méditerranée (EID-Méd), Montpellier, France. [15] Institut Sénégalais de Recherches Agricoles (ISRA), Laboratoire National de l'Elevage et de Recherches Vétérinaires, Dakar, Sénégal. [16] CNRS, Université de Montpellier, UMR 5290 Maladies Infectieuses & Vecteurs-Ecologie, Génétique, Ecologie, Contrôle (MIVEGEC). [17] IRD, UR 224 MIVEGEC, BP 64501, Agropolis, 34 394 Montpellier cedex 5, 25 IRD, (INTERTRYP), UMR177 IRD-CIRAD, TA A-17/G, Campus International de Baillarguet, Montpellier, France.

*Corresponding authors

*Cirad, UMR Contrôle des Maladies Animales Exotiques et Emergentes, Campus International de Baillarguet, TA-A15/G, 34398 Montpellier cedex 5, France, stephanie.jacquet@cirad.fr, claire.garros@cirad.fr

[$] by alphabetical order.

Abstract

The influence of a northward expansion of *Culicoides imicola* Kieffer in driving recent and unprecedented outbreaks of *Culicoides*-borne arboviruses in southern Europe has been a significant point of contention. We combined entomological survey with movement simulation of air-borne particles and population genetics to reconstruct the chain of events that led to a newly colonized French area nestles at the northern foot of the Pyrenees. Simulating the movement of air-borne particles evidenced frequent wind-transport events allowing, within at most 36 hours, the immigration of midges from north-eastern Spain and Balearic Islands and rare events allowing their immigration from Corsica. Population genetics analyses discriminated Corsica as the origin of the new population and completed the puzzle by identifying a couple of successive colonization events within west-Mediterranean basin. Our findings are of considerable importance when trying to understand the invasion of new territories by expanding species.

Keywords: *Culicoides*, distribution range, population genetics, entomological survey, Mediterranean region

Introduction

Rapid changes in the geographic distribution of arthropod species, including incursion into new regions, can have major ecological and economic impacts [1]. Among vectors of human or livestock-associated arboviruses, the most high profile recent examples of this phenomenon have been among the Aedenine mosquitoes, particularly *Stegomyia albopicta* (= *Aedes albopictus*) (Skuse), *Hulecoeteomyia japonica* (= *Ae. japonicus*) (Theobald) and *Hu. koreica* (= *Ae. koreicus*)[6] (Edwards) [2, 3, 4]. Several lifecycle characteristics of these species facilitate their long-distance dispersal; most importantly their possibility to diapause at egg stages that allows adaptation to periods of desiccation, thus to the exploitation of transient water sources. This has facilitated long-distance migration via global trades of plants and used tires [5, 6]. From then, the establishment of new populations depends on the suitability of climatic and environmental conditions at the place of arrival. Rapid expansions in distribution associated with global trade have not generally been reported for the genus *Culicoides* Latreille (Diptera: Ceratopogonidae) [7, 8], although this has been hypothesized for *C. jamaicensis* Edwards [9] and *C. belkini* (Wirth and Arnaud) [10].

At a more local scale, however, the distribution of major vector species of arboviruses can change according to environmental parameters and in turn influence disease distribution. An example is the primary Australian vector of bluetongue virus (BTV) *C. brevitarsis* Kieffer. The southern limit of distribution of this species on the central coastal region of New South Wales varies significantly with climatic variables and this in turn determines the limit of BTV and Akabane virus (AKAV) distribution each year [11, 12]. Within Europe, it has been hypothesized that changes in the northern limits of *C. imicola* Kieffer in the Mediterranean basin have occurred and coincided with an unprecedented expansion of BTV in this region [13, 14, 15]. This is however challenged by recent genetic analyses that supported a long time presence of *C. imicola* in the Mediterranean basin [16, 17]. *Culicoides imicola* is the primary afrotropical vector species of BTV and African horse sickness virus see review in [18, 19].

A key challenge for assessing the recent invasive hypothesis was that systematic data regarding the distribution of *C. imicola* prior to BTV incursions were rarely available. Entomological evidences of *C. imicola* presence in southern Europe (i.e. Balearic Islands, Italy, France, and continental Greece) dated from less than fifteen years [13, 20, 21, 22, 23]. No prior

[6] The Aedini classification has been revised with most subgenera recognized now as genera. We have followed the classification and species names as updated by Harbach on http://mosquito-taxonomic-inventory.info.

extensive surveys are available to indicate that these territories were *C. imicola*-free before the first records. That being said, established populations of the species seem to have expanded their range by colonizing new habitats at the northern limit of the distribution range. Indeed, reinforcement of entomological surveys allowed recording the presence of the species in Catalonia, Spain [24] and in Var department, France [13]; the observed low abundance of captured insects [25] and physiological status [13] suggest a recent northward expansion of these populations at the northern edge distribution of *C. imicola*.

In the Iberian Peninsula, the first recorded BTV outbreaks occurred in the 1960's but confirmed presence of *C. imicola* populations was first reported in 1983 in Spain [26] and soon afterwards in Portugal [27]. The latitude 40°N (i.e., that of Madrid) was then described as the northernmost limit of *C. imicola* with high abundances and continuous distribution characterizing the south-west quarter of the Iberian Peninsula [25, 28, 29]. *Culicoides imicola* was observed in the Balearic Islands in 2001-2002 [20]. In 2002, the first detection of *C. imicola* in a coastal site of Catalonia (~41-42°N) marked a new incursion step toward the northern expansion of the species distribution [24]. The authors hypothesized that this establishment in Catalonia resulted from a windborne dispersal event from the Balearic Islands where *C. imicola* was found at high abundance [24].

Culicoides imicola was recorded in Corsica in 2000 [21] and in the south-east of continental France (Var department) in 2003 [13]. The establishment of *C. imicola* in the Var department was subsequently confirmed through extensive trapping surveys. There, the local expansion of the species distribution was estimated as 14.5 km/year; it was hypothesized to be limited by physical barriers and the limitation of both suitable larval habitats and suitable hosts for blood-feeding [13].

This apparent colonization event and the presence of populations of *C. imicola* in neighbouring countries to mainland France has led to the question of whether incursions of this species will continue to occur into new areas [30, 31]. Indeed, a recent ecoclimatic niche model predicted that additional habitats will become suitable for *C. imicola* colonization in Western Europe under climate change scenarios and predicted northward range expansion along the Spanish and French border [32]. As part of a risk assessment of this scenario the potential expansion of *C. imicola* from Catalonia to the south of France (Pyrénées-Orientales department) was therefore investigated from 2002 onwards [13]. Three individuals where captured in the Pyrénées-Orientales department in 2008, questioning on the presence of *C.*

imicola in this region [13]. This paper subsequently records a new wave of range expansion and the establishment of *C. imicola* in a new area of the French mainland. We then use a unique combination of population genetics, genetic assignment analyses and modeling of long-distance dispersal to trace the origin of these populations in relation to neighboring areas.

Results

Entomological surveys

Within the entire study area, 2,375 nights of trapping were conducted from 2008 to 2012 at 15 sentinel sites along the French-Spanish border (with traps surveyed on a yearly basis) and at 18 monitoring sites in France and Spain (with traps surveyed on monthly or weekly bases) (Supplementary Table 1). In 2012, trapping at the two sentinel sites 25 and 29 (Fig. 1) was stopped before the abundance peak of *C. imicola* (early autumn) (Supplementary Table 1). In Spain, *C. imicola* was collected at 10 of the 12 monitoring sites including two sites (Piera and Susqueda) that had remained positive along four consecutive years (Supplementary Table 1, Fig. 2). In France, *C. imicola* was reported once at a monitoring site during the five years of survey (monitoring site at St-Jean-Pla-de-Corts, site 9). *Culicoides imicola* was trapped in 6 out of the 15 sentinel sites with the highest records observed in 2012 (11 individuals / night). Maximum catches of *C. imicola* were relatively low at sentinel sites (< 11 females / night) and monitoring sites (< 24 females / night), except at two Spanish monitoring sites (Caldes and Susqueda) where more than 250 individuals were regularly collected (i.e., 250 individuals / night from 2009 to 2011), indicating established *C. imicola* populations (Fig. 2).

Within population genetic diversity

We genotyped a total of 483 *C. imicola* adults sampled over 16 sites (Fig.1). A total of 1,107 base pairs of mitochondrial genes COI (474 bp) and CytB (633 bp) were sequenced for 132 of these individuals. The analysis of the concatenated mitochondrial data provided a total of 31 haplotypes, among which two (H2 and H7) were dominant and distributed across the populations. The level of genetic variability within populations was comparable among sites ($0.67 \pm 0.12 \leq H_d \leq 0.95 \pm 0.04$) (Supplementary Table 2).

Population genetic structure

The microsatellite pairwise allelic test failed to detect linkage disequilibrium among loci within sample-sites. All populations were at Hardy-Weinberg equilibrium where F_{IS} estimate

ranges from -0.038 to 0.140 (Supplementary Table 3). While tests based on the IAM mutation model suggested potential signatures of past genetic bottlenecks in samples collected in Algeria, Var, Corsica, Pyrénées-Orientales and Sardinia, those based on the most realistic TPM and SMM mutation models were only significant for Roquebrune-sur-Argens (Var department, France) under the TPM model (Supplementary Table 3).

The microsatellite Bayesian clustering analysis defined two geographical groups, ΔK was clearly maximum for K = 2 (ΔK_{max} = 33): a "western cluster" including Morocco, Spain, Portugal and Majorca, and a "central cluster" consisting of Algeria, Corsica, Sardinia, Pyrénées-Orientales and Var departments (Fig. 3). This spatial genetic structure was consistent with that obtained with the microsatellite Neighbor-joining tree (Fig. 3). This was further supported by the median-joining mitochondrial haplotype network, which displayed strong genetic relationships between Pyrénées-Orientales, Sardinian, Algerian and French populations (Var department and Corsica) while the Spanish populations were genetically closer to those in Portugal and Morocco (Fig. 4). These genealogical relationships were also supported by the Bayesian phylogenetic tree (Fig. 5) and the mitochondrial pairwise F_{ST} values (Supplementary Table 4). Interestingly, the Bayesian clustering analysis and microsatellite neighbor-joining tree suggested that Catalonian population (Girona) is genetically similar to all other continental Spanish populations. Likewise, midges from the Balearic Islands (Majorca) were most closely related to Moroccan and Continental Spanish populations.

Considering the hierarchy of sampling, significant differentiation was detected between both genetic clusters (Fcluster-total = 0.016; P = 0.0001) but also within clusters (Fpopulations-clusters = 0.012; P = 0.0001).

Despite the geographical distances involved, pairwise F_{ST} estimates based on microsatellite data remained relatively low ($F_{ST} \leq 0.07$; Table 2). The genetic differentiation tests were significant for several pairwise comparisons; and particularly when estimating among two populations that did not belong to the same genetic cluster inferred by STRUCTURE (Table 2).

Genetic inference of colonization pathways

We tested the sources and routes of colonization of *C. imicola* in Pyrénées-Orientales using ABC methods. Our results support the scenario involving Corsica as the source of Pyrénées-Orientales populations. More specifically, the most probable scenario entails a succession of three colonization events: the colonization of Sardinia by North African individuals, followed

by the colonization of Corsica by Sardinian founders, and that of Pyrénées-Orientales by Corsican emigrants (P = 0.62, 95% CI= [0.60 - 0.64]; Fig. 6, Supplementary Table 5). The type I and type II errors associated to this scenario were evaluated as 0.28 and 0.06, respectively (Supplementary Table 5). Model checking was carried out for the selected scenario. None of the summary statistics (used and unused for ABC inferences) displayed low probability (i.e. P < 0.05), indicating that the selected scenario fits well the observed data (Supplementary Table 6). This is also confirmed by a Principal Component analysis (PCA): PCA points simulated from the posterior predictive distribution grouped together closely and centered on the target point corresponding to the real dataset (Supplementary Fig. 1).

Long-distance dispersal model outputs

The areas of the study region most likely to have been source regions of windborne *C. imicola* were assessed using the NAME model. The resulting air frequency map shows that air arriving at the entry point (Saint-Jean-Pla-de-Corts, site 9 in Fig. 1) during the full studied time period (1st of August to 31st of October 2003 to 2008) frequently came from north-eastern Spain and Balearic Islands (Fig. 7, left panel). At some periods however, rare wind-borne transport events lead northern Corsica (Fig. 7, right panel) to be the most likely sources for *C. imicola*. Air only occasionally arrived at the trap site from Corsica, other parts of southern France, parts of Italy and northern coast of Africa within the 36 hour time limit.
The individual trajectory maps described the same situation. Full 36-hour back-trajectories for all particles together are presented for each day during the full observation period in supplementary file video clip 1.

Discussion

This study reports a second incursion of *C. imicola* in continental France up-north the apparent northern edge of the species distribution. By using a combination of standard population genetics and approximate Bayesian computation methods, we were able to determine that this newly discovered population was not closely related to the nearby (~80 km south) populations settled in Catalonia. Instead, the newly settled *C. imicola* population was showed to be closely related at both nuclear and mitochondrial genetic loci from the far more distant populations (360 to 1,000 km east or south-east) settled in the Var department, Corsica, Sardinia and Algeria. Corsica was further supported as the most likely source of introduction by the ABC analyses, suggesting that establishment of *C. imicola* in Pyrénées-

Orientales could have occurred through long-distance dispersal from abundant populations in the island (> 500 km from the mainland sampling site). However, other potential population sources such as smaller populations in the Var department or yet undiscovered populations (despite entomological surveillance in this area) between these on the southern coast of France cannot be totally discounted.

Research on the dispersal activity of *Culicoides* is divided into two main areas of focus. Long-distance semi-passive flights on prevailing winds over water bodies have been investigated as a means of both predicting and retrospectively identifying sources of incursions see [33 for a review]. In the current study, we used NAME to simulate the potential for *Culicoides* dispersal to Pyrénées-Orientales and found that trajectories centered primarily on directly surrounding areas, including north-eastern Spain and Balearic Islands. These trajectories also sometimes comprised simulated particles originating from distant areas including northern Corsica and Sardinia, suggesting that midges' dispersal from these sources were possible, but related to rare wind-transport events during the period of abundance of this species.

Although the Pyrénées is a limited elevated mountainous chain, it appears to shape the *C. imicola* population genetic structure more than expected. NAME has been most successfully applied to trajectory simulations over water bodies and would require adaptations to be applicable for local-scale movements over land due to the influence of topographical complexity. Abundance of population sources is also a key factor to take into account. The probability to reach a point by long-distance dispersal depends on the number of active midges that will spread and then survive during transportation. The low abundances observed in Catalonia (maximum catch ~ 12,000 individuals per night), Balearic Islands (mean number 5 – 26 individuals per night per trap) [34] and the Var department (> 100 individuals per night and maximum catch > 4,001 individuals per year) [13] compared to Sardinia and Corsica (30,000-100,000 individuals per night) [13, 35], suggest that these populations unlikely to act as a seed source. A combination of high abundance and favorable winds may support the dispersion of midges from Corsica reaching Pyrénées-Orientales.

Combining the results provided by the NAME model and genetics approach suggests that long-distance dispersal events contribute to *C. imicola* introduction and colonization of new areas. Our genetic analyses also allowed assessing the origin of the Catalonian populations. We discounted the previous hypothesis of the Catalonian population being sourced from the Balearic Islands via windborne dispersal [24]. The microsatellite neighbor-joining tree as well as

the Bayesian clustering analysis indicates instead that the Catalonian population is genetically closer to any other continental Spanish populations than to the insular Balearic population. Moreover, North-Africa appears as a much more likely source of the Balearic populations than Sardinia, which hosts *C. imicola* populations closely related to the French ones.

A second major area of current research in *Culicoides* flight is active dispersal in random directions that can reach 2.21 km daily. This has been investigated recently in northern Europe using capture, mark recapture (CMR) techniques based on fluorescent dusts [36, 37] or immunomarking [38]. Historically, the maximum distance that a recapture has made in this type of study is at 6 km in the peculiar case of *Culicoides mohave* Wirth in the USA [39], a species breeding in desert area. Interestingly, the speed of colonization recorded for *C. imicola* populations over land in the Var region appears to be limited [13]. This may be a consequence of low population density in the Var region [13] and landscape barriers to population spread. The inland limit of *C. imicola* in the Var region in France appears to be restricted by the South Alps. This is consistent with intensive surveys at several sites along the French Mediterranean coast that failed to detect *C. imicola* outside this region between 2002 and 2010 [13]. Nonetheless, more targeted surveys of the southern coast of France for further *C. imicola* populations would be useful in ensuring that the range of this species has not been overlooked in these areas. The investigation of landscape barriers to dispersal of *Culicoides* remains a relatively poorly investigated area. Studies of local-scale landscape ecology could fall below the resolution of genetic techniques, such as microsatellite analysis. In this regard, the use of genome-wide single nucleotide polymorphisms (SNPs), accessed via next-generation sequencing methods, may provide greater resolution at a local scale and advance our understanding of population processes [40]. This may in turn enable improvements in the accuracy of predictive models for *Culicoides* dispersal over land through integration of meteorological, landscape and activity-based parameters [33].

The influence of globalized transport on *Culicoides* dispersal and colonization of new areas remains poorly understood. The introduction of infected *Culicoides* into Europe via these routes has been cited as one of many potential points of entry of arboviruses, but direct data remains extremely limited [41]. *Culicoides* have been recorded as being present at low number on aircraft (number unknown) [8] or ships (~ 1 adult / ship) [7], even if such estimates are probably underestimated due to the logistical challenges of sampling. Recent modeling analyses showed that the risk of introduction of infected *Culicoides* via transport and trade networks to Spain from other European countries is low [42, 43] although these studies are

largely based upon very poorly defined parameters. In the current study, Corsica, the Var department, Algeria and Sardinia share no major ruminant or equine trade links with Pyrénées-Orientales, suggesting that windborne dispersal remains the most likely migration means among these localities.

Except in two sites in Spain, the observed *C. imicola* abundance remains very low in the French and Spanish study sites, and no massive expansion is observed, as also observed in the Var department [13]. The role of adverse meteorological conditions (wind, rain) on *Culicoides* population dynamics has been described and may have influenced our results on species abundance. This probably explains the overall low number of *Culicoides* collected in 2009 in France (the week of prospection was particularly rainy and windy). The relatively limited abundance in monitoring sites compared to other parts of the *C. imicola* distribution area e.g. [44] could be explained by climatic conditions that might be less suitable in this region and/or by the fact that this region is presumably the northern edge of *C. imicola* distribution.

Our work highlights that observation bias related to entomological surveys could lead to misinterpretation of routes and population sources of colonization, especially when the targeted species is a small size and highly passive dispersive species. Our results are consistent with the hypothesis of an introduction by winds, into Pyrenées-Orientales from Sardinia. The combination of independent approaches using population genetic analysis and modeling of long-distance dispersal of *Culicoides* confirm the importance of windborne transport for the spread of exotic species and infected females. Facing numerous signals of long dispersal of *Culicoides* populations, one should now estimate the frequency of these events, especially when outbreaks are declared in Northern Africa while free statuses are maintained in continental areas.

Methods

Entomological surveys and species identification

Thirty-three sites in France and Spain were sampled for *Culicoides* from 2008 to 2012 (Fig. 1). Two levels of sampling effort can be distinguished (Supplementary Table 1, Fig. 1, Fig. 2): monitoring sites were used in the national surveillance network for *Culicoides* populations in the two countries and operated throughout the year on a weekly or monthly basis; sentinel sites in the Pyrénées-Orientales department (France) were visited once a year to survey *C. imicola* expansion from the 2008 detection point (Supplementary Table 1, Fig. 2). Surveys of sentinel sites were carried out during early autumn (September/October) to match

the abundance peak of *C. imicola* [13]. Sampling was carried out using ultra-violet light-suction traps (Onderstepoort design) in France and miniature CDC black light traps in Spain, in close proximity to animal shelters containing sheep, cattle or horses and operated from dusk to dawn. Collections were stored in 90% ethanol prior to species identification. Morphological identification of *C. imicola* within samples was carried out to species level using wing pattern [21, 45].

Population genetics

DNA extraction and amplification. A total of 483 *C. imicola* individuals from 16 localities in North Africa and south-western Europe were used for microsatellite analyses, and a portion of the mitochondrial genes Cytochrome oxydase subunit I (COI) and Cytochrome b (CytB) were sequenced for 132 successful genotyped individuals (Table 1, Fig. 1). Microsatellite data as well as COI and CytB sequences from eight of the localities were previously published in [16] (see details in Table 1). Genomic DNA was extracted from single adult *C. imicola* using a NucleoSpin96 Tissue Kit (Macherey-Nagel, Duren, Germany) according to the manufacturer's instructions. Nuclear genotyping was conducted at 9 microsatellite markers previously developed for *C. imicola* by Mardulyn et al [17] (Supplementary Table 7) and following the protocol described in [16]. Insects were sequenced for the mitochondrial genes COI and CytB using the primers C1J1718/C1N2191 and CytB_12329F/CytB_13038R, respectively, as described in [16].

Sequence analyses. All the sequences were edited and aligned with ClustalW algorithm implemented in the software GENEIOUS v.6.0.5 (Biomatters, www.geneious.com). COI and CytB data sets were analysed separately and showed the same pattern but with a lower resolution. We thus combined COI and CytB data for all analyses.The genetic diversity was estimated by computing the number of haplotypes (H), haplotype diversity (H_d) and nucleotide diversity (π) using DNASP v.5 [46]. The relationships and the geographical distribution of genetic variation among sites were explored with a median-joining network [47] conducted in Network v.4.6.1.2 (www.fluxus-engineering.com) on the concatenated COI and CytB dataset. Genealogical relationships were further investigated by a Bayesian phylogenetic inference as implemented in MRBAYES v.3.2.2 [48]. The software JMODELTEST v.2.1.3 [49] was used to assess the best-fit substitution model based on the Akaike Information Criterion (AIC). The phylogenetic tree was estimated after 1 million generations of four Markov chains ran twice and sampled every 100 generations. Chain convergence was checked with Tracer

v.1.6 software [50] and the first 2,500 generations were discarded as burn-in phase. Finally, population structure was assessed by computing pairwise F_{ST} values between populations.

Microsatellite analyses. The genotypic profile of each individual was characterized with the software GeneMapper® 4.0 (AppliedBiosystems). Linkage disequilibrium between all pairs of loci was tested using FSTAT v2.9.3.2 [51]. Within-population departure from Hardy-Weinberg proportions was investigated by estimating the inbreeding coefficient (F_{IS}). The significance of this estimator was assessed by randomizing alleles among individuals within samples (10,000 permutations). To visualize the genetic relationships between the sampled sites, we constructed a neighbor-joining (NJ) tree [52] based on the pairwise genetic distances of Cavalli-Sforza and Edwards using the software POPULATIONS v.1.2.30 (http://bioinformatics.org/~tryphon/populations/). The robustness of nodes was evaluated by carrying out 1,000 bootstrap replicates.

The Bayesian approach implemented in STRUCTURE v.2.3.3 [53] was used to infer spatial genetic structure. We assumed an admixture model with correlated allele frequencies [54] and used the sampling locations (Locprior model) as priors' information [55]. For each value of the number (K) of clusters set between 1 and 14 (number of sampled sites), we performed 10 independent runs of 10^6 Markov chain Monte Carlo (MCMC) iterations with a burn-in of 10^5. The most probable number of clusters was inferred using ΔK method [56].

The relative importance of the genetic clusters previously inferred by STRUCTURE and the populations in differentiation was assessed with the multilocus hierarchical F-statistics Fpopulations-clusters and Fclusters-total, respectively. This analysis was performed with hierfstat package [57]. These tests were based on 10 000 permutations of either culicoides genotypes among populations and within clusters (H0: 'Fpopulations-cluster = 0'), or populations among clusters (H0: 'Fclusters-total = 0'). Genetic differentiation among samples was further assessed through the Weir and Cockerham [58]'s unbiased estimates F_{ST} and the significance was tested using the exact G test over 10,000 permutations of genotypes among samples as implemented in FSTAT v2.9.3.2 [51].

In populations that have undergone a sharp decrease in effective population size, the loss of alleles is faster than the decline of genetic diversity (H_S). This results in an increase of heterozygosity across loci. The program BOTTLENECK allows testing of this event in a representative sample of individuals [59]. It has been shown that past bottleneck events will be detected with a high degree of sensitivity using the Infinite Allele Mutation (IAM) model, moderately with the two-phase model (TPM) and dimly with the Stepwise Mutation Model

(SMM) [60]. We therefore performed the unilateral Wilcoxon test under the three proposed mutation models [60]. For the TPM model the proportion of SMM was set to 70% and the variance to 30 (default values). The significance was assessed by performing 10,000 replicates.

Inference of colonization pathways. Microsatellite data were used to investigate the source of *C. imicola* individuals in Pyrénées-Orientales (Continental France) and test hypotheses regarding the observed genetic clusters using approximate Bayesian computation (ABC). Our hypotheses addressed three potential sources of *C. imicola*: Catalonia, Corsica, Sardinia or North Africa. We tested four demographic scenarios presented in Supplementary Table 4 and Fig. 6 with DIYABC software v.2.0.4 [61, 62]. Data were simulated under demographic, historical and mutational parameter values used as priors' information given in Supplementary Table 8. We assumed 10 generations per year [63], a divergence time starting 40 generations ago with 10,000 generations of uncertainty, and a mutation rate ranging from 10^{-6} to 10^{-4}. Genetic variation within and between populations was summarized using a set of statistics implemented in DIYABC including the mean number of alleles, the mean expected heterozygosity [64], the mean allelic size variance, the Garza-Williamson's M (mean ratio of the number of alleles over the range of allele sizes) [65], pairwise F_{ST} values [66] and the classification index (*mean individual assignment likelihood*) [67]. The posterior probabilities for each of the competing scenarios were calculated by a polychotomous logistic regression [61, 62] on 1% of the simulated data sets similar to the observed data set. Confidence in the selected scenario was evaluated by analyzing 100 simulated pseudo-observed data sets (pods) with the same number of loci and individuals as our data set. The parameter values drawn from prior distribution (Supplementary Table 8) and LDA-transformed summary statistics were used to calculate type I and II errors. These latters refer to the probability of excluding the selected scenario when it is true and the probability of selecting the scenario when it is false, respectively. Mean type II error was calculated over the competing scenarios. Finally, we assessed the goodness of fit of the selected scenario by using the model checking option of DIYABC software [61], which allows evaluating whether the selected scenario and associated posteriors distributions match well with the observed genetic data of *C. imicola*. As recommended by Cornuet et al. (2010), we used as test statistics the DIYABC summary statistics not used for model selection in previous ABC treatments. Because this analysis may suffer from non-independence between the summary statistics, we also performed a principal component analysis (PCA) in the space of the summary statistics.

Model of long-distance biting midge dispersal

Possible windborne incursion of *C. imicola* into the study region were assessed using the Numerical Atmospheric-dispersion Modelling Environment (NAME) Lagrangian model, designed to simulate the release, transport, mixing and transformation of airborne gases or particulates and their subsequent depletion or removal from the atmosphere [68]. The release and dispersion of hundreds of thousands of model particles allows for representation of the stochastic nature of the atmosphere. The motions of the particles are determined by the ambient three-dimensional wind flow with a random component superimposed to simulate turbulence. The underlying meteorological data necessary to drive the dispersion model was taken from the UK Met Office's Unified Model [69]. For Aug to Oct 2003 to 2008, the horizontal resolution of the Unified Model over Europe was 12 km with a temporal resolution of 1 hour.

NAME was chosen over other dispersion models as it has been previously used to describe wind-borne incursion events that correlate with the timing and location of outbreaks of BTV in Europe [33, 70] and compared favorable against another complex dispersion model, MATCH, for outbreaks in Sweden [71]. Simpler wind trajectory models have also been used to assess transport of *Culicoides* in the atmosphere [72, 73, 74]. These studies only follow the path taken by one trajectory at very low temporal and spatial resolution (typically 6 hourly at a horizontal resolution of 0.25x0.25°) and therefore cannot account for the stochastic nature of the atmosphere. Other Lagrangian particle-dispersion models are also available, such as the HYSPLIT model used by [75] to assess incursions of *Culicoides* into Australia. However the underlying meteorological data that is freely available to use with this model for our study period and region is only available at 3-hourly intervals with a horizontal resolution of 1°. These scales would not be adequate for modelling the transport of *Culicoides* within the Mediterranean basin.

In this study, the model was run in backwards mode to simulate the source of winds potentially transporting *C. imicola* to the trap location in Pyrénées-Orientales. In backwards mode the wind direction is reversed and the model steps backwards through time. Saint-Jean-Pla-de-Corts (Site 9, Fig. 1) was selected as the entry point in 2008 as this was the first location where *C. imicola* was recorded. The period from 1 August to 31 October covering the peak of *C. imicola* abundance was assumed to be the period most likely for an introduction to the trap location and we thus modeled particles movement for this period from 2003 to 2008. A large number of model particles (30,000) were released in the model from the trap location for each day in the time window and tracked backwards for 36 hours (assumed to be the

maximum flight time for *C. imicola*). At the end of each day's simulation period the total number of particles present in each box of a 0.25°x0.25° grid defined over the region were calculated. The greater the number of particles present in each grid box, the greater the proportion of air arriving at the trap site from that source. To assess where air most frequently arrived from during the likely introduction window, the relative probabilty of pixels as source points for Pyrénées-Orientales was mapped throughout the region (Fig. 7). It was calculated as the total number of particles received in each grid cell from the individual daily simulations divided by the total number of particles received by all the grid cells not located over the sea (which cannot be a source for culicoides populations) for a given period of time. In addition individual trajectories taken by 100 particles on each day in the time window were also calculated and examined to analyse the routes taken by individual air streams. Clustering of trajectories due to a dominant wind pattern can be identified, with some individual trajectories being taken in a very different direction due to turbulence or a separate synoptic system. In Figure 7, we illustrate the fact that in some periods (e.g. 10-20 Oct. 2008), the pattern is very much different from the general pattern (mean values for 2003 to 2008). To illustrate the modeling process, supplementary file video clip 1 presents the 36-hour trajectories with a one hour time step for days 11/09/2008 and video clip 2 shows the full 36-hour back-trajectories for all particles together for each day during the full observation period (2003 to 2008).

Acknowledgements

The authors would like to thank the people that helped with *Culicoides* trappings in France, Spain, Portugal and Northern Africa (EID-Med agents, farmers, Thierry Baldet (Cirad), Catherine Cêtre-Sossah (Cirad), Jean-Claude Delécolle (University of Strasbourg)), and *Culicoides* identification (Xavier Allène (Cirad), Ignace Rakotoarivony (Cirad), Jean-Claude Delécolle (University of Strasbourg)).

Contributions

SJ, KH, TB, HG, CG designed the study. SJ genotyped the samples. SJ, KH, LB, SC, CS, AHD, JB, TB, HG, and CG analyzed the data. NP, ST, SL, MD, YL, JL, MM, IPDF, DWR, and M-LS-R collected the *C. imicola* samples. CC, JB, TB and HG, contributed to the manuscript firstly written by SJ, KH and CG. All authors read and commented the final manuscript version.

Accession numbers

The COI and CytB sequences were deposited in the GenBank Database under the accession number KT026989 to KT027189 and KT027190 to KT027376, respectively.

Competing financial interests

The authors declare no competing financial interests.

Funding information

This work was granted by the Chercheur d'Avenir 2011 call from the Languedoc-Roussillon region, France and the Ministère de l'Agriculture. SJ was partially funded by the Chercheur d'Avenir grant and was supported by the Ecole Internationale de Recherche d'Agreenium (EIR-A). SC, LB and CS were funded by Defra project SE: 4211. This study was partially funded by EU grant FP7-613996 VMERGE and is catalogued by the VMERGE00X Steering Committee as VMERGE000 (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.

References

1. Sanders C, Mellor PS, Wilson AJ. Invasive arthropods. *Rev sci tech Off int Epiz* **29**, 273-286 (2010).

2. Bonizzoni M, Gasperi G, Chen X, James AA. The invasive mosquito species *Aedes albopictus*: current knowledge and future perspectives. *Trends Parasitol* **29**, 460-468 (2013).

3. Kaufman MG, Fonseca DM. Invasion biology of *Aedes japonicus japonicus* (Diptera: Culicidae). *Annu Rev Entomol* **59**, 31-49 (2014).

4. Schaffner F, Medlock JM, Van Bortel W. Public health significance of invasive mosquitoes in Europe. *Clin Microbiol Infect* **19**, 685-692 (2013).

5. Medlock JM, *et al.* A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. *Vector Borne Zoonotic Dis* **12**, 435-447 (2012).

6. Scholte EJ, *et al.* Accidental importation of the mosquito *Aedes albopictus* into the Netherlands: a survey of mosquito distribution and the presence of dengue virus. *Med Vet Entomol* **22**, 352-358 (2008).

7. Nie WZ, Li JC, Li DX, Gratz NG. A preliminary report of introduced living biting midges (Diptera: Ceratopogonidae) found on entry ships arriving at Qinhuangdao Port, China. *Med Entomol Zool* **56**, 359-361 (2005).

8. Reye EJ. The problems of biting midges (Diptera: Ceratopogonidae) in Queensland. *Australian J Entomol* **3**, 1-6 (1964).

9. Meiswinkel R, Labuschagne K, Goffredo M. Christopher Columbus and *Culicoides*: was *C. jamaicensis* Edwards, 1922 introduced into the Mediterranean 500 years ago and later re-named *C. paolae* Boorman 1996? *Vet Ital* **40**, 340-344 (2004).

10. Duval J, Riviere F, Pichon G. Quelques aspects bio-écologiques de *Culicoides belkini* (Wirth et Arnaud, 1969) (Diptera : Ceratopogonidae). *Cah ORSTOM sér Ent med parasitol* **16**, 273-277 (1978).

11. Bishop AL, Spohr LJ, Barchia IM. Effects of altitude, distance and waves of movement on the dispersal in Australia of the arbovirus vector, Culicoides brevitarsis Kieffer (Diptera: Ceratopogonidae). *Prev Vet Med* **65**, 135-145 (2004).

12. Bishop AL, Kirkland PD, McKenzie HJ, Barchia IM. The dispersal of *Culicoides brevitarsis* in eastern New South Wales and associations with the occurrences of arbovirus infections in cattle. *Aust Vet J* **73**, 174-178 (1996).

13. Venail R, *et al.* Assessing diversity and abundance of vector populations at a national scale: example of *Culicoides* surveillance in France after bluetongue virus emergence. In: *Parasitology Research Monographs* (eds Mehlhorn H) pp. 77-102. Berlin, Springer-Verlag (2012).

14. Purse BV, Mellor PS, Rogers DJ, Samuel AR, Mertens PP, Baylis M. Climate change and the recent emergence of bluetongue in Europe. *Nat Rev Microbiol* **3**, 171-181 (2005).

15. Mellor PS, Wittmann EJ. Bluetongue virus in the Mediterranean Basin 1998-2001. *Vet J* **164**, 20-37 (2002).

16. Jacquet S, et al. Colonization of the Mediterranean Basin by the vector biting midge species Culicoides imicola: an old story. *Mol Ecol* **24**, 5707-5725 (2015).

17. Mardulyn P, *et al.* 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 (2013).

18. Mellor PS. African horse sickness: transmission and epidemiology. *Vet Res* **24**, 199-212 (1993).

19. Mellor PS, Boorman J, Baylis M. *Culicoides* biting midges: Their role as arbovirus vectors. *Ann Rev Entomol* **45**, 307-340 (2000).

20. Miranda MA, Borras D, Rincon C, Alemany A. Presence in the Balearic Islands (Spain) of the midges *Culicoides imicola* and *Culicoides obsoletus* group. *Med Vet Entomol* **17**, 52-54 (2003).

21. Delécolle J-C, de La Rocque S. Contribution à l'étude des *Culicoides* de Corse. Liste des espèces recensées en 2000/2001 et redescription du principal vecteur de la fièvre cattarhale ovine: C. imicola Kieffer, 1913 (Diptera: Ceratopogonidae). *Bull Soc Entomo France* **107**, 371-379 (2002).

22. Goffredo M, Conte A, Cocciolito R, Meiswinkel R. The distribution and abundance of *Culicoides imicola* in Italy. *Vet Ital* **47**, 22-32 (2003).

23. Patakakis MJ. *Culicoides imicola* in Greece. *Vet Ital* **40**, 232-234 (2004).

24. Sarto i Monteys V, Ventura D, Pagès N, Aranda C, Escosa R. Expansion of *Culicoides imicola*, the main bluetongue virus vector in Europe, into Catalonia, Spain. *Vet Rec* **156**, 415-417 (2005).

25. Acevedo P, *et al.* A broad assessment of factors determining *Culicoides imicola* abundance: modelling the present and forecasting its future in climate change scenarios. *PloS one* **5**, e14236 (2010).

26. Mellor PS, Boorman J, Wilkinson PJ, Martinez-Gomez F. Potential vectors of bluetongue and African horse sickness in Spain. *Vet Rec* **112**, 229-230 (1983).

27. Mellor PS, Jennings DM, Wilkinson PJ, Boorman J. *Culicoides imicola*: a bluetongue virus vector in Spain and Portugal. *Vet Rec* **116**, 589-590 (1985).

28. Ortega MD, Mellor PS, Rawlings P, Pro MJ. The seasonal and geographical distribution of *Culicoides imicola, C. pulicaris* group and *C. obsoletus* group biting midges in central and southern Spain. *Arch Virol Suppl* **14**, 85-91 (1998).

29. Rawlings P, Pro MJ, Pena I, Ortega MD, Capela R. Spatial and seasonal distribution of *Culicoides imicola* in Iberia in relation to the transmission of African horse sickness virus. *Med Vet Entomol* **11**, 49-57 (1997).

30. Goffredo M, Conte A, Meiswinkel R. Distribution and abundance of *Culicoides imicola,* Obsoletus Complex and Pulicaris Complex (Diptera: Ceratopogonidae) in Italy. *Vet Ital* **40**, 270-273 (2004).

31. Goldarazena A, Romón P, Aduriz G, Balenghien T, Baldet T, Delécolle JC. First record of *Culicoides imicola*, the main vector of bluetongue virus in Europe, in the Basque Country (northern Spain). *Vet Rec* **162**, 820-821 (2008).

32. Guichard S, Guis H, Tran A, Garros C, Balenghien T, Kriticos DJ. Worldwide niche and future potential distribution of *Culicoides imicola*, a major vector of bluetongue and African horse sickness viruses. *PLoS One* **9**, e112491 (2014).

33. Burgin LE, Gloster J, Sanders C, Mellor PS, Gubbins S, Carpenter S. Investigating incursions of bluetongue virus using a model of long-distance *Culicoides* biting midge dispersal. *Transbound Emerg Dis* **60**, 263-272 (2013).

34. Miranda M, Rincón C, Borràs D. Seasonal abundance of *Culicoides imicola* and *C. obsoletus*. *Vet Ital* **40**, 293 (2004).

35. Ippoliti C, *et al.* Can landscape metrics help determine the *Culicoides imicola* distribution in Italy? *Geospat health* **8**, 267-277 (2013).

36. Kirkeby C, Bødker R, Stockmarr A, Lind P, Heegaard PM. Quantifying dispersal of european culicoides (Diptera: Ceratopogonidae) vectors between farms using a novel mark-release-recapture technique. *PLoS One* **8**, e61269 (2013).

37. Kluiters G, Swales H, Baylis M. Local dispersal of palaearctic *Culicoides* biting midges estimated by mark-release-recapture. *Parasit Vectors* **8**, 86 (2015).

38. Sanders CJ, Carpenter S. Assessment of an immunomarking technique for the study of dispersal of *Culicoides* biting midges. *Infect Genet Evol* **28**, 583-587 (2014).

39. Brenner RJ, Wargo MJ, Stains GS, Mulla MS. The dispersal of *Culicoides mohave* (Diptera: Ceratopogonidae) in the desert of Southern California. *Mosq News* **44**, 343-350 (1984).

40. Rasic G, Filipovic I, Weeks AR, Hoffmann AA. Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, *Aedes aegypti*. *BMC Genomics* **15**, 275 (2014).

41. Carpenter S, Wilson AJ, Mellor PS. *Culicoides* and the emergence of bluetongue virus in northern europe. *Trends Microbiol* **17**, 172-178 (2009).

42. Napp S, Garcia-Bocanegra I, Pagès N, Allepuz A, Alba A, Casal J. Assessment of the risk of a bluetongue outbreak in Europe caused by *Culicoides* midges introduced through intracontinental transport and trade networks. *Med Vet Entomol* **27**, 19-28 (2013).

43. de Vos CJ, Hoek CA, Nodelijk G. Risk of introducing African horse sickness virus into the Netherlands by international equine movements. *Prev Vet Med* **106**, 108-122 (2012).

44. Venter GJ, *et al. Culicoides* species abundance and potential over-wintering of African horse sickness virus in the Onderstepoort area, Gauteng, South Africa. *J S Afr Vet Assoc* **85**, 1-6 (2014).

45. Delécolle JC. Nouvelle contribution à l'étude systématique et iconographique des espèces du genre *Culicoides* du Nord-est de la France (eds). Université de Strasbourg (1985).

46. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452 (2009).

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

48. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755 (2001).

49. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**, 772 (2012).

50. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v1.6, Available from http://beast.bio.ed.ac.uk/Tracer. (2014).

51. Goudet J. Fstat version 1.2: a computer program to calculate Fstatistics. *J Heredity* **86**, 485-486 (1995).

52. Takezaki N, Nei M. Genetic Distances and Reconstruction of Phylogenetic Trees From Microsatellite DNA. *Genetics* **144**, 389-399 (1996).

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

54. Falush D, Stephens M, Pritchard JK. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* **164**, 1597-1587 (2003).

55. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Res* **9**, 1322-1332 (2009).

56. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* **8**, 2611-2620 (2005).

57. Goudet J. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes* **5**, 184-186 (2005).

58. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population-Structure. *Evolution* **38**, 1358-1370 (1984).

59. Piry S, Luikart G, Cornuet JM. Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. *J Heredity* **90**, 502-503 (1999).

60. Cornuet JM, Luikart G. Description and Power Analysis of Two Tests for Detecting Recent Population Bottlenecks from Allele Frequency Data. *Genetics* **144**, 2001-2014 (1996).

61. Cornuet JM, Ravigné V, Estoup A. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *Bioinformatics* **11**, 401 (2010).

62. Cornuet JM, *et al.* Inferring population history with DIYABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**, 2713-2719 (2008).

63. Braverman Y, Linley JR. Parity and voltinism of several *Culicoides* spp. (Diptera: Ceratopogonidae) in Israel, as determined by two trapping methods. *J Med Entomol* **25**, 121-126 (1988).

64. Nei M. *Molecular evolutionary genetics*. New York, Columbia University Press (1987).

65. Garza JC, Williamson EG. Detection of reduction in population size using data from microsatellite loci. *Mol Ecol* **10**, 305-318 (2001).

66. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358-1370 (1984).

67. Pascual M, *et al.* Introduction history of *Drosophila subobscura* in the New World: a microsatellite-based survey using ABC methods. *Mol Ecol* **16**, 3069-3083 (2007).

68. Jones AR, Thomson DJ, Hort M, Devenish B. *The U.K. Met Office's next-generation atmospheric dispersion model, NAME III', in Borrego C. and Norman A.-L.(Eds) Air Pollution Modeling and its Application XVII (Proceedings of the 27th NATO/CCMS*

International Technical Meeting on Air Pollution Modelling and its Application), Springer, pp. 580 (2007).

69. Davies T, *et al.* A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q J Roy Meteor Soc* **131**, 1759-1782 (2005).

70. Gloster J, Burgin L, Witham C, Athanassiadou M, Mellor P. Bluetongue in the United Kingdom and northern Europe in 2007 and key issues for 2008. *Vet Rec* **162**, 298-302 (2008).

71. Agren EC, Burgin LE, Lewerin SS, Gloster J, Elvander M. Possible means of introduction of bluetongue virus serotype 8 (BTV-8) to Sweden in August 2008: comparison of results from two models for atmospheric transport of the *Culicoides* vector. *Vet Rec* **167**, 484-488 (2010).

72. Hendrickx G, *et al.* A wind density model to quantify the airborne spread of *Culicoides* species during north-western Europe bluetongue epidemic, 2006. *Prev Vet Med* **87**, 162-181 (2008).

73. Ducheyne E, *et al.* Quantifying the wind dispersal of *Culicoides* species in Greece and Bulgaria. *Geospat Health* **1**, 177-189 (2007).

74. Ducheyne E, Lange M, Van der Stede Y, Meroc E, Durand B, Hendrickx G. A stochastic predictive model for the natural spread of bluetongue. *Prev Vet Med* **99**, 48-59 (2011).

75. Eagles D, Walker P, Zalucki M, Durr P. Modelling spatio-temporal patterns of long-distance *Culicoides* dispersal into northern Australia. *Prev Vet Med* **110**, 312-322 (2013).

Figures and Tables

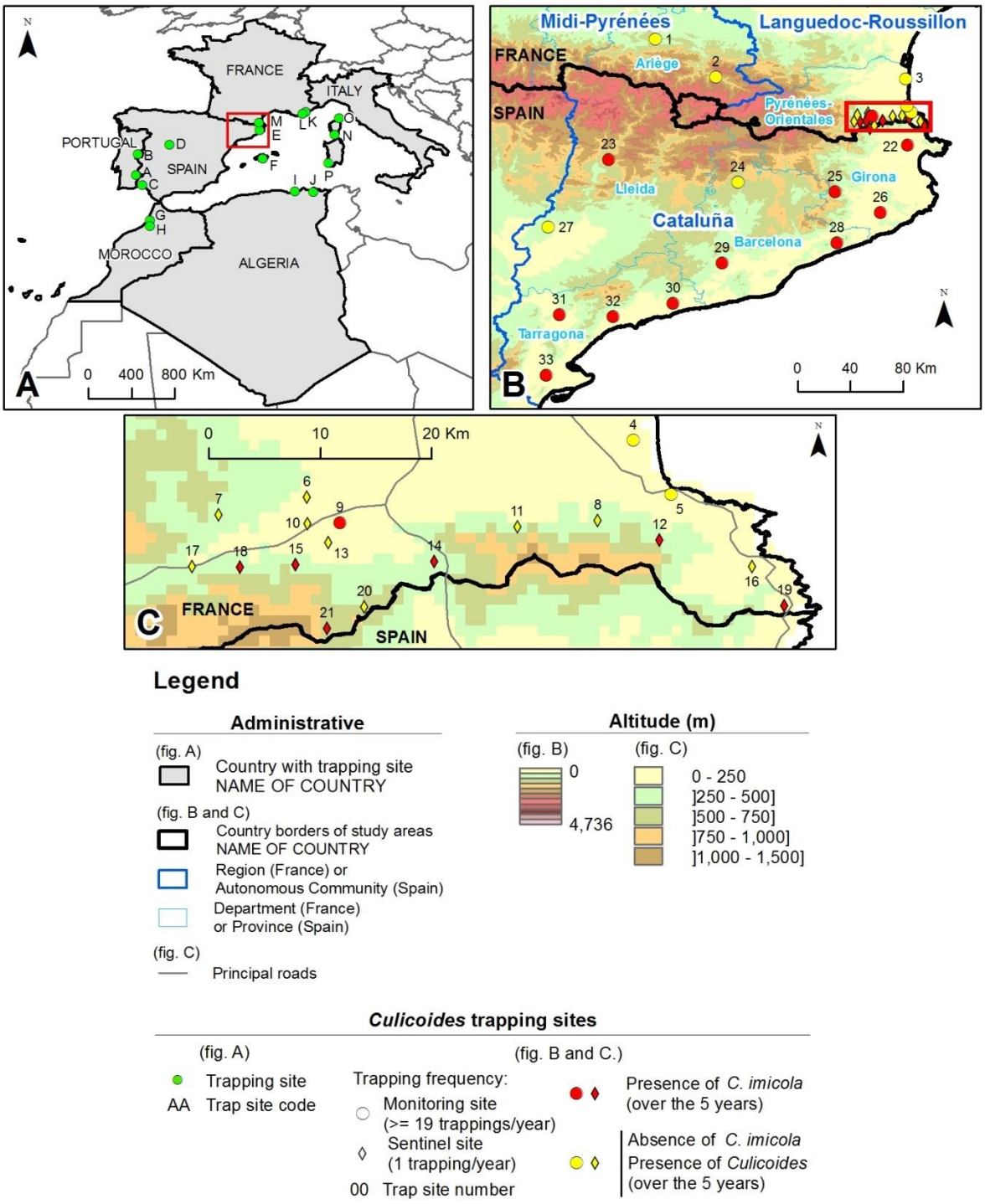


Figure 1 | Sampling sites for population genetic analyses (A) and entomological surveys (B and C). Code sites are detailed in Table 1 and Supplementary Table 1. Maps were generated using ArcGIS software v10.2.2 (ESRI, Redlands, CA).

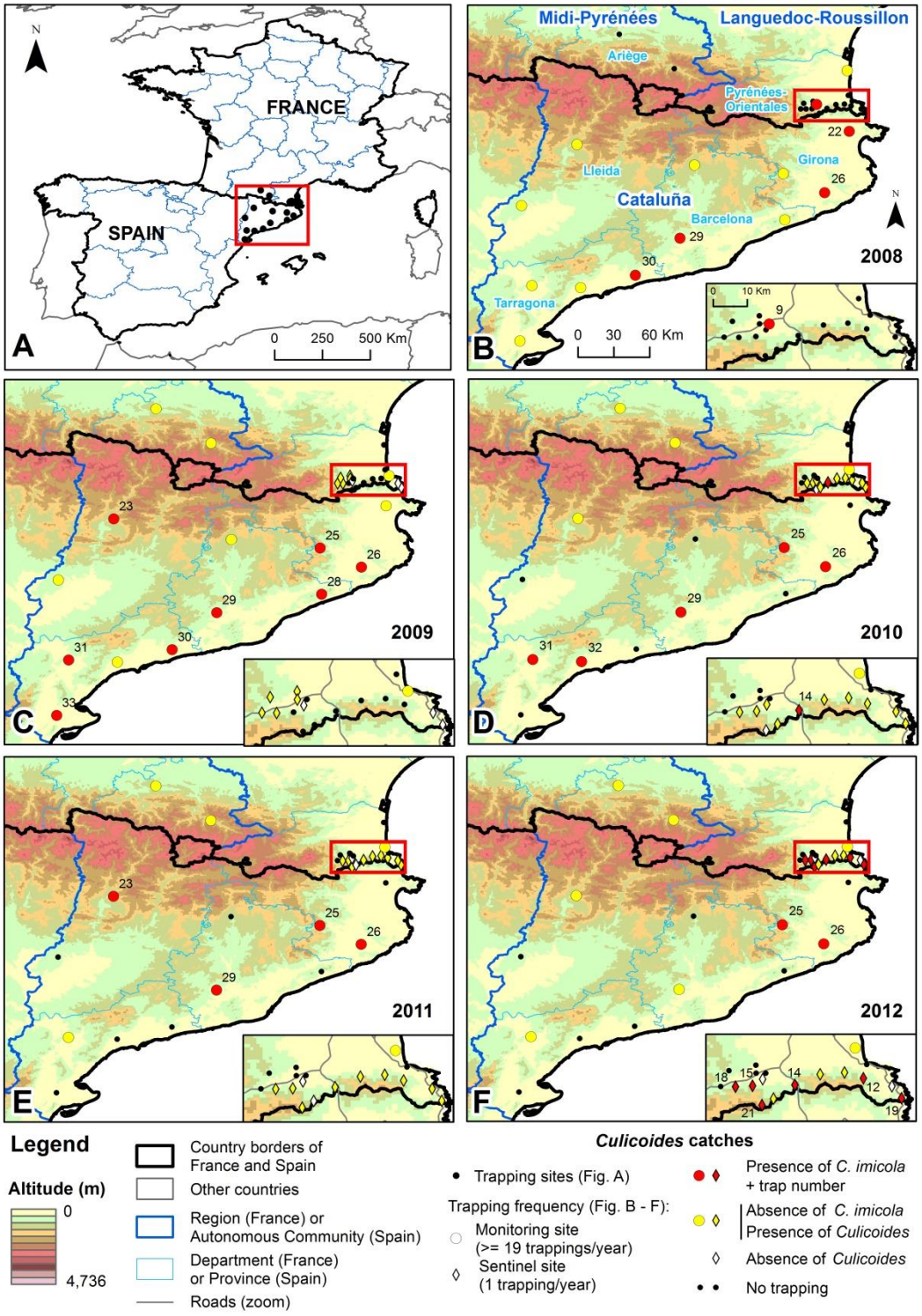


Figure 2 | Presence/absence map of *C. imicola* in Pyrénées-Orientales and Catalonia from 2008 to 2012. Code sites are detailed in Supplementary Table 1. Maps were generated using ArcGIS software v10.2.2 (ESRI, Redlands, CA).

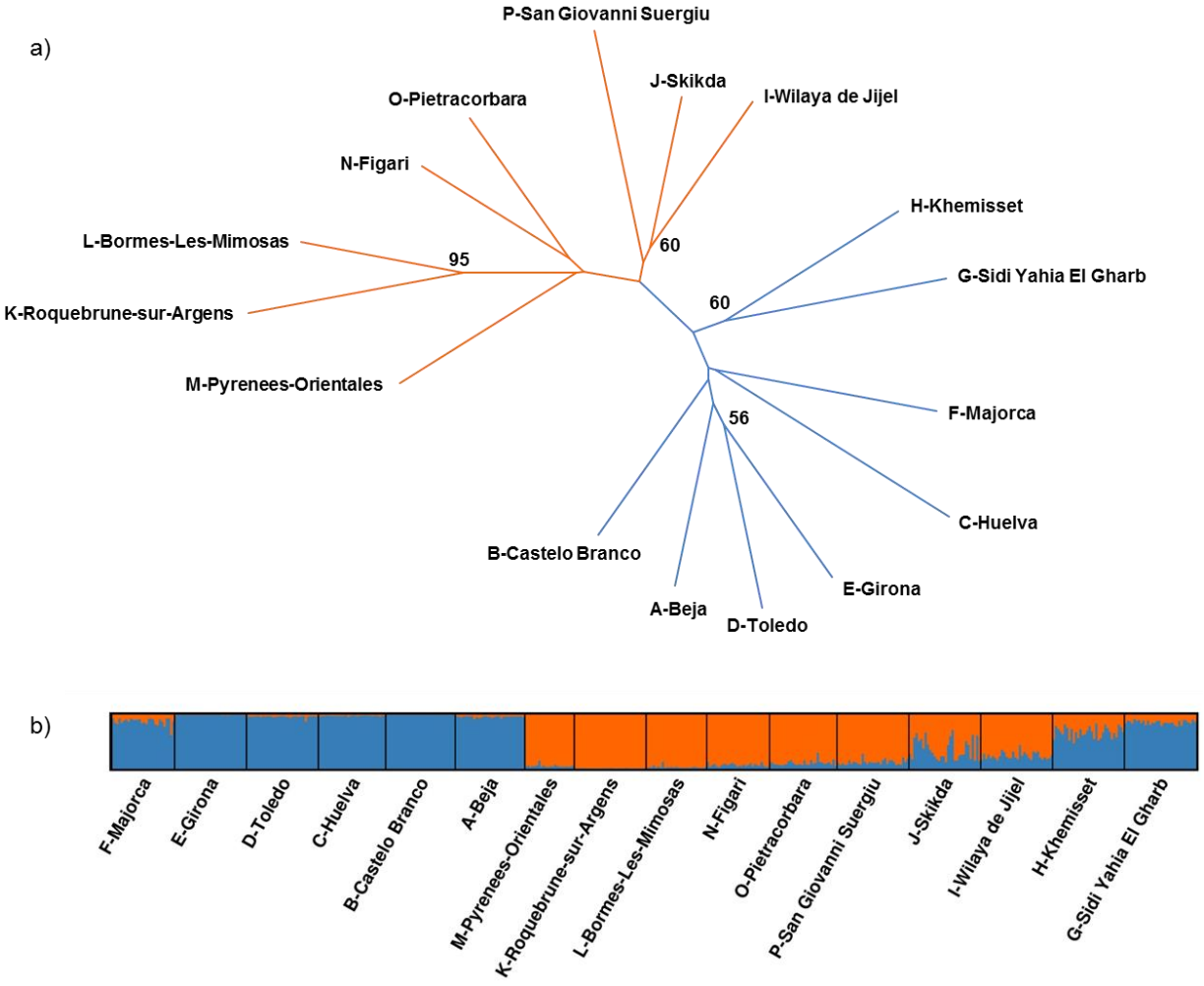


Figure 3 | Microsatellite neighbour-joining tree and genetic clustering of *C. imicola* population samples. (a) The neighbor-joining tree is based on genetic distance of Cavalli-Sforza & Edwards (1967). Bootstrap values are calculated over 1,000 replicates (only values > 60% are shown). (b) Each vertical line represents an individual, and each color represents a cluster. Individuals are grouped by sampling location: Algeria (Skikda, Wilaya de Jijel), Balearic Islands (Majorca), Continental France (Pyrénées-Orientales, Roquebrune-sur-Argens, Bormes-les-Mimosas), Continental Spain (Girona, Toledo, Huelva), Corsica (Figari, Pietracorbara), Morocco (Khemisset, Sidi Yahia El Gharb), Portugal (Beja, Castelo Branco), Sardinia (San Giovanni Suergiu).

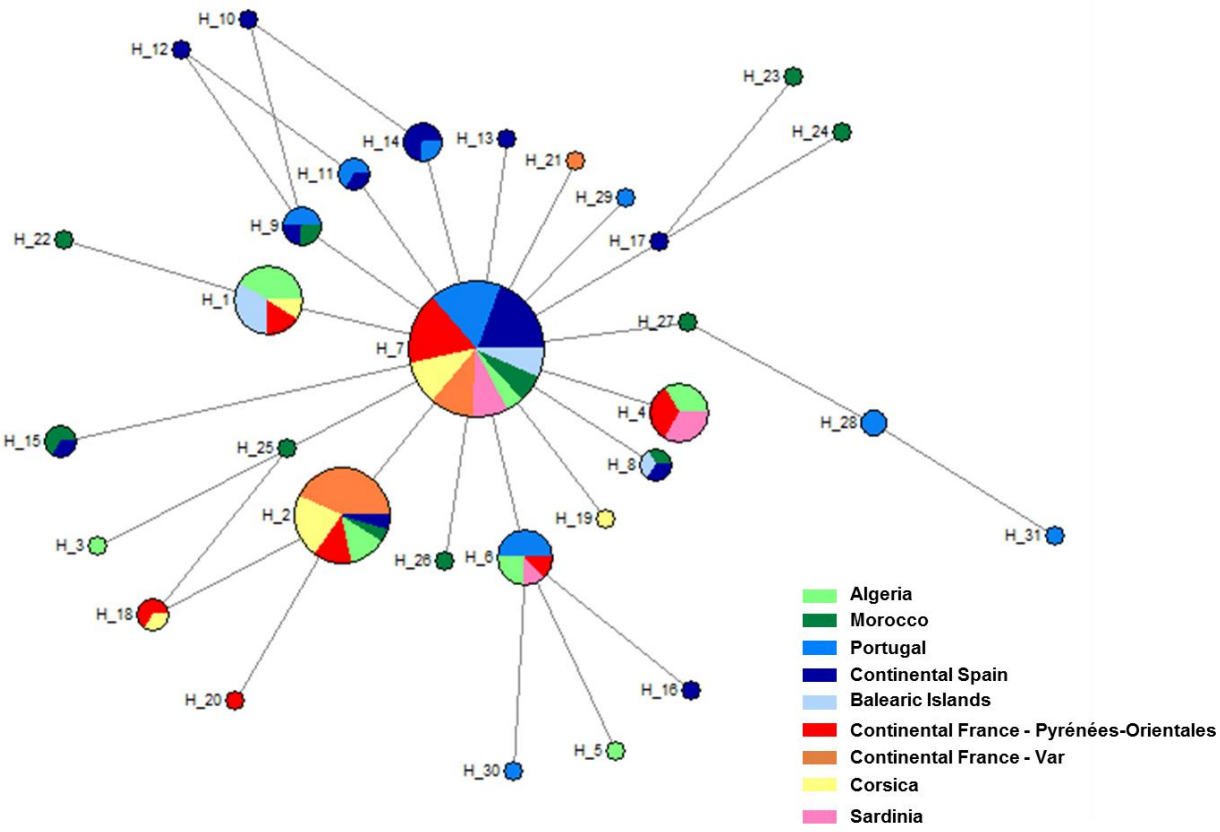


Figure 4 | Median-joining haplotype network. The size of the circles is proportional to the number of individuals with that haplotype. The length of the branches separating haplotypes is proportional to the number of mutational steps between them.

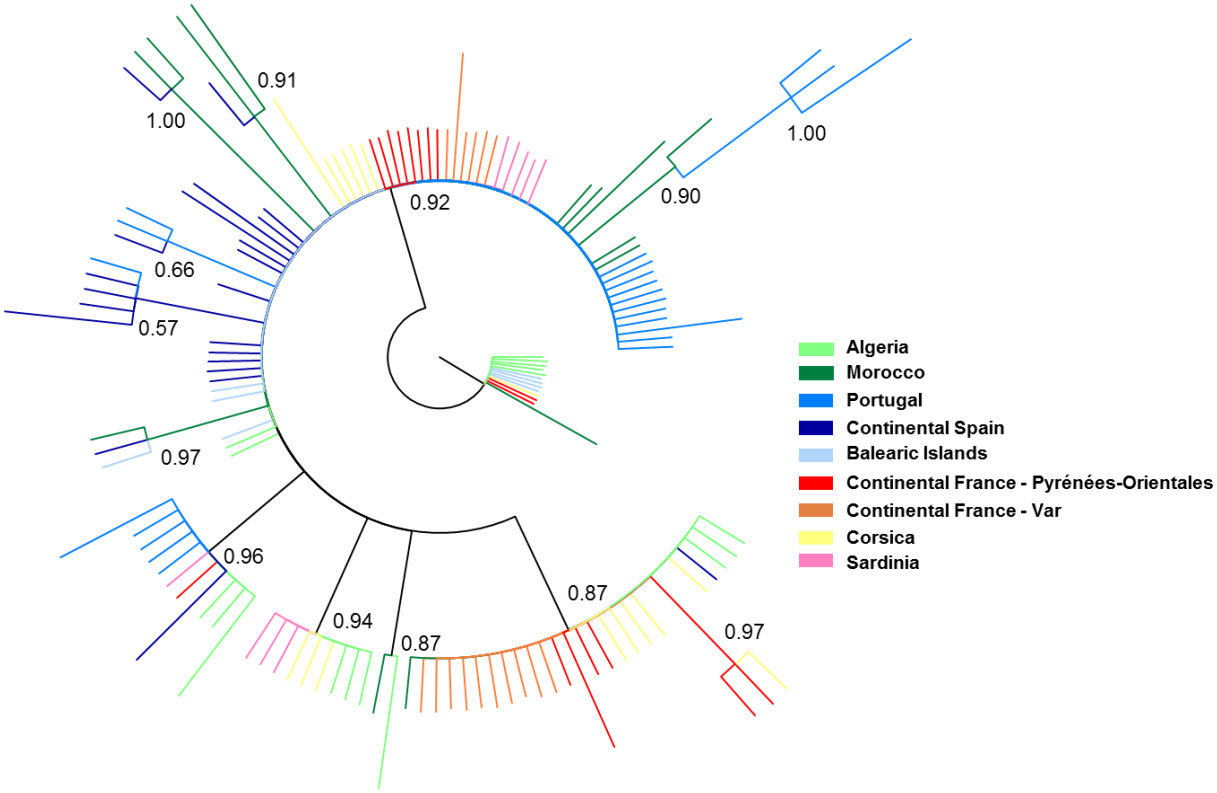


Figure 5 | Mitochondrial Bayesian phylogenetic tree. Numbers represent the posterior probability and each color refers to a geographical region.

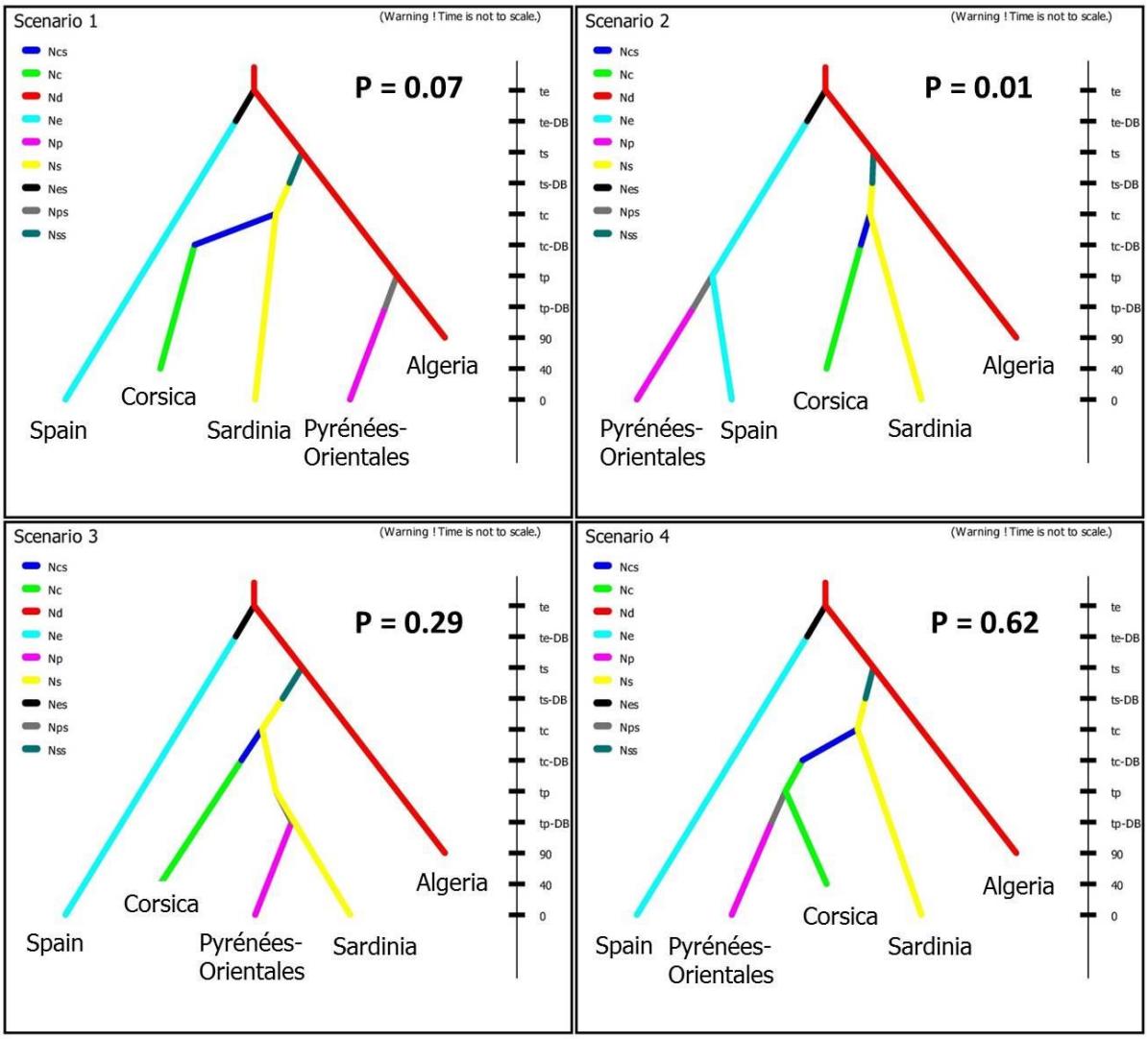


Figure 6 | Graphical representation of the tested scenarios regarding colonization sources of *C. imicola* in the Pyrénées-Orientales. Microsatellite data were used and data were simulated using an approximate Bayesian computation (ABC) approach. The y-axis represents the time of events (not to scale), time 0 being the most recent sampling date. Nc, Ns, Nd, Ne and Np refer respectively to the effective population sizes, stable over the time, of the populations from Corsica, Sardinia, Algeria, Catalonia and Pyrénées-Orientales and Ncs, Nss, Nes and Nps refer to the effective number founder for Corsica, Sardinia, Catalonia and Pyrénées-Orientales populations. P refers to the probability obtained for each scenario. Details of all scenarios and parameters are shown in Supplementary Table 5 and 6.

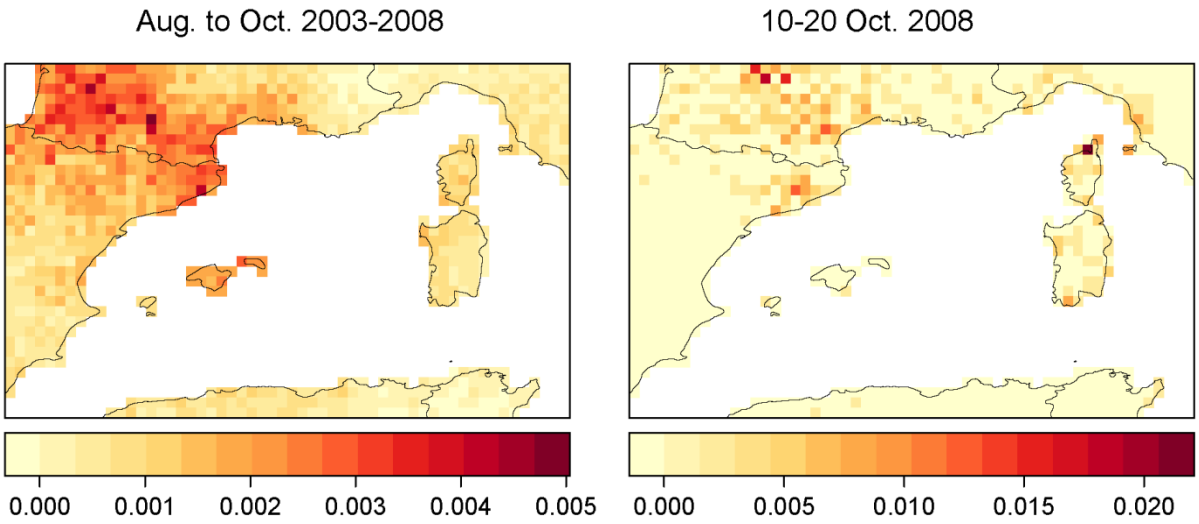


Figure 7 | Source of winds potentially transporting *C. imicola* to the trap location in Pyrénées-Orientales. To generate this map, the NAME dispersion model was run in backwards mode for 36H each day from 1 August to 31 October for the period 2003 to 2008, using 30,000 particles (left panel). We also present the results of the simulations for the period 10-20 Oct. 2008 only (right panel). The probability of pixels as source points for Pyrénées-Orientales was calculated as the total number of particles received in each pixel from the individual daily simulations divided by the total particles received by all the grid cells not located over the sea for each time period. Maps were generated using R software v3.2.2.

Table 1 | Geographical locations, sampling dates and number of *C. imicola* individuals typed for the population genetics analysis.

Country	Location	code	Collection year	Nmic	N_{MtDNA}
Algeria	Skikda	J	2003	32	9
	Wilaya de Jijel	I	2003	32	8
Morocco	Khemisset	H	2002	32	6
	Sidi Yahia El Gharb	G	2004	32	8
Portugal	Castelo Branco	B	2010	31	8
	Beja	A	2010	31	6
Balearic Islands, Spain	Majorca	F	2012	28	8
Continental Spain	Girona, Catalonia	E	2012	32	6
	Toledo, Castilla-La-Mancha	D	2012	32	8
	Huelva, Andalusia	C	2012	30	8
Continental France	Pyrénées-Orientales	M	2012	22	17
	Roquebrune-sur-Argens	K	2008	32	8
	Bormes-les-Mimosas	L	2008	27	8
Corsica, France	Figari	N	2008	28	8
	Pietracorbara	O	2008	30	8
Sardinia, Italy	San Giovanni Suergiu	P	2012	32	8

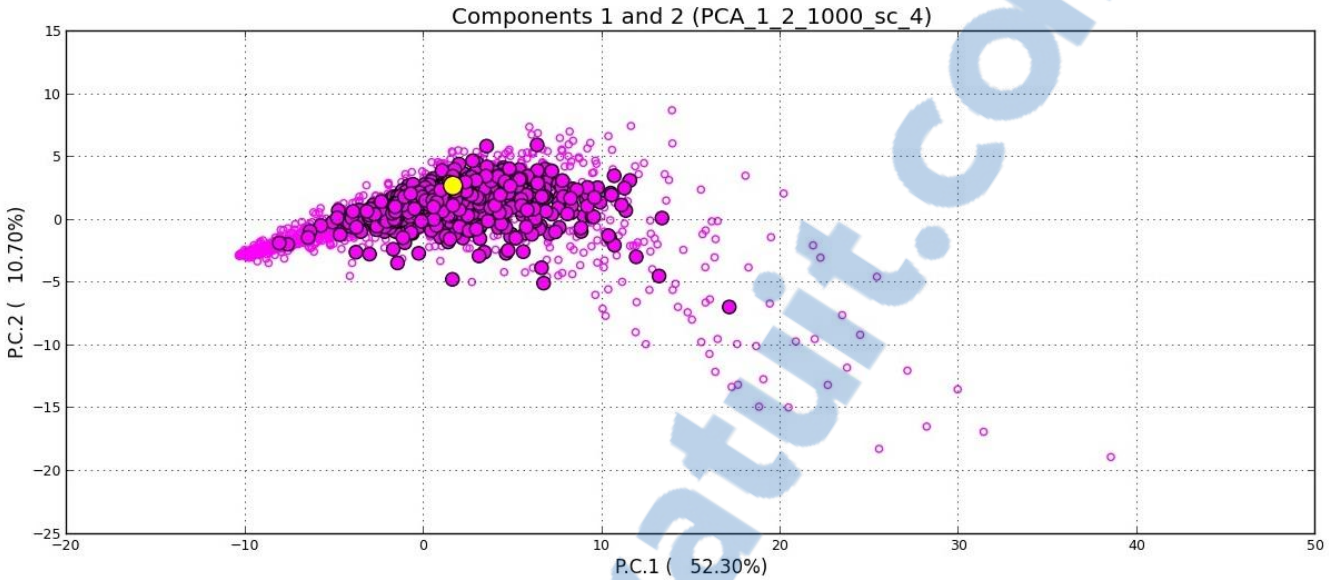
N_{mic} and N_{MtDNA} refer, respectively, to the number of individuals typed for microsatellite analyses and the number mitochondrial sequences obtained. In Pyrénées-Orientales, due to the very small number of individuals collected in each trap, the sample consists of a mix of samples collected in 2012 in different locations (Reynes, Maureillas, Ceret ; see Supplementary Table 1).

Table 2 | Pairwise F_{ST} values between *C. imicola* populations samples.

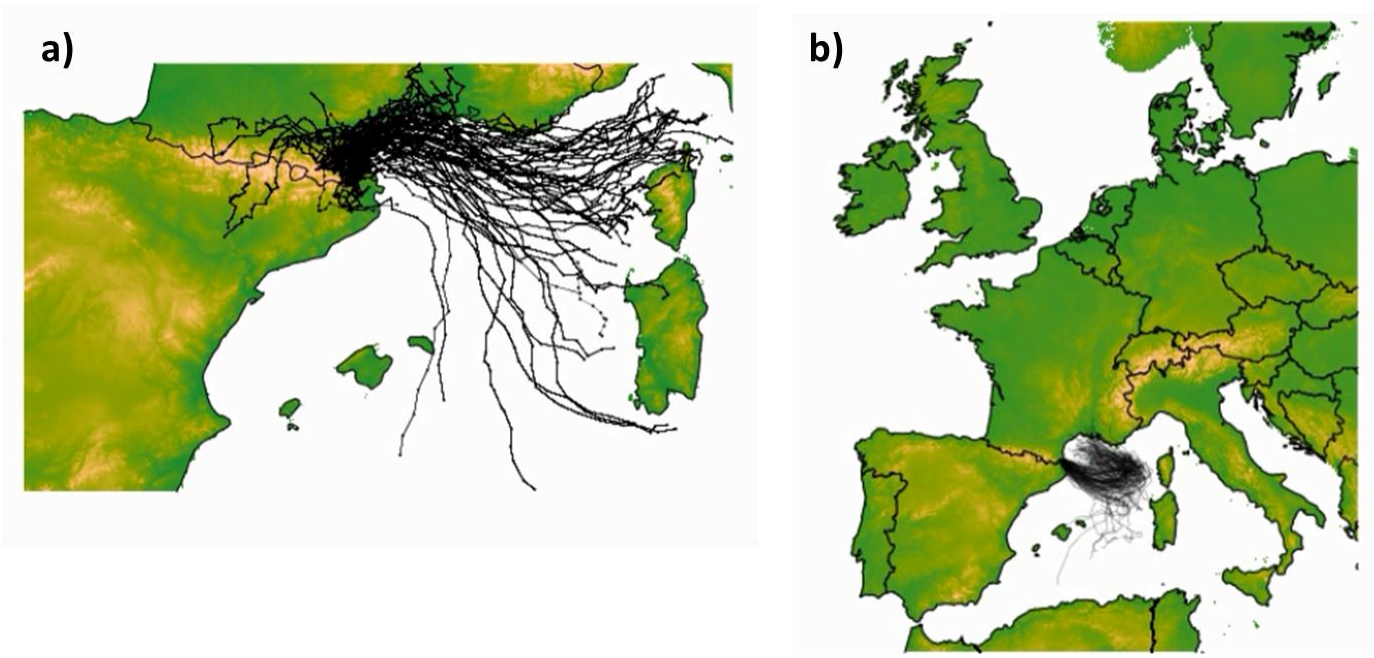
	Western cluster							Central cluster							
Locations	E-Girona	D-Toledo	C-Huelva	B-Castelo Branco	A-Beja	H-Khemisset	G-Sidi Yahia El Gharb	M-Pyrénées-Orientales	K-Roquebrune-sur-Argens	L-Bormes-les-Mimosas	N-Figari	O-Pietracorbara	P-San Giovanni Suergiu	J-Skikda	I-Wilaya de Jijel
F-Majorca	0.0095	0.0113	0.0086	0.0057	-0.0015	0.0057	-0.0017	**0.0242**	**0.0375**	**0.0320**	0.0126	0.0054	**0.0392**	0.0118	**0.0158**
E-Girona		-0.0033	**0.0177**	-0.0002	-0.0002	0.0089	0.0093	**0.0341**	**0.0620**	**0.0446**	**0.0326**	**0.0203**	**0.0352**	**0.0265**	**0.0308**
D-Toledo			**0.0233**	-0.0040	-0.0038	0.0051	0.0089	**0.0236**	**0.0570**	**0.0366**	**0.0276**	**0.0168**	**0.0336**	**0.0250**	**0.0340**
C-Huelva				0.0066	0.0127	**0.0243**	**0.0152**	**0.0295**	**0.0627**	**0.0643**	**0.0236**	**0.0228**	**0.0423**	**0.0244**	**0.0309**
P-Castelo Branco					-0.0012	0.0033	0.0023	**0.0185**	**0.0529**	**0.0430**	**0.0118**	**0.0104**	**0.0172**	0.0102	**0.0244**
P-Beja						0.0079	0.0056	0.0164	**0.0485**	**0.0342**	0.0143	**0.0082**	**0.0262**	0.0155	**0.0220**
H-Khemisset							-0.0074	**0.0142**	**0.0405**	**0.0322**	**0.0139**	**0.0145**	**0.0228**	0.0082	**0.0251**
G-Sidi Yahia El Gharb								0.0144	**0.0290**	**0.0250**	0.0087	0.0068	**0.0286**	0.0021	**0.0193**
M-Pyrénées-Orientales									**0.0231**	0.0237	-0.0003	0.0110	0.0149	0.0170	0.0097
K-Roquebrune-sur-Argens										0.0100	**0.0249**	**0.0171**	**0.0527**	**0.0342**	**0.0279**
L-Bormes-les-Mimosas											0.0247	**0.0210**	**0.0472**	**0.0305**	**0.0259**
N-Figari												-0.0045	0.0099	0.0043	0.0090
O-Pietracorbara													0.0219	0.0056	0.0064
P-San Giovanni Suergiu														0.0044	0.0109
J-Skikda															0.0002

F_{ST} values are grouped according to the genetic clusters inferred by STRUCTURE v.2.3.3.: western cluster (Spain, Portugal, Morocco) and central cluster (Algeria, Continental France, Corsica). The first letter in front of each location name refers to corresponding country: A – B, Portugal; C – E, Continental Spain; F, Balearic Islands; G – H, Morocco; I – J, Algeria; K – M, Continental France; N – O, Corsica; P, Sardinia. Significant values, at the adjusted nominal level (5%) for multiple comparison of 0.000476, are highlighted in bold.

Supplementary Information



Supplementary Figure 1 | Principal component analysis (PCA) in the space of summary statistics performed under the most probable scenario (Scenario 4). The observations are the simulated datasets and the variables are the summary statistics. Small blue circles correspond to simulated statistics with parameters drawn from the prior distributions, the large blue dots correspond to the simulated datasets drawn from the posteriors distributions and the yellow dot represents the real *C. imicola* data set.



Supplementary Figure 2 | NAME simulations of individual trajectories. Individual 36 hours back-trajectories for 2008-09-11, when particles mainly originated east from Pyrénées-Orientales (Fig. 2a). In this figure, points represent the position of each particle every hour. Full 36-hour back-trajectories for all particles together are also presented for each day during the full observation period (Fig. 2b).

Supplementary Table 1 | Details of the entomological surveys realized in Spain and in France from 2008 and 2012.

		2008					2009					2010					2011					2012				
				Maximum *C. imicola* collection					Maximum *C. imicola* collection					Maximum *C. imicola* collection					Maximum *C. imicola* collection					Maximum *C. imicola* collection		
code	Location	Nb night collection during the year	Total nb of *Culicoides* collected	Collection date (DD /MM/YY)	Nb of *C. imicola*	Nb of other *Culicoides* collected	Nb night collection (DD /MM/YY)	Total nb of *Culicoides* collected	Collection date (DD /MM/YY)	Nb of *C. imicola*	Nb of other *Culicoides* collected	Nb night collection during the year	Total nb of *Culicoides* collected	Collection date (DD /MM/YY)	Nb of *C. imicola*	Nb of other *Culicoides* collected	Nb night collection during the year	Total nb of *Culicoides* collected	Collection date (DD /MM/YY)	Nb of *C. imicola*	Nb of other *Culicoides* collected	Nb night collection during the year	Total nb of *Culicoides* collected	Collection date (DD /MM/YY)	Nb of *C. imicola*	Nb of other *Culicoides* collected
Monitoring sites																										
Spain																										
23	Aramunt	52	590	-	0	-	52	375	02/07/2009	2	5	52	7,594	-	0	-	52	6,801	25/08/2011	1	38	52	3,188	-	0	-
26	Caldes	52	30,905	07/08/2008	129	699	52	9,801	10/09/2009	1,464	37	52	6,786	16/09/2010	11,627	358	52	22,817	01/09/2011	6,344	194	52	1,705	11/10/2012	739	0
31	Garcia	52	297	-	0	-	52	163	03/09/2009	1	0	52	184	20/05/2010	1	18	52	843	-	0	-	52	236	-	0	-
29	Piera	52	1,303	11/09/2008	1	28	52	2,001	08/10/2009	4	340	52	50,026	09/09/2010	7	5,207	52	8,551	03/11/2011	1	214	26	729	-	0	-
25	Susqueda	52	24,702	-	0	-	52	17,904	16/07/2009	296	4,654	52	40,123	01/07/2010	283	4,250	52	26,218	26/05/2011	279	4,103	26	3,604	24/05/2012	2	420
24	Avia	52	30	-	0	-	52	729	-	0	-															
32	Montbrio	52	10	-	0	-	52	64	-	0	-	19	3	29/04/2010	1	2										
27	Alguaire	52	29	-	0	-	52	110	-	0	-															
33	Amposta	52	76	-	0	-	52	667	31/12/2009	1	16															
30	Bonastre	52	44	21/08/2008	4	4	52	63	15/10/2009	24	12															
28	Sant Iscle	52	25,763	-	0	-	52	2,391	15/10/2009	3	54															
22	Vilanova Muga	52	6,486	27/03/2008	1	13	52	7,887	-	0	-															

France																												
9	**St Jean Pla de Corts**	20	144	**07/10/2008**	**3**	20																						
2	Vebre						25	8,972	-	0	-	26	23,788		0		25	2,305		0		27	1,155		0			
1	Clermont						23	740	-	0	-	26	446		0		26	5,306		0		27	3,069		0			
5	Argeles Sur Mer						23	3,080	-	0	-																	
4	Argeles Sur Mer											25	92		0		24	81		0		21	62		0			
3	Toreilles	21	229	-	0	-																						
Sentinel sites																												
France																												
6	Vives						1	7	19/10/2009	0	7																	
10	Ceret						1	16	19/10/2009	0	16																	
18	**Reynes**						1	1	19/10/2009	0	1	1	408	11/10/2010	0	408	1	105	18/10/2011	0	105	1	466	**25/09/2012**	**8**	466		
15	**Ceret**											1	186	11/10/2010	0	186	1	194	18/10/2011	0	194	1	2	**26/09/2012**	**3**	2		
17	Amélie Les Bains Palalda						1	1	19/10/2009	0	1																	
7	Taillet						1	1	19/10/2009	0	1																	
21	**Maureillas Las Illas**											1	0	11/10/2010	0	0	1	109	18/10/2011	0	109	1	10	**25/09/2012**	**11**	10		
20	Maureillas Las Illas											1	6	11/10/2010	0	6	1	0	18/10/2011	0	0	1	20	25/09/2012	0	20		
14	**Albere**											1	6	11/10/2010	1	6	1	3	18/10/2011	0	3	1	1	**26/09/2012**	**2**	1		

19	**Cerbere**				1	0	21/10/2009	0	0	1	5	14/10/2010	0	5	1	3	19/10/2011	0	3	1	9	**24/09/2012**	**3**	9
16	Banyuls				1	0	19/10/2009	0	0	1	2	14/10/2010	0	2	1	2	19/10/2011	0	2	1	0	24/09/2012	0	0
13	Montesquieu Des Alberes				1	0	20/10/2009	0	0	1	9	13/10/2010	0	9	1	0	18/10/2011	0	0	1	0	26/09/2012	0	0
11	Laroque Des Alberes									1	1	13/10/2010	0	1	1	37	17/10/2011	0	37	1	7	24/09/2012	0	7
12	**Collioure**									1	1	14/10/2010	0	1	1	14	19/10/2011	0	14	1	103	**25/09/2012**	**1**	103
8	Sorede									1	41	13/10/2010	0	41	1	1	19/10/2011	0	1	1	53	24/09/2012	0	53

For *C. imicola* collections, we reported only the data for the maximum abundance collection date. Nb = number. Locations and collection dates in bold are positive sites for *C. imicola* in the Pyrénées-Orientales department.

Supplementary Table 2 | COI and Cytb concatenated genetic diversity and haplotype composition within sample site.

Site	N	H	Hd ± SD	π ± SD	Haplotype composition																														
					H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23	H24	H25	H26	H27	H28	H29	H30	H31
Algeria	17	7	0.86 ± 0.05	0.0015 ± 0.0002	5	3	1	3	1	2	2																									
Morocco	14	11	0.95 ± 0.04	0.0020 ± 0.0002		1					3	1	1						2							1	1	1	1	1	1				
Portugal	15	7	0.81 ± 0.09	0.0013 ± 0.0003						2	7		1		2																		1	1	1
Continental Spain	22	11	0.83 ± 0.07	0.0013 ± 0.0002		2					9		1	1	1	1	1	3	1	1	1														
Balearic Islands, Spain	8	3	0.67 ± 0.12	0.0007 ± 0.0001	4						3	1																							
Pyrénées-Orientales, France	17	6	0.75 ± 0.09	0.0010 ± 0.0002	2	3				1	8											2		1											
Var, France	16	3	0.54 ± 0.09	0.0056 ± 0.0001		10				5															1										
Corsica, France	16	6	0.85 ± 0.02	0.0013 ± 0.0001	1	5		3			5											1	1												
Sardinia, Italy	8	3	0.68 ± 0.12	0.0007 ± 0.0002				3		1	4																								

For each site, the number of samples (N), the number of haplotypes (H), haplotype (Hd) and nucleotide (π) diversity and their standard deviations (SD) are indicated.

Supplementary Table 3 | P-values inferred by the Wilcoxon test of bottleneck, F_{IS} values within samples and effective population sizes

Country	Sites	code	IAM	TPM	SMM	F_{IS}
Algeria	Skikda	J	0.0185*	0.5000	0.9355	0.107
	Wilaya de Jijel	I	0.0136*	0.3671	0.8984	0.050
Morocco	Khemisset	H	0.4101	0.9179	0.9951	0.030
	Sidi Yahia El Gharb	G	0.5898	0.9179	0.9970	0.140
Portugal	Castelo Branco	B	0.1250	0.5898	0.9179	0.096
	Beja	A	0.0644	0.5898	0.8984	0.122
Balearic Islands, Spain	Majorca	F	0.1015	0.5898	0.9814	0.097
Continental Spain	Girona	E	0.0644	0.5898	0.1503	0.080
	Toledo	D	0.0097**	0.4101	0.8984	0.010
	Huelva	C	0.0019**	0.0644	0.5449	0.025
Continental France	Pyrénées-Orientales	M	0.0068**	0.1503	0.3671	-0.015
	Roquebrune-sur-Argens	K	0.0019**	0.0097**	0.3261	-0.038
	Bormes-les-Mimosas	L	0.0097**	0.1503	0.4550	0.024
Corsica, France	Figari	N	0.0097**	0.2480	0.8496	0.036
	Pietracorbara	O	0.0019**	0.0644	0.8750	0.112
Sardinia, Italy	San Giovanni Suergiu	P	0.0185*	0.2481	0.7871	0.052

For the bottleneck analysis, p-values correspond to the Wilcoxon unilateral test under each mutation model (Infinite Allele Mutation: IAM, Two-phase model: TPM and Stepwise Mutation Model: SMM). Statistical significance is tested over 10000 replicates (**P < 0.01, *P < 0.05). P-values obtained for the Hardy-Weinberg test were adjusted for multiple comparisons at the nominal level (5%) of 0.00037.

Supplementary Table 4 | Mitochondrial pairwise F_{ST} values between *C. imicola* populations.

	Algeria	Sardinia, Italy	Corsica, France	Pyrénées-Orientales, France	Var, France	Morocco	Portugal	Balearics Island, Spain
Sardinia, Italy	0.03188							
Corsica, France	0.04440	0.00444						
Pyrénées-Orientales, France	0.04421	**0.19543**	0.04354					
Var, France	**0.15972**	**0.34694**	0.10000	0.03702				
Morocco	**0.07275**	**0.11014**	0.04071	0.00905	**0.03303**			
Portugal	0.06163	**0.07438**	**0.07590**	0.02891	**0.07054**	0.02240		
Balearics Island, Spain	0.02336	**0.28571**	**0.24953**	**0.19914**	**0.47986**	0.12475	**0.19556**	
Continental Spain	**0.12118**	**0.12695**	**0.07619**	**0.05231**	0.05430	0.01229	0.03447	**0.20952**

Significant values are highlighted in bold.

Supplementary Table 5 | Approximate Bayesian computation (ABC) results.

Description of tested scenarios	Posterior probability	95% credibility interval	Type I error	Type II error
Independent introductions from North Africa into Spain, Sardinia, Corsica and Pyrénées-Orientales	0.07	[0.01-0.13]	0.16	0.06
Introduction from North Africa into Spain, then colonization of Pyrénées-Orientales out from Spain	0.01	[0.00-0.06]	0.18	0.06
Introduction from North Africa into Sardinia, then colonization of Pyrénées-Orientales out from Sardinia	0.29	[0.26-0.32]	0.21	0.09
Introduction from North Africa into Sardinia, then introduction in Corsica out from Sardinia, and then colonization of Pyrénées-Orientales out from Corsica	**0.62**	**[0.60-0.64]**	**0.28**	**0.06**

Type 1 error is the probability of selecting another scenario when the chosen scenario is true. Type 2 error is the mean probability of selecting the chosen scenario when it is false. The selected (most probable) scenario is highlighted in bold (see also Fig. S7).

Supplementary Table 6 | ABC model checking of the most probable scenario (Scenario 4).

summary statistics	observed value	proportion (simulated<observed)	summary statistics	observed value	proportion (simulated<observed)	summary statistics	observed value	proportion (simulated<observed)
NAL_1_1	5.0000	0.6330	V2P_1_1&4	1.3863	0.3315	DAS_1_2&3	0.3945	0.3635
NAL_1_2	4.3333	0.5850	V2P_1_1&5	1.3480	0.3255	DAS_1_2&4	0.3845	0.3985
NAL_1_3	4.6667	0.5440	V2P_1_2&3	1.3258	0.3310	DAS_1_2&5	0.4048	0.3855
NAL_1_4	5.0000	0.5915	V2P_1_2&4	1.3718	0.3355	DAS_1_3&4	0.3984	0.4425
NAL_1_5	4.4444	0.5445	V2P_1_2&5	1.3245	0.3420	DAS_1_3&5	0.4049	0.4220
HET_1_1	0.6101	0.6745	V2P_1_3&4	1.3911	0.3290	DAS_1_4&5	0.3989	0.4565
HET_1_2	0.6113	0.7160	V2P_1_3&5	1.3621	0.3420	DM2_1_1&2	0.0504	0.0805
HET_1_3	0.5909	0.6200	V2P_1_4&5	1.4067	0.3375	DM2_1_1&3	0.0468	0.1135
HET_1_4	0.6074	0.6310	FST_1_1&2	0.0332	0.2965	DM2_1_1&4	0.0354	0.0825
HET_1_5	0.5939	0.6340	FST_1_1&3	0.0352	0.4550	DM2_1_1&5	0.0605	0.1440
VAR_1_1	1.3344	0.3385	FST_1_1&4	0.0261	0.4365	DM2_1_2&3	0.0616	0.2345
VAR_1_2	1.2770	0.3340	FST_1_1&5	0.0320	0.3400	DM2_1_2&4	0.0450	0.0860
VAR_1_3	1.3534	0.3460	FST_1_2&3	0.0146	0.3340	DM2_1_2&5	0.0639	0.3585
VAR_1_4	1.4478	0.3500	FST_1_2&4	0.0159	0.1545	DM2_1_3&4	0.0173	0.0165 (*)
VAR_1_5	1.3576	0.3595	FST_1_2&5	-0.0004	0.2160	DM2_1_3&5	0.0776	0.3590
MGW_1_1	0.9854	0.8330	FST_1_3&4	0.0042	0.0420 (*)	DM2_1_4&5	0.0579	0.1545
MGW_1_2	0.9000	0.6515	FST_1_3&5	0.0119	0.3580			
MGW_1_3	1.0286	0.9160	FST_1_4&5	0.0049	0.0590			
MGW_1_4	0.9712	0.7930	LIK_1_1&2	1.0667	0.5670			
MGW_1_5	0.9677	0.8430	LIK_1_1&3	1.0523	0.6180			
N2P_1_1&2	5.7778	0.6655	LIK_1_1&4	0.9948	0.6055			
N2P_1_1&3	5.7778	0.6080	LIK_1_1&5	1.0554	0.5920			
N2P_1_1&4	5.7778	0.5920	LIK_1_2&1	0.9423	0.4770			
N2P_1_1&5	5.6667	0.6085	LIK_1_2&3	0.8331	0.4720			
N2P_1_2&3	5.0000	0.5110	LIK_1_2&4	0.8503	0.3980			
N2P_1_2&4	5.3333	0.5390	LIK_1_2&5	0.8065	0.5155			
N2P_1_2&5	4.7778	0.5125	LIK_1_3&1	1.0029	0.5390			
N2P_1_3&4	5.5556	0.5470	LIK_1_3&2	0.9266	0.5165			

N2P_1_3&5	5.0000	0.4750	LIK_1_3&4	0.9104	0.4740
N2P_1_4&5	5.4444	0.5520	LIK_1_3&5	0.9307	0.5560
H2P_1_1&2	0.6210	0.6675	LIK_1_4&1	0.9566	0.5355
H2P_1_1&3	0.6120	0.6430	LIK_1_4&2	0.9286	0.4315
H2P_1_1&4	0.6183	0.6590	LIK_1_4&3	0.8742	0.4170
H2P_1_1&5	0.6134	0.6520	LIK_1_4&5	0.9132	0.4460
H2P_1_2&3	0.6038	0.6490	LIK_1_5&1	0.8989	0.4915
H2P_1_2&4	0.6145	0.6485	LIK_1_5&2	0.7852	0.5140
H2P_1_2&5	0.6022	0.6685	LIK_1_5&3	0.8042	0.4940
H2P_1_3&4	0.6010	0.5915	LIK_1_5&4	0.8158	0.4165
H2P_1_3&5	0.5966	0.6160	DAS_1_1&2	0.3730	0.3740
H2P_1_4&5	0.6033	0.5920	DAS_1_1&3	0.3790	0.3770
V2P_1_1&2	1.3115	0.3100	DAS_1_1&4	0.3770	0.3605
V2P_1_1&3	1.3455	0.3230	DAS_1_1&5	0.3826	0.4000

All the summary statistics (used and unused for model selection) were used. For each statistic, the observed value and probability (simulated < observed) are given. For a detail description of the summary statistics, see the DIYABC v.2.0.4 manual. Significant probabilities are represented with an asterisk.

Supplementary Table 7 | Primers characteristics.

Loci	Motif	Forward	Reverse	Allele size range	T (°C)
68	(GT)n	CTTTTCCGTTTCTTTTTATTTCTTT	GTTTCTTTCTGGTCGCGTTGGTTGCTG	99-105	60
12b	(CT)n	TTATGTGTGTATGTTAGCAAGGTCA	GTTTCTTCTTCGGATCAAAGAAATTTTGCC	129-139	50
3b	(AC)n	ATGCGGATGTTTGAAGTG	GTTTCTTTTTTGTGTCTTATTGCCC	154-175	50
31	(CAA)n	TTCTGTTCGGCTGTTGCGTT	GTTTCTTCTTTTTACGTGGTGGTCATTC	152-168	60
41b	(CT)n	GAGGAGGAGGTAGAA	GTTTCTTTCTATTAGTCAATGGTG	155-166	50
35t	(AC)n	TTTGTAAAAGCCAGTTCAACCG	GTTTCTTATCGAACGAAGGAAATAACCAC	171-194	60
88b	(AC)n	TTTGTTCGATTTGTAGTG	GTTTCTTCCTCTCTTTCATTCGC	243-256	50
16	(TG)n	TTGCCTTTGCTTGTTGAGGATG	GTTTCTTTCCTCTTTAAAATCACTGACGTG	292-299	60
88	(CAT)n	GTTGGTGCTTTGTTGTGTTGT	GTTTCTTTTTCTTTTTCTCCTTTTTGTTTCTTTC	344-348	50

Supplementary Table 8 | Prior distributions of demographic, historical and mutation parameters used for ABC inferences.

Parameter description	Parameter	Prior distribution
Stable effective population size	N	Loguniform [1000; 200000]
Time of colonization events in Pyrénées-Orientales	t_p	Uniform [40; 10000]
Time of colonization events in Sardinia, Corsica and Spain	t_s, t_c, t_e	Uniform [140; 10000]
Duration of bottleneck	D_B	Uniform [0; 50]
Founding effective population sizes	N_s	Loguniform [2; 100]
Mean mutation rate	µmicrosatellites	Loguniform [1E-6; 1E-4]
Mean parameter of the geometric distribution of length	P	Uniform [1E-1; 3E-1]
Mean mutation rate for nucleotide instability	SNI	Uniform [1E-8; 1E-4]

Effective population sizes (N) are expressed in number of diploid individuals and times of events (t) in number of generations going back to the past. Conditions among the parameters used during the simulations were ts ≥ tp, te ≥ tp, tc ≥ tp.

Spatio-temporal genetic variation of the biting midge vector species *Culicoides imicola* (Ceratopogonidae) Kieffer in France

Stéphanie Jacquet[a,e,f*], Karine Huber[b], Hélène Guis[a], Marie-Laure Setier-Rio[c], Maria Goffredo[d], Xavier Allène[a], Ignace Rakotoarivony[a], Christine Chevillon[e,f], Jérémy Bouyer[a], Thierry Baldet[a], Thomas Balenghien[a], Claire Garros[a]

a. Cirad, UMR15 CMAEE, 34398 Montpellier, France
b. INRA, UMR1309 CMAEE, 34398 Montpellier, France; Cirad, UMR15 CMAEE, 34398 Montpellier, France
c. EID Méditérranée, 34184 Montpellier, France
d. Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise 'G. Caporale', 64100 Teramo, Italy
e. CNRS, Université de Montpellier, UMR 5290 Maladies Infectieuses & Vecteurs-Ecologie, Génétique, Ecologie, Contrôle (MIVEGEC)
f. IRD, UR 224 MIVEGEC, BP 64501, Agropolis, 34 394 Montpellier cedex 5, 25 IRD, (INTERTRYP), UMR177 IRD-CIRAD, TA A-17/G, Campus International de Baillarguet, Montpellier, France

* Corresponding author

Cirad, UMR Contrôle des Maladies Animales Exotiques et Emergentes, Campus International de Baillarguet, TA-A15/G, 34398 Montpellier cedex 5, France, stephanie.jacquet@cirad.fr

Email adresses :

SJ: stephanie.jacquet@cirad.fr

KH: karine.huber@cirad.fr

HG: helene.guis@cirad.fr

MLSR: mlsetierrio@eid-med.org

MG: m.goffredo@izs.it

XA: xavier.allene@cirad.fr

IR: ignace.rakotoarivony@cirad.fr

CC: christine.chevillon@ird.fr

JB: jeremy.bouyer@cirad.fr

ThB: thierry.baldet@cirad.fr

TB: thomas.balenghien@cirad.fr

CG: claire.garros@cirad.fr

Abstract

Background Introduction of vector species into new areas represents a main driver for the emergence and worldwide spread of vector-borne diseases. This poses a substantial threat to livestock economies and public health. *Culicoides imicola* Kieffer, a major vector species of economically important animal viruses, is described with an apparent range expansion in Europe where it has been recorded in south-eastern continental France, its known northern distribution edge. This questioned on further *C. imicola* population extension and establishment into new territories. Studying the spatio-temporal genetic variation of expanding populations can provide valuable information for the design of reliable models of future spread.

Methods Entomological surveys and population genetics approaches were used to assess the spatio-temporal population dynamics of *C. imicola* in France. Entomological surveys (2-3 consecutive years) were used to evaluate population abundances and local spread in continental France (28 sites in the Var department) and in Corsica (4 sites). We also genotyped at nine microsatellite loci insects from 3 locations in the Var department over 3 years (2008, 2010 and 2012) and from 6 locations in Corsica over 4 years (2002, 2008, 2010 and 2012).

Results Entomological surveys confirmed the establishment of *C. imicola* populations in Var department, but indicated low abundances and no apparent expansion there within the studied period. Higher population abundances were recorded in Corsica. Our genetic data suggested the absence of spatio-temporal genetic changes within each region but a significant increase of the genetic differentiation between Corsican and Var populations through time. The lack of intra-region population structure may result from strong gene flow among populations. We discussed the observed temporal variation between Corsica and Var as being the result of genetic drift following introduction, and/or the genetic characteristics of populations at their range edge.

Conclusions Our results suggest that local range expansion of *C. imicola* in continental France may be slowed by the low population abundances and unsuitable climatic and environmental conditions.

Keywords: *Culicoides*, distribution range, spatio-temporal, population genetics, entomological survey, Mediterranean basin

Background

Range expansions are increasingly frequent in the history of many, if not most, species [1]. The growing rate of such phenomena and the associated consequences could be terribly alarming when concerning arthropod vector species. Introduction of vector species into new areas can lead to the emergence and spread of human and animal vector-borne pathogens, posing substantial threat to public health and livestock economy. Although most biological invasions are linked to anthropogenic influences such as intercontinental commerce and travel (e.g. ship, boat, aircraft, highways), species introductions can also occur by natural dispersal of the organisms or passive transport by winds [e.g. 2, 3, 4]. The dynamics of range expansion results from the interaction of ecological and evolutionary factors. These factors are most often analyzed at a spatial scale while the temporal scale is overlooked. Assessing the drivers involved in range expansion can help designing reliable predictive models of future spread. In particular, the study of spatio-temporal genetic variation of expanding populations can yield insights into their dispersion rates and patterns of spread, which in turn can prompt the processes underlying their establishment and persistence [5].

In this study, we investigated the spatio-temporal population dynamics of the biting midge *Culicoides imicola* Kieffer (Diptera: Ceratopogonidae) at its northern distributional edge in France. This species is known as the main afrotropical vector of Orbiviruses including bluetongue virus (BTV) and African horse sickness virus (AHSV) in Africa and Middle-East [6]. *Culicoides imicola* was described as expanding its range northward in the Mediterranean basin, however recent genetic studies suggest the species has been present and established for a long time in the Mediterranean region [7, 8]. Nevertheless, if the settlement of *C. imicola* populations in most parts of the Mediterranean region is an old story, extensive entomological surveillance performed in continental France since 2002 following BTV emergence in Corsica Island suggests a recent northward expansion of the species at the northern edge of its distribution.

Following the record of *C. imicola* in the island of Sardinia (Italy), the species was collected for the first time in the French island of Corsica in 2000 [9]. These findings were confirmed by subsequent extensive entomological surveys undertaken from 2002 onwards [10-12]. Populations are found to be widely distributed throughout Corsica notably on the littoral, with high abundances (reaching 12,000 captured individuals per night during the first observations) [10], suggesting that *C. imicola* is well established in the island. Given the short distance

between continental France and Corsica (180 km), a wide entomological surveillance network was implemented in 2002 in continental France along the Mediterranean coast to investigate the spread from Corsica. In 2003, the first *C. imicola* individuals were captured (two individuals) in south-eastern continental France, in Var and Alpes-Maritimes departments [10]. Despite subsequent additional surveys around the two positive sites, settled populations of *C. imicola* were not recorded. In 2004, five more individuals were collected in the Var department; the next years witnessed records of other settled populations of *C. imicola* [10]. The first collected individuals of *C. imicola* in the Var department coincided with a year of high abundance of the species in the north-eastern part of Corsica [10]. The authors hypothesized that Corsican emigrants of the species had most likely colonized the Var department through its high passive wind-mediated dispersal capacity. This was supported by genetic analyses based on microsatellite and mitochondrial markers which indicated strong genealogical relationships between populations of *C. imicola* from the Var department and Corsica [13].

To date, the species distribution has reached its most northern distribution edge in the south-east of the Var department. Since the first records, population abundance is low in this zone (maximum catch < 200 individuals collected per night) [10]. In 2008, a new record of *C. imicola* in south-western continental France, in the littoral of the Pyrénées-Orientales department, confirmed the species ability to continue expanding its range and colonize new habitats [13]. A combination of high population abundances and suitable wind dispersal capacity has been shown to contribute to the species successful range expansion [13]. Besides, an ecoclimatic niche model predicted new suitable habitats under climate change for *C. imicola* populations in Europe [14].

In this paper, we further investigated the processes underlying the range expansion of *C. imicola* at a local scale in France. We updated the distribution and abundance of *C. imicola* in Corsica and Var department. We described the fine scale spatio-temporal genetic structuring of this vector species' populations. Particularly, we focused on the levels of genetic diversity and differentiation between the range margin (Var department) and a more central distribution (Corsica and Sardinia) of *C. imicola*. The results were jointly interpreted to describe *C. imicola* population dynamics in these regions.

Materials and Methods

Entomological surveys

To carry on the evaluation of the settlement (presence/absence in sampled sites) and spatial expansion of *C. imicola* in Var and Alpes Maritimes departments, we continued the yearly surveillance network presented in [10]. Entomological surveys were carried out in twenty-eight and twenty-nine sites respectively in 2011 and 2012 (Figure 1, Table S2). Population abundance in the Var department was assessed for the period 2003-2014. Moreover, four sites were sampled in Corsica in 2010 and 2012 (Figure 3, Table S3) in order to survey population abundances. Most of these sites have been surveyed between 2003-2010 and 2002-2008 in the Var department and Corsica, respectively, in [10]. All sites in the Var department were sampled once a year during the end of summer/early autumn (September/October) to match the abundance peak of *C. imicola* [10]. In Corsica, sites were surveyed weekly (February-April, November-December) and monthly (January, May-October) with a single night collection.

For genetic analyses, *C. imicola* samples collected from six sites over four years in Corsica (2002, 2008, 2010 and 2012), and from three sites over four years in the Var department (2006, 2008, 2010 and 2012) were selected (Figure 4, Table 1). One location sampled in 2012 from Sardinia was also included (Table 1). Adult midges were caught using ultra-violet light-suction traps (Onderstepoort design) placed near animal shelters containing sheep, cattle or horses. Specimens were stored in 90% ethanol and *C. imicola* individuals were identified and sexed under a binocular microscope using morphological determination keys [9].

Microsatellite genotyping

Genomic DNA was extracted from each single midge 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 insect (~ 32 individuals per site) was genotyped at nine microsatellite markers previously developed for *C. imicola* [8] (Table S1) following the protocol described in [7].

Microsatellite analyses

Genetic diversity and equilibrium testing

Linkage disequilibrium between all pairs of loci and deviations from Hardy-Weinberg, i.e. significant deviation of the inbreeding coefficient F_{IS} from zero, were tested with FSTAT

v2.9.3.2 [15]. The significance of F_{IS} was assessed by randomizing alleles among individuals within samples (10,000 permutations). Level of genetic diversity within samples per year was quantified by computing the allelic richness (A_R) and the mean observed heterozygosity (H_O) [16] with FSTAT v2.9.3.2 [15]. We used a non-parametric Mann-Whitney-Wilcoxon statistical test in R software v.3.1.2 [17] to check for any differences in levels of allelic richness between Corsican and Var populations.

To test whether populations are in mutation-equilibrium, we performed tests of heterozygosity excess [18] implemented in BOTTLENECK v.1.2.02 [19]. It has been shown that past bottleneck events will be detected with a high degree of sensitivity using the Infinite Allele Mutation (IAM) model, moderately with the Two-phase Model (TMM) and dimly with the Stepwise Mutation Model (SMM) [18]. Heterozygosity excess was tested under all three mutation models. For the TPM model the proportion of SMM was set to 70% and the variance to 30 (default values). The significance was assessed by performing 10,000 replicates. The Wilcoxon's signed-rank statistics were used to evaluate any deviation of the observed heterozygosity from the expectation under mutation-drift equilibrium.

Population structure

Genetic relationships between samples were visualized by a neighbor-joining tree (NJ) tree [20] based on the pairwise genetic distances of Cavalli-Sforza and Edwards [21] using the software POPULATIONS v.1.2.30 (http://bioinformatics.org/~tryphon/populations/). The robustness of the tree topology was evaluated by bootstrapping 10,000 times over loci.

Genetic variation and population structuring was also investigated by undertaking a Principal Component Analysis (PCA) using PCA-GEN v.1.2.1 software [22]. This analysis correlates genotypes and allele frequencies among individuals, without any assumption of equilibrium within populations, to define variables (components) that can characterize the neutral genetic variation among populations. The statistical significance associated with each axis was evaluated after 10 000 randomizations.

The genetic structure of *C. imicola* was then assessed using the Bayesian clustering method implemented in STRUCTURE v.2.3.3 software [23] which ascertains population membership of individuals according to their genotypes. We assumed an admixture model and correlated allele frequencies and used the sampling location as prior information. Each run consisted of a burn-in of 10^5 Markov chain Monte Carlo (MCMC) iterations, followed by 10^6 iterations. Ten replicates were carried out for each number K of genetic clusters set between 1 and 9. The accurate number of clusters was inferred with the ΔK method [24]. The relative importance of

the genetic clusters previously inferred by STRUCTURE and the population differentiation was assessed with the multi-locus hierarchical F-statistics. We tested if the genetic patterns were explained by the geography and/or the collection dates by grouping hierarchically the samples according their origin (Corsica *vs* Var department), then their collection year. We then performed the test separately within Corsica and Var department: samples were grouped hierarchically according to their collection year to assess any genetic significant differentiation during the time studied. This analysis was performed with Hierfstat package [25]. These tests were based on 10 000 permutations of either *Culicoides* genotypes among populations, within groups (i.e. collection dates) and within clusters (i.e. Corsica and Var department; H0: 'Fpopulations-cluster = 0'), or populations among clusters (H0: 'Fclusters-total = 0').

The software FSTAT v2.9.3.2 was used for assessing the level of genetic differentiation among populations through the Weir and Cockerham [26]'s unbiased estimates of F_{ST}.

These analyses were performed per year over the total sampling set available (encompassing thus Corsica and Var). A significant deviation of F_{ST} from zero was tested using the exact G test over 10,000 permutations of genotypes among samples. Global F_{ST} value was estimated over all sampled sites (Corsica and Var) in either 2008 or 2012 (i.e., two years when the sampling sets included the exact same localities). A statistical non-parametric Wilcoxon signed test with continuity correction implemented in R software v.3.1.2 [17] was used to test whether the genetic differentiation between Corsican and Var populations has significantly increased over time. Specifically we tested if the genetic differentiation (based on F_{ST} values) in 2012 was significantly higher than in 2008.

Isolation by distance

To test whether patterns of neutral genetic structure of *C. imicola* Corsican and Var populations were related to geographic distance, we performed partial Mantel tests implemented in the online program GENEPOP v.4.2 [27, 28]. These tests were conducted separately for the years 2008 and 2012. The significance of the regression between the logarithmic geographic distances and the pairwise values of $F_{ST}/ (1- F_{ST})$ was assessed with 100,000 permutations.

Results

Entomological surveys

A total of twenty-eight and twenty-nine sites were surveyed in 2011 and 2012 respectively (Table S2). In 2011, three out of the twenty-eight sites studied were previously recorded positive in [10] and were positive for a second consecutive year (Figure 1). In 2012, eight new positive sites were recorded (Figure 1). The compiled maximum catch per year of *C. imicola* between 2003 and 2014 showed low abundances in the Var department since the maximum did not exceed 4,500 individuals per site and night. In Corsica, *C. imicola* population abundances were always high in the most southern site, as it was recorded before in [10], reaching a maximum of about 160,000 *C. imicola* caught per night in 2010. The three other sites exhibited relatively low abundance, never exceeding 2,500 individuals caught per night for the three studied years (Figure 3).

For genetic analyses a total of 809 individuals from six sites sampled over four years (2002, 2008, 2010 and 2012) in Corsica and from three sites collected over four years (2006, 2008, 2010 and 2012) in the Var department were genotyped and analysed (27-32 individuals per site; Table 1).

Microsatellite analyses

Within population genetic diversity through time

Pairwise allelic tests of linkage disequilibrium indicated that loci were physically unlinked and statistically independent within populations (p-value > 0.05 after Bonferroni correction). Fisher's exact test revealed that genotypic frequencies were in accordance with Hardy-Weinberg equilibrium for all populations (p-value > 0.05 after Bonferroni correction), F_{IS} values ranged from -0.075 to 0.140 (Table 2).

Levels of genetic diversity remained stable over populations and over years (Table 2). In Corsica, the allelic richness (A_R) varied from 4.188 to 4.725 and the mean heterozygosity (H_O) ranged from 0.523 to 0.616. Both measures were comparable in Sardinia with an allelic richness equal to 4.512 and a mean heterozygosity of 0.560. In the Var department, the allelic richness extended from 3.481 to 4.214 and the mean observed heterozygosity varied from 0.515 to 0.609. Interestingly, the allelic richness was statistically lower in the Var department than in Corsica ($p < 0.0001$).

Spatio-temporal genetic changes in population structure

A spatial genetic structure was detected between, on the one hand, the group formed of Corsican and Sardinian population, and on the other hand, the continental populations from the Var department. The Bayesian clustering analysis defined these two geographical groups (ΔK maximum for K = 2) (Figure 5a). The same pattern was also evidenced in the neighbor-joining tree and the Principal Component Analysis (PCA) which grouped the sampled sites according to their geographical origin, without any effect of the collection year (Figure 5b, 5c).The PCA showed a significant overall F_{ST} value of 0.034, and indicated that the first two axes were significant and explained 55.81% of the total inertia: axis 1 accounted for 45.10% of the variance (F_{ST} = 0.016, p-value = 0) and axis 2 accounted for 10.71% of the variance (F_{ST} = 0.004, p-value = 0).

Considering the hierarchy of sampling, significant differentiation was detected for all levels: between the two genetic clusters Corsica and Var (Fclusters-total = 0.0021; P = 0.0085), within clusters (i.e. between collection years; Fyears-clusters = 0.0024; P = 0.0001) and within group (i.e. between samples collected at the same year; Fpopulations-years = 0.0247; P = 0.0022). When performing the analysis within Corsica, the collection year did not significantly account for the observed pattern (i.e. no significant difference between collection year; Fyears-total = 0.0032; P = 0.9755), however significant differentiation was detected among samples for each year (Fpopulations-years = 0.0343; P = 0.0002). The same results were obtained for Var department (Fyears-total = 0.0079; P = 0.6877 and Fpopulations-years = 0.0071; P = 0.0021).

These results are globally consistent with the genetic differentiation tests, which indicated significant genetic differentiation between island and continental populations (Table S4). Such differentiation is likely to have increased over time as indicated by the F_{ST} values which are twice as high in 2012 compared to 2008 (Table 3) and the significant Wilcoxon signed rank test (p = 0.0122). This was also evident in the estimates of global F_{ST} over all populations per year: values were higher in 2012 (F_{ST} = 0.032) than 2008 (F_{ST} = 0.014). Interestingly, populations from Sardinia were significantly differentiated to those from the Var department but were not differentiated to the Corsican populations (Table S4).

Assessing the levels of genetic differentiation over all collected years within regions indicated that sampled populations within Corsica and the Var department were not markedly structured. Indeed, the computed F_{ST} values were relatively low, ranging respectively from -0.0104 to 0.0290 in Corsica and -0.0029 to 0.0261, in the Var department (Table S4). None of the pairwise comparisons was significant within Corsica (p-value > 0.0042), and only four out of the thirty-six comparisons within the Var department were significant (p-value < 0.0014,

after Bonferonni correction). These results did not support any evident changes in genetic differentiation within each region during the time period studied.

Equilibrium testing

While tests based on the IAM mutation model suggested potential signatures of past genetic bottlenecks in nearly all sampled sites over years, those based on the most realistic TPM and SMM mutation models were only significant for one Corsican site sampled in 2010 (Figari) and two sites in the Var department, including Roquebrune-sur-Argens (V1, over the four collection years and V2 in 2010) (Table 2).

Isolation by distance

Patterns of isolation by distance among *C. imicola* sampled populations were observed both in 2008 and in 2012 (p-value = 0.00608 and 0.00719, respectively; Figure 6).

Discussion

While a recent study claims the historical presence of *C. imicola* in the Mediterranean region [7], continuous northward expansion of this vector species has been predicted as a result of projected climate change in the future [14]. Particularly, introduction of the species in the Var department in continental France, through wind-mediated dispersal from Corsica, highlights the potential of *C. imicola* to expand its range and colonize new habitats.

Assessing the temporal variation in population structure provides insights into population dynamics and information on the major forces driving genetic changes at a short temporal scale. In this study, we investigated the spatio-temporal patterns of genetic variation of the vector species, *C. imicola* in the Var department and in Corsica. Using nine highly polymorphic markers, two main conclusions can be drawn from our results: (i) the genetic composition and diversity within region did not significantly change in the time period studied but (ii) the genetic differentiation among Var and Corsican populations has significantly increased through time.

Our results suggest that the genetic diversity as well as the genetic structure within Corsica and the Var department have remained relatively stable during the time period studied. Indeed, the principal component analysis indicated that virtually none of the variation could be attributed to temporal samples within region, but was explained by spatial differentiation between regions. The observed low population structure within region is an expected response to considerable gene flow between populations. Species from the genus of *Culicoides* are

described as weak active flyers because of their small size. The daily flight distance for European *Culicoides* species in the Palearctic region has been estimated to be up to 2.21 km [29, 30], but is still unknown for *C. imicola*. In the Var department, the average spread of *C. imicola* has been estimated to 14.5km/year, suggesting a limited local expansion of the species [10]. The limited density of *C. imicola*, the presence of physical barriers and potential unsuitable environmental conditions probably impair its ability to successfully disperse. In contrast, our findings indicate that *C. imicola* may passively and/or actively disperse at the local scale, allowing relatively high gene flow among populations. It is worth noting that the number of sampled sites in the Var department was small and that the sites were geographically close to each other, and thus may not allow a proper uncover of the genetic structuring. However, despite the hilly topography of Corsica and the geographical distances between the sampled sites, low genetic differentiation was also observed among Corsican populations supporting the existence of gene flow among them. Nevertheless, it is possible that our set of markers have failed to assess the population structure at this scale. It would thus be interesting to further investigate the fine-scale spatio-temporal population variation with more markers or by performing genomic analysis based on highly polymorphic markers such as single nucleotide polymorphisms (SNPs). This would allow assessing the impact of landscape on population structuring. It would also be interesting to estimate the spread of *C. imicola* Corsican populations which are more abundant, using direct methods in order to evaluate the dispersal ability of the species in the landscape.

A significant increase in the genetic differentiation between Var and Corsican populations was observed through time. This genetic pattern could be explained by two hypotheses. First, it could reflect the genetic changes in population structuring due to founder effects following introduction. Indeed, incursion of species in novel environments is often associated with a loss of genetic diversity when the gene pool in the new habitat is provided by a small number of founding individuals [31]. The newly founded population may then experience strong genetic drift resulting in genetic differentiation among populations [31]. Our results support this scenario since signatures of demographic bottlenecks were detected in two Var populations and one Corsican population. In addition, the levels of allelic richness were significantly lower in the Var department while the mean heterozygosity was comparable among all populations. These results are consistent with the loss of genetic diversity associated with founder effects, as allelic richness is expected to be more sensitive to the effects of bottlenecks than is heterozygosity [31-33]. Thus, a limited number of emigrants

midges carried by the wind from Corsica Island into the Var department could have been the target of genetic drift, leading to the observed genetic differentiation.

Complementarily, the observed patterns could result from the processes operating at the distributional edge of *C. imicola* populations. Indeed, it is widely appreciated that towards the range edge, habitats and ecological factors can influence the demography of populations, leading to habitat fragmentation and low population abundance and density [1, 34]. As a result, genetic diversity is likely to decrease while genetic differentiation is expected to increase in marginal populations compared with central populations [34]. Such patterns have been reported for many taxa including plants [e.g. 35, 36], insects [e.g. 37] and reptiles [e.g. 38]. Despite regular entomological surveys, *C. imicola* populations have not been found in neighboring areas of the south-east coastal region of the Var department, its present consensually admitted northern edge [10]. In addition, levels of *C. imicola* abundance in this region have remained very low since its first records (maximum of 4,500 individuals caught per night over 2003-2014), compared with other reported local abundance such as in southern-Spain (20,000-43,000 maximum catch/night) [39], Corsica (maximum catch of ~ 160,000 in 2010) [10], and Sardinia (10,000-65,000 maximum catch/night) [40]. Interestingly, lower levels of *C. imicola* abundance have also been reported in northern Spain, i.e. Catalonia, Basque country (< 1,000 maximum catch/night) [41, 42] and northern continental Italy, i.e. north of Tuscany (< 100 maximum catch/night) [40], marking respectively the north-western and north-eastern limits of the species distribution in the western Mediterranean area (the population recorded in south-western Continental France in the Pyrénées-Orientales department is still in investigation to assess the establishment the population [13]).The patchy distribution and low population abundance of *C. imicola* populations are most likely driven by local habitat conditions including climate, topography, soil (type and moisture) and host availability [6, 39, 43], of which the species is highly dependent. Thus, environmental conditions in the Var department may be less suitable for *C. imicola* than in southern Corsica for example. As a consequence, the resulting low population size in the Var department may be more subject to demographic stochastic events and genetic drift. This may explain the observed genetic differentiation between Corsican and Var populations; however we would also expect higher genetic differentiation between Var populations compared with Corsican populations. Yet, our results indicate low levels of genetic differentiation within Var. As discussed, our sampling may not allow a proper uncover of the spatial genetic structure within the Var department. Alternatively, these low levels of differentiation could simply result from

a recent introduction of *C. imicola* there; if the incursion of the species has occurred in 2004, populations may not have enough time for a significant differentiation.

Interestingly, Corsican and Sardinian populations were genetically similar. The high capacity of *C. imicola* to passively disperse by winds may allow high gene flow among these two nearby islands, as it has been evidenced on longer distances in the Mediterranean region [13]. In contrast, the increase of genetic differentiation between Var and Corsican populations could indicate that despite the high dispersal ability of the species, migrations between Corsica and the Var department, if there are at all, may not be sufficient to homogenize the genetic composition of the continental populations. Nonetheless, these findings should be further investigated and confirmed by increasing the time period studied.

Conclusion

Our study highlights the processes underlying contemporary range expansions and driving population dynamics at a local scale. Our results suggest that local range expansion of *C. imicola* in continental France may be slowed by low population abundances and unsuitable environmental conditions. Despite the high ability of *C. imicola* to passively disperse through winds, our results indicate that the presence of large water bodies may restrict this process allowing the genetic differentiation between Corsican and Var populations. Further analyses based on a wider temporal scale would help for a better understanding of the mechanisms shaping *C. imicola* distribution and therefore assessing local *C. imicola*-borne disease epidemiology. Nevertheless, our findings yield information for the design of predictive models of future spread.

Competing interests

The authors declare that they have no competing interests

Author's contributions

SJ, KH, CG designed the study. SJ genotyped the samples. SJ, KH, and CG analyzed the data. MLSR, MG, XA, IR, ThB collected *C. imicola* samples. CC, JB, ThB, TB and HG, contributed to the manuscript firstly written by SJ, KH and CG. All authors read and commented the final manuscript version.

Acknowledgements

This work was granted by the Chercheur d'Avenir 2011 call from the Languedoc-Roussillon region, France and the Ministère de l'Agriculture.

References

1. Excoffier L, Foll M, Petit RJ: **Genetic consequences of range expansions**. *Annual Review of Ecology, Evolution, and Systematics* 2009, **40**:481-501.

2. Tatem AJ, Hay SI, Rogers DJ: **Global traffic and disease vector dispersal**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(16):6242-6247.

3. Harrington LC, Scott TW, Lerdthusnee K, Coleman RC, Costero A, Clark GG, Jones JJ, Kitthawee S, Kittayapong P, Sithiprasasna R: **Dispersal of the dengue vector *Aedes aegypti* within and between rural communities**. *The American Journal of Tropical Medicine and Hygiene* 2005, **72**(2):209-220.

4. Bishop AL, Spohr LJ, Barchia IM: **Effects of altitude, distance and waves of movement on the dispersal in Australia of the arbovirus vector, *Culicoides brevitarsis* Kieffer (Diptera: Ceratopogonidae)**. *Preventive Veterinary Medicine* 2004, **65**(3-4):135-145.

5. Peterson AT, Vieglais DA: **Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem A new approach to ecological niche modeling, based on new tools drawn from biodiversity informatics, is applied to the challenge of predicting potenti0al species' invasions**. *BioScience* 2001, **51**(5):363-371.

6. Mellor PS, Boorman J, Baylis M: ***Culicoides* biting midges: their role as arbovirus vectors**. *Annual review of entomology* 2000, **45**:307-340.

7. Jacquet S, Garros C, Lombaert E, Walton C, Restrepo J, Allene X, Baldet T, Cetre-Sossah C, Chaskopoulou A *et al* : **Colonization of the Mediterranean Basin by the vector biting midge species *Culicoides imicola*: an old story**. *Molecular Ecology* 2015, DOI: 10.1111/mec.13422.

8. Mardulyn P, Goffredo M, Conte A, Hendrickx G, Meiswinkel R, Balenghien T, Sghaier S, Lohr Y, Gilbert M: **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**. *Molecular Ecology* 2013, **22**(9):2456-2466.

9. Delécolle J-C, De La Rocque S: **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* 2002, **107**(4):371-379.

10. Venail R, Balenghien T, Guis H, Tran A, Setier-Rio ML, Delécolle JC, Mathieu B, Cetre-Sossah C, Martinez D, Languille J *et al*: **Assessing diversity and abundance of vector populations at a national scale: example of *Culicoides* surveillance in France after bluetongue virus emergence**. In: *Arthropods as vectors of emerging diseases.* Edited by Mehlhorn H. Berlin Springer-Verlag; 2012, 77-103.

11. Balenghien T, Delécolle J-C, Setier-Rio M-L, Rakotoarivony I, Allène X, Venail R, Delécolle D, Lhoir J, Mathieu B, Chavernac D: **Vecteurs du virus de la fièvre catarrhale ovine: suivi des populations de culicoides en 2011 en France**. *Bulletin Epidémiologique Santé Animale Alimentation* 2012, **54**:35-40.

12. Balenghien T, Rakotoarivony I, Allène X, Perrin JB, Garros C : **L'activité des populations de *Culicoides* en Corse en 2013**. *Santé animale-alimentation* 2014, 41.

13. Jacquet S, Huber K, Pagès N, Talavera S, Burgin LE, Carpenter S, Sanders C, Dicko AH, Djerbal M, Lhor Y *et al*. **Range expansion of the Bluetongue vector *Culicoides imicola* in continental France thanks to meteorological events** (submitted).

14. Guichard S, Guis H, Tran A, Garros C, Balenghien T, Kriticos DJ: **Worldwide niche and future potential distribution of Culicoides imicola, a major vector of bluetongue and African horse sickness viruses.** *PloS one* 2014, **9**(11):e112491.

15. Goudet J: **FSTAT (Version 1.2): A computer program to calculate F-statistics**. *Journal of Heredity* 1995, **86**(6):485-486.

16. Nei M: **Molecular evolutionary genetics**. New York: Columbia University Press; 1987.

17. R Development Core Team: **R: A language and environment for statistical computing.** *R Foundation for Statistical Computing, Vienna, Austria*; 2013.

18. Cornuet JM, Luikart G: **Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data**. *Genetics* 1996, **144**(4):2001-2014.

19. Piry S, Luikart G, Cornuet JM: **BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data**. *Journal of Heredity* 1999, **90**(4):502-503.

20. Takezaki N, Nei M: **Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA**. *Genetics* 1996, **144**(1):389-399.

21. Cavalli-Sforza LL, Edwards AW: **Phylogenetic analysis. Models and estimation procedures**. *American journal of human genetics* 1967, **19**(3 Pt 1):233-257.

22. Goudet J: **PCA-GEN (Version 1.2) a Program to Perform Principal Component Analysis (PCA) on Gene Frequency Data**. *Institute of Ecology, University of Lausanne, Switzerland* 1999.

23. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data**. *Genetics* 2000, **155**(2):945-959.

24. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study**. *Molecular Ecology* 2005, **14**(8):2611-2620.

25. Goudet J: **Hierfstat, a package for R to compute and test hierarchical F-statistics**. *Molecular Ecology Notes* 2005, **5**(1):184-186.

26. Weir BS, Cockerham CC: **Estimating F-Statistics for the Analysis of Population-Structure**. *Evolution* 1984, **38**(6):1358-1370.

27. Raymond M, Rousset F: **An exact test for population differentiation**. *Evolution* 1995, **49**(6)1280-1283.

28. Rousset F: **genepop'007: a complete re-implementation of the genepop software for Windows and Linux**. *Molecular Ecology Resourcesces* 2008, **8**(1):103-106.

29. Kluiters G, Swales H, Baylis M: **Local dispersal of Palaearctic *Culicoides* biting midges estimated by mark-release-recapture**. *Parasites & Vectors* 2015, **8**:86.

30. Kirkeby C, Bodker R, Stockmarr A, Lind P, Heegaard PM: **Quantifying dispersal of european *Culicoides* (Diptera: Ceratopogonidae) vectors between farms using a novel mark-release-recapture technique**. *PloS one* 2013, **8**(4):e61269.

31. Nei M, Maruyama T, Chakraborty R: **The bottleneck effect and genetic variability in populations**. *Evolution* 1975, **29**(1):1-10.

32. Leberg PL: **Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis**. *Evolution* 1992, **46**(2):477-494.

33. Spencer C, Neigel J, Leberg P: **Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks**. *Molecular Ecology* 2000, **9**(10):1517-1528.

34. Sexton JP, McIntyre PJ, Angert AL, Rice KJ: **Evolution and ecology of species range limits**. *Ecology Evolution and Systematics* 2009, **40**:415-436.

35. Diekmann OE, Serrao EA: **Range-edge genetic diversity: locally poor extant southern patches maintain a regionally diverse hotspot in the seagrass *Zostera marina***. *Molecular Ecology* 2012, **21**(7):1647-1657.

36. Faugeron S, Martínez EA, Correa JA, Cardenas L, Destombe C, Valero M: **Reduced genetic diversity and increased population differentiation in peripheral and overharvested populations of *Gigartina skottsbergii* (Rhodophyta, Gigartinales) in southern Chile**. *Journal of Phycology* 2004, **40**(3):454-462.

37. Krauss J, Schmitt T, Seitz A, Steffan-Dewenter I, Tscharntke T: **Effects of habitat fragmentation on the genetic structure of the monophagous butterfly *Polyommatus coridon* along its northern range margin**. *Molecular Ecology* 2004, **13**(2):311-320.

38. Böhme MU, Schneeweiß N, Fritz U, Schlegel M, Berendonk TU: **Small edge populations at risk: genetic diversity of the green lizard (*Lacerta viridis viridis*) in Germany and implications for conservation management**. *Conservation Genetics* 2007, **8**(3):555-563.

39. Acevedo P, Ruiz-Fons F, Estrada R, Marquez AL, Miranda MA, Gortazar C, Lucientes J: **A broad assessment of factors determining *Culicoides imicola* abundance: modelling the present and forecasting its future in climate change scenarios**. *PloS one* 2010, **5**(12):e14236.

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

41. Goldarazena A, Romón P, Aduriz G, Balenghien T, Baldet T, Delécolle JC: **First record of *Culicoides imicola*, the main vector of bluetongue virus in Europe, in the Basque Country (northern Spain)**. *Veterinary Record* 2008, **162**(25):820-821.

42. Sarto i Monteys V, Ventura D, Pagès N, Aranda C, Escosa R: **Expansion of *Culicoides imicola*, the main bluetongue virus vector in Europe, into Catalonia, Spain**. *Vet Rec* 2005, **156**:415-417.

43. Conte A, Ippoliti C, Savini L, Goffredo M, Meiswinkel R: **Novel environmental factors influencing the distribution and abundance of *Culicoides imicola* and the Obsoletus Complex in Italy**. *Veterinaria Italiana* 2007, **43** (3): 571 – 580.

44. Medlock JM, Hansford KM, Schaffner F, Versteirt V, Hendrickx G, Zeller H, Bortel WV: **A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options**. *Vector-borne and zoonotic diseases* 2012, **12**(6):435-447.

Figures and Tables

Figure 1 Distribution area of *C. imicola* in the Var department in 2011 and 2012. Presence/absence maps from 2003 to 2010 were previously published in [10].

Figure 2 Maximum catch per year of *C. imicola* in the Var department from 2003 to 2014.

Figure 3 Collection sites in Corsica with maximum catches of *Culicoides imicola* per year in 2010 and 2012. Population abundance maps from 2003 to 2009 were previously published in [10].

Figure 4 Sampled sites of *C. imicola* used for population genetics analyses.

Figure 5 Population genetic structure results of *Culicoides imicola*. (a) Genetic clustering of *C. imicola* samples. Each vertical line represents an individual and each color represents a cluster. (b) Microsatellite neighbor-joining tree based on genetic distance of Cavalli-Sforza & Edwards (1967). Bootstrap values are calculated over 1,000 replicates (only values > 50% are shown). (c) Principal Component Analysis based on microsatellite allelic frequencies.

Figure 6 Results of the Mantel tests for isolation by distance (IBD).

Table 1 Geographical location of *C. imicola* sampled sites for the population genetics study.

Table 2 Genetic diversity and Bottleneck results based on microsatellite data for each sampled site. The allelic richness (A_R), genetic diversity (H_S), observed (H_O) and expected (H_E) heterozygosity and F_{IS} are presented for each population. Results of bottleneck tests are presented for the Infinite Allele Model (IAM), Two-phase Model (TPM) and Stepwise Mutation Model (SMM).

Table 3 Pairwise F_{ST} values between Corsican and Var *C. imicola* population samples.

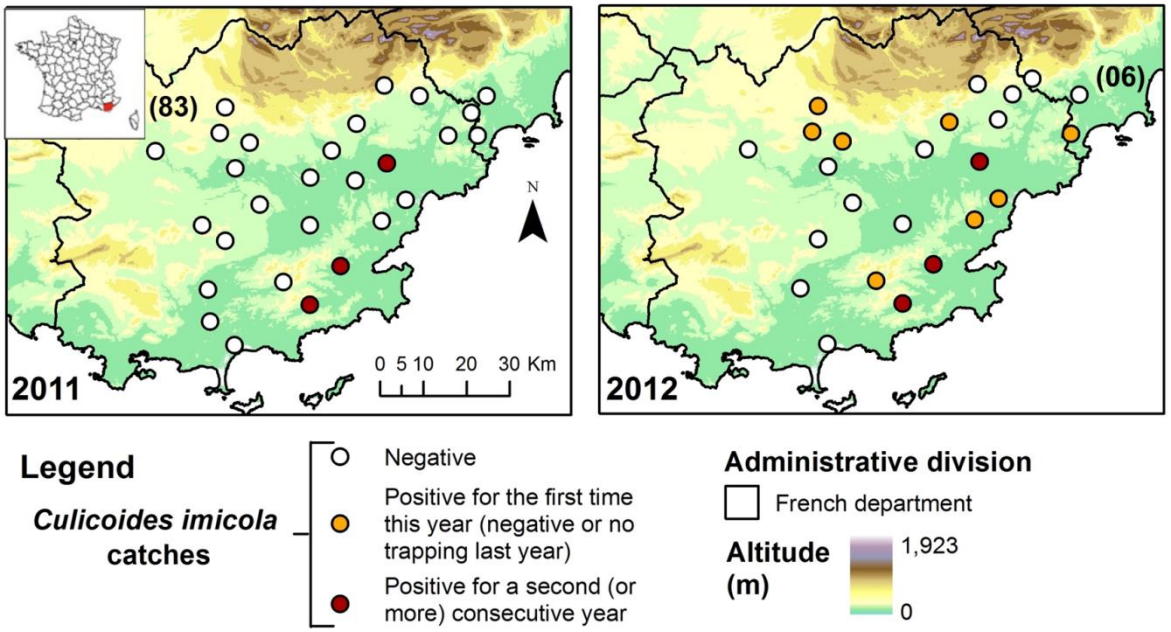


Figure 1 Distribution area of *C. imicola* in the Var department (83) and in the Alpes-Maritimes department (06) from 2011 and 2012. Presence/absence maps from 2003 to 2010 were previously published in [10].

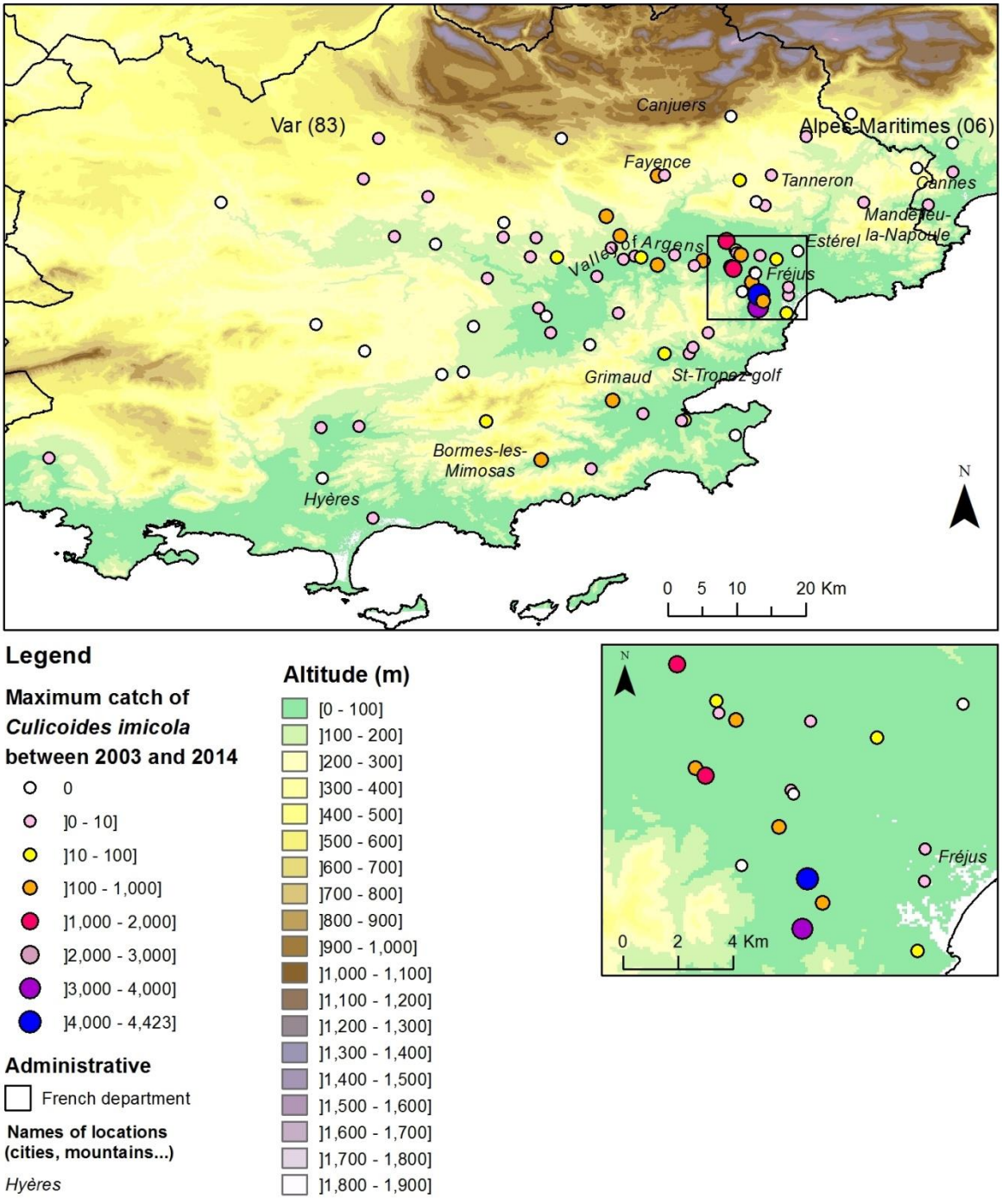


Figure 2 Maximum catch per year of *C. imicola* in the Var department from 2003 to 2014. Population abundance maps from 2003 to 2010 were previously published in [10].

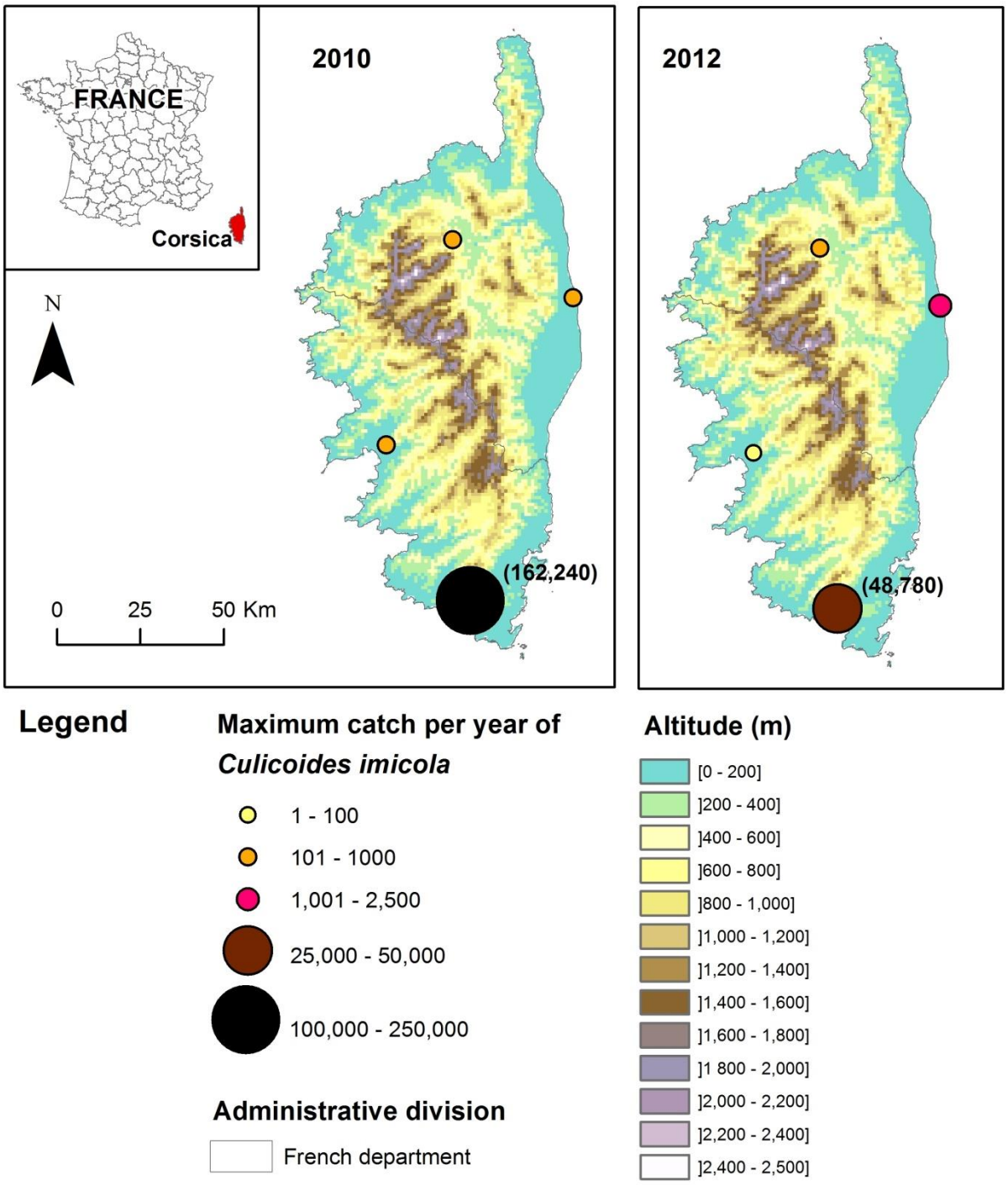


Figure 3 Collection sites in Corsica with maximum catches of *C. imicola* per year in 2010 and 2012. Population abundance maps from 2003 to 2009 were previously published in [10].

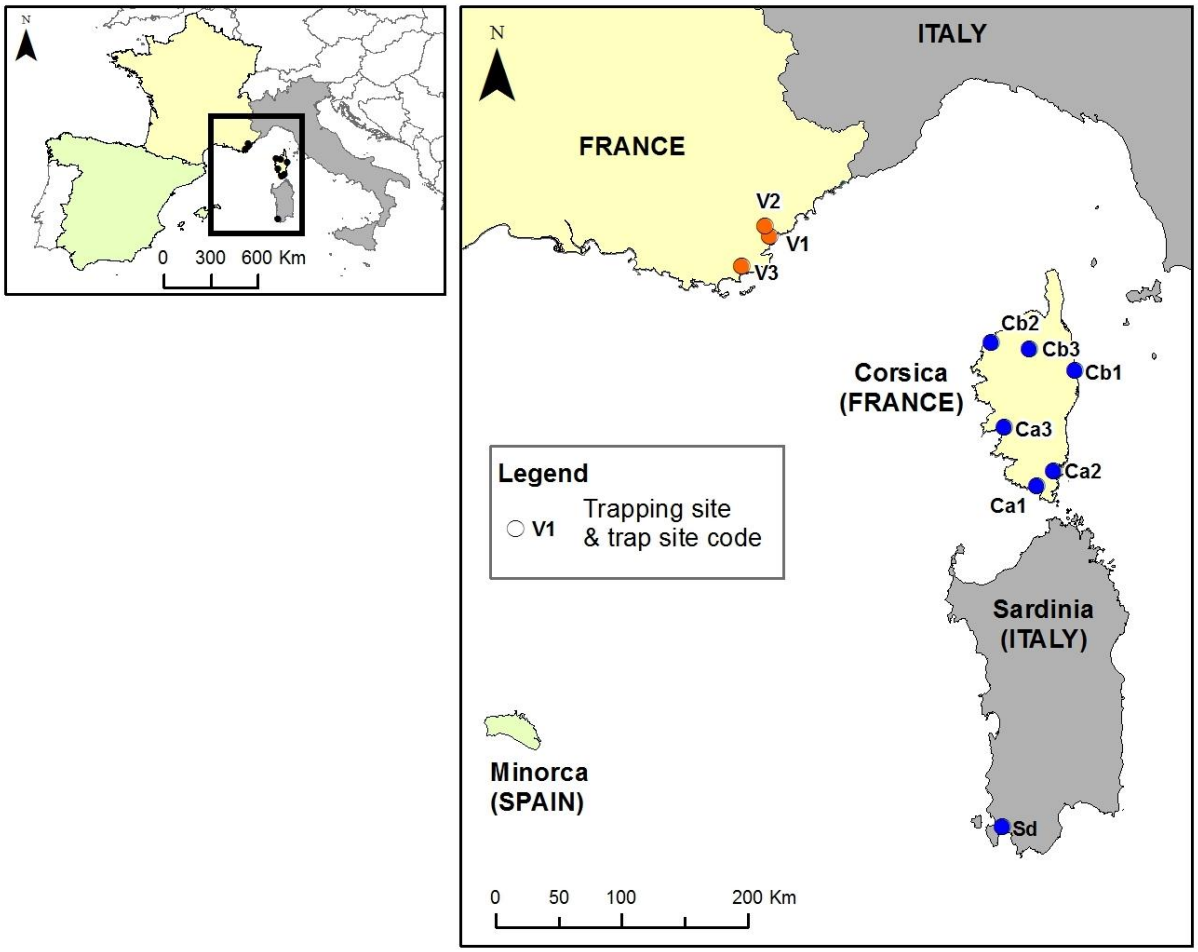


Figure 4 Sampled sites of *C. imicola* used for population genetics analyses.

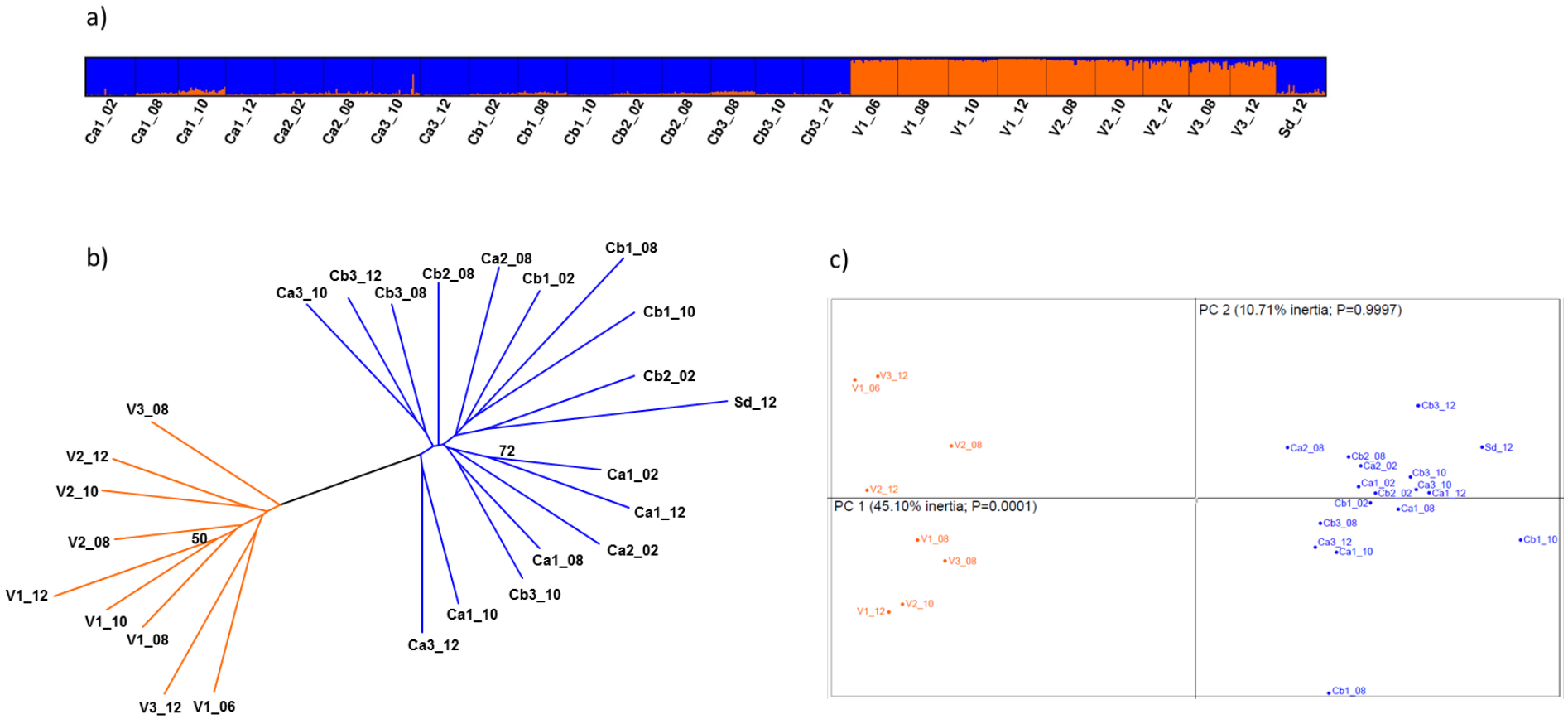


Figure 5 Population genetic structure results of *C. imicola*. (a) Genetic clustering of *C. imicola* samples. Each vertical line represents an individual and each color represents a cluster. (b) Microsatellite neighbor-joining tree based on genetic distance of Cavalli-Sforza & Edwards (1967). Bootstrap values are calculated over 1,000 replicates (only values > 50% are shown). (c) Principal Component Analysis based on microsatellite allelic frequencies.

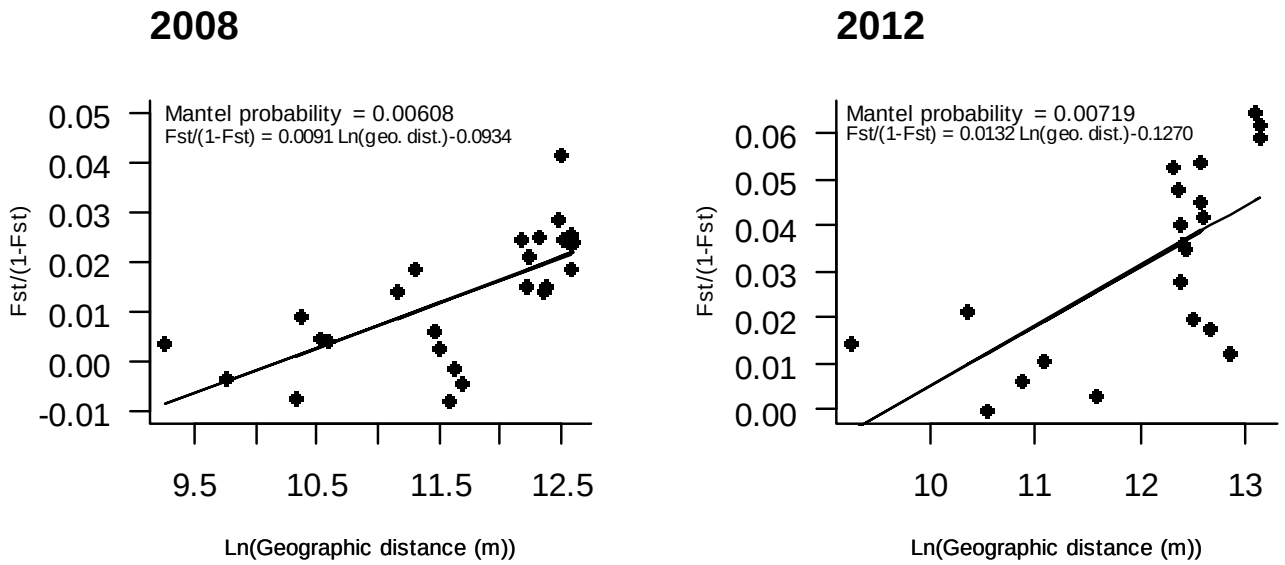


Figure 6 Results of the Mantel tests for Isolation by distance (IBD). The tests were performed for all Corsican and Var department sampled sites for the collection years 2008 and 2012.

Table 1 Geographical location of *C. imicola* sampled sites for the population genetics study.

					2002		2006		2008		2010		2012	
Country	**Locations**	**LON**	**LAT**	**Code**	**Collection date**	**N**	**Collection date**	**N**	**Collection date**	**N**	**Collection date**	**N**	**Collection date**	**N**
France, Corsica	Figari	9.08	41.50	Ca1	10/09/2002	32	-	-	18/09/2008	28	07/09/2010	31	19/09/2012	32
France, Corsica	Porto-Vecchio	9.25	41.59	Ca2	10/09/2002	32	-	-	19/11/2008	32	-	-	-	-
France, Corsica	Bastelicaccia	8.82	41.94	Ca3	-	-	-	-	-	-	10/05/2010	31	13/11/2012	32
France, Corsica	San Giuliano	9.54	42.29	Cb1	12/11/2002	32	-	-	18/09/2008	32	07/09/2010	30	-	-
France, Corsica	Calvi	8.76	42.54	Cb2	28/10/2002	32	-	-	01/09/2008	32	-	-	-	-
France, Corsica	Moltifao	9.12	42.47	Cb3	-	-	-	-	26/08/2008	29	10/08/2010	31	20/11/2012	31
France, Var department	Roquebrune-sur-Argens	6.68	43.40	V1	-	-	20/06/2006	32	30/09/2008	32	10/09/2010	32	18/09/2012	32
France, Var department	Roquebrune-sur-Argens	6.64	43.49	V2	-	-	-	-	30/09/2008	32	10/09/2010	31	18/09/2012	30
France, Var department	Bormes-Les-Mimosas	6.40	43.20	V3	-	-	-	-	28/08/2008	27	-	-	20/09/2012	30
Italy, Sardinia	San Giovanni Suergiu	39.13	8.52	Sd	-	-	-	-	-	-	-	-	06/11/2012	32

N is the number of individuals typed for microsatellite analyses.
The dashes indicate the absence of data

Table 2 Genetic diversity and Bottleneck results based on microsatellite data for each sampled site.

Country	Locations	Code	A_R	H_O	H_E	F_{IS}	IAM	TPM	SMM
France-Corsica	Figari	Ca1_02	4.390 ± 1.311	0.534 ± 0.132	0.569 ± 0.100	0.061	0.082	0.633	0.981
France-Corsica	Figari	Ca1_08	4.483 ± 1.583	0.573 ± 0.156	0.594 ± 0.121	0.037	**0.010**	0.285	0.850
France-Corsica	Figari	Ca1_10	4.285 ± 1.188	0.582 ± 0.154	0.590 ± 0.141	0.014	**0.005**	**0.024**	0.935
France-Corsica	Figari	Ca1_12	4.263 ± 1.310	0.574 ± 0.144	0.592 ± 0.125	0.031	**0.019**	0.150	0.590
France-Corsica	Porto-Vecchio	Ca2_02	4.541 ± 1.481	0.552 ± 0.112	0.572 ± 0.104	0.034	0.082	0.633	0.976
France-Corsica	Porto-Vecchio	Ca2_08	4.431 ± 1.512	0.555 ± 0.110	0.599 ± 0.095	0.074	**0.010**	0.180	0.850
France-Corsica	Bastelicaccia	Ca3_10	4.562 ± 1.688	0.616 ± 0.218	0.603 ± 0.138	-0.022	**0.014**	0.367	0.820
France-Corsica	Bastelicaccia	Ca3_12	4.568 ± 1.365	0.549 ± 0.182	0.594 ± 0.159	0.077	**0.019**	0.082	0.976
France-Corsica	San Giuliano	Cb1_02	4.240 ± 0.999	0.566 ± 0.154	0.574 ± 0.129	0.014	**0.010**	0.367	0.850
France-Corsica	San Giuliano	Cb1_08	4.476 ± 1.194	0.661 ± 0.117	0.616 ± 0.087	-0.075	**0.010**	0.285	0.787
France-Corsica	San Giuliano	Cb1_10	4.293 ± 1.369	0.550 ± 0.177	0.578 ± 0.116	0.050	**0.010**	0.326	0.850
France-Corsica	Calvi	Cb2_02	4.368 ± 1.327	0.601 ± 0.150	0.611 ± 0.110	0.018	**0.007**	0.180	0.674
France-Corsica	Calvi	Cb2_08	4.264 ± 0.979	0.523 ± 0.134	0.607 ± 0.115	0.140	**0.002**	0.150	0.820
France-Corsica	Moltifao	Cb3_08	4.219 ± 1.358	0.546 ± 0.120	0.603 ± 0.110	0.096	**0.001**	0.064	0.752
France-Corsica	Moltifao	Cb3_10	4.121 ± 0.867	0.611 ± 0.162	0.605 ± 0.131	-0.010	**0.005**	0.064	0.590

France-Corsica	Moltifao	Cb3_12	4.307 ± 1.330	0.568 ± 0.197	0.576 ± 0.154	0.014	0.064	0.410	0.850
France-Var department	Roquebrune-sur-Argens	V1_06	3.680 ± 0.794	0.560 ± 0.100	0.561 ± 0.082	0.103	**0.005**	**0.020**	0.590
France-Var department	Roquebrune-sur-Argens	V1_08	3.501 ± 0.926	0.609 ± 0.112	0.584 ± 0.094	-0.044	**0.002**	**0.010**	0.326
France-Var department	Roquebrune-sur-Argens	V1_10	3.690 ± 1.366	0.603 ± 0.150	0.587 ± 0.100	-0.029	**0.001**	**0.005**	0.213
France-Var department	Roquebrune-sur-Argens	V1_12	3.566 ± 1.176	0.573 ± 0.125	0.592 ± 0.091	0.033	**0.001**	**0.001**	0.082
France-Var department	Roquebrune-sur-Argens	V2_08	3.525 ± 0.361	0.515 ± 0.131	0.556 ± 0.127	0.075	**0.003**	0.064	0.545
France-Var department	Roquebrune-sur-Argens	V2_10	3.744 ± 1.091	0.596 ± 0.146	0.581 ± 0.106	-0.027	**0.003**	**0.014**	0.367
France-Var department	Roquebrune-sur-Argens	V2_12	4.031 ± 0.957	0.591 ± 0.103	0.570 ± 0.093	-0.038	**0.005**	0.285	0.875
France-Var department	Bormes-Les-Mimosas	V3_08	3.820 ± 1.045	0.558 ± 0.147	0.572 ± 0.122	0.024	**0.014**	0.150	0.633
France-Var department	Bormes-Les-Mimosas	V3_12	4.067 ± 1.045	0.519 ± 0.160	0.558 ± 0.117	0.072	**0.005**	0.410	0.981
Italy-Sardinia	San Giovanni Suergiu	Sd_12	4.513 ± 1.926	0.560 ± 0.156	0.590 ± 0.143	0.052	**0.018**	0.248	0.787

The allelic richness (A_R), observed (H_O) and expected (H_E) heterozygosity and F_{IS} are presented for each population. A_R is based on the minimum sample size of 27 diploid individuals. Results of bottleneck tests are presented for the Infinite Allele Model (IAM), Two-phase Model (TPM) and Stepwise Mutation Model (SMM).
The number at the end of each sample code corresponds to the collection year.
Significant results of bottleneck tests are indicated in bold.

Table 3 Pairwise F_{ST} values between Corsican and Var *C. imicola* population samples.

	Cb3_08	V1_08	V2_08	V3_08
Ca1_08	-0.0076	**0.0247**	**0.0251**	**0.0259**
Cb3_08		**0.0240**	**0.0138**	**0.0144**
V1_08			0.0034	0.0089
V2_08				0.0047

	Cb3_12	V1_12	V2_12	V3_12
Ca1_12	0.0027	**0.0506**	**0.0400**	**0.0431**
Cb3_12		**0.0501**	**0.0457**	**0.0385**
V1_12			0.0139	0.0206
V2_12				-0.0004

Pairwise F_{ST} values were computed for 2008 (atop) and 2012 (below). Significant F_{ST} are represented in bold.

Supplementary Information

Table S1 Primers used for the amplification of the microsatellite loci in *C. imicola* (Mardulyn *et al*. 2013).

Loci	Motif	Forward / Reverse	Allele size range (bp)	Ta (°C)
68	(GT)n	CTTTTCCGTTTCTTTTTATTTCTTT GTTTCTTTCTGGTCGCGTTGGTTGCTG	101-105	60
12b	(CT)n	TTATGTGTGTATGTTAGCAAGGTCA GTTTCTTCTTCGGATCAAAGAAATTTTGCC	133-139	50
3b	(AC)n	ATGCGGATGTTTGAAGTG GTTTCTTTTTTGTGTCTTATTGCCC	154-175	50
31	(CAA)n	TTCTGTTCGGCTGTTGCGTT GTTTCTTCTTTTTACGTGGTGGTCATTC	162-166	60
41b	(CT)n	GAGGAGGAGGTAGAA GTTTCTTTCTATTAGTCAATGGTG	162-166	50
35t	(AC)n	TTTGTAAAAGCCAGTTCAACCG GTTTCTTATCGAACGAAGGAAATAACCAC	181-188	60
88b	(AC)n	TTTGTTCGATTTGTAGTG GTTTCTTCCTCTCTTTCATTCGC	243-256	50
16	(TG)n	TTGCCTTTGCTTGTTGAGGATG GTTTCTTTCCTCTTTAAAATCACTGACGTG	292-299	60
88	(CAT)n	GTTGGTGCTTTGTTGTGTTGT GTTTCTTTTTCTTTTTCTCCTTTTTGTTTCTTTC	344-348	50

Table S2 Details of the entomological surveys realized in Var department in 2011 to 2012.

		2011		2012	
Code site	Location	Collection date	Nb of *C. imicola* collected per trap per night	Collection Date	Nb of *C. imicola* collected per trap per night
06BT0	Mandelieu-La Napoule	04/10/11	0	18/09/12	1
06BT1	Saint-Cézaire-sur-Siagne	-	-	19/09/12	0
06BT03	Pégomas	03/10/11	0	18/09/12	0
83BT01	Roquebrune-sur-Argens	05/10/11	11	17/09/12	134
83BT16	Grimaud	05/10/11	7	19/09/12	52
83BT31	Sainte-Maxime	05/10/11	0	19/09/12	9
83BT33	Bormes-Les-Mimosas	04/10/11	27	18/09/12	701
83BT35	Hyères	04/10/11	0	18/09/12	0
83BT40	Taradeau	05/10/11	0	17/09/12	0
83BT43	Les Adrets-de-L'Estérel	03/10/11	0	18/09/12	0
83BT44	Tanneron	03/10/11	0	18/09/12	0
83BT46	Saint-Paul-en-Forêt	-	-	20/09/12	0
83BT50	Draguignan	05/10/11	0	20/09/12	0
83BT56	Entrecasteaux	06/10/11	0	20/09/12	0
83BT57	Le Cannet-des-Maures	05/10/11	0	19/09/12	0
83BT64	Cabasse	03/10/11	0	17/09/12	0
83BT66	Collobrières	04/10/11	0	18/09/12	25
83BT69	Besse-sur-Issole	03/10/11	0	17/09/12	0
83BT70	Camps-la-Source	03/10/11	0	-	-
83BT71	Cuers	04/10/11	0	20/09/12	0
83BT72	Salernes	06/10/11	0	20/09/12	1
83BT74	Seillans	04/10/11	0	20/09/12	0
83BT75	Callian	04/10/11	0	19/09/12	0
83BT76	Aups	06/10/11	0	20/09/12	3
83BT77	Sillans-la-Cascade	06/10/11	0	20/09/12	2

83BT78	Le Muy	05/10/11	0	17/09/12	0
83BT79	Callas	06/10/11	0	19/09/12	2
83PL1	Roquebrune-sur-Argens	05/10/11	0	17/09/12	232
83PL2	Barjols	05/10/11	0	17/09/12	0
83PS6	Hyères	04/10/11	0	18/09/12	0

The dashes indicate no collection.

Table S3 Details of the entomological surveys realized in Corsica from 2010 and 2012.

		2010		**2012**	
Code	**Locations**	**Collection date**	**Maximum catch per trap per night**	**Collection date**	**Maximum catch per trap per night**
Ca1	Figari	11/08/2010	162240	17/07/2012	48780
Ca3	Bastelicaccia	08/09/2010	717	13/11/2012	58
Cb1	San-Giuliano	04/10/2010	927	15/10/2012	1597
Cb3	Moltifao	10/08/2010	145	20/11/2012	379

Table S4 Pairwise F_{ST} values between Corsican and Var populations of *C. imicola* for all collected years.

	Ca1_08	Ca1_10	Ca1_12	Ca2_02	Ca2_08	Ca3_10	Ca3_12	Cb1_02	Cb1_08	Cb1_10	Cb2_02	Cb2_08	Cb3_08	Cb3_10	Cb3_12	V1_06	V1_08	V1_10	V1_12	V2_08	V2_10	V2_12	V3_08	V3_12	Sd_12
Ca1_02	-0.0104	-0.0029	-0.0081	0.0001	0.0004	0.0030	0.0027	-0.0050	0.0109	-0.0015	0.0027	-0.0019	-0.0057	-0.0033	0.0046	**0.0363**	**0.0306**	**0.0329**	**0.0407**	**0.0212**	**0.0282**	0.0262	0.0205	**0.0334**	0.0157
Ca1_08		-0.0075	-0.0090	0.0025	-0.0036	-0.0029	0.0002	-0.0091	0.0054	-0.0042	-0.0081	-0.0045	-0.0084	-0.0104	0.0024	**0.0408**	**0.0245**	**0.0351**	**0.0382**	0.0236	**0.0289**	0.0307	0.0246	**0.0363**	0.0097
Ca1_10			-0.0048	0.0030	0.0104	-0.0054	-0.0067	0.0026	0.0046	0.0047	0.0007	0.0028	-0.0048	0.0019	0.0075	**0.0418**	**0.0277**	**0.0335**	**0.0317**	**0.0186**	0.0181	**0.0307**	0.0227	**0.0306**	**0.0215**
Ca1_12				0.0046	0.0038	-0.0021	0.0062	-0.0020	0.0123	0.0021	-0.0012	0.0023	0.0020	-0.0040	0.0027	**0.0468**	**0.0362**	**0.0424**	**0.0506**	**0.0332**	**0.0334**	**0.0400**	**0.0352**	**0.0431**	0.0194
Ca2_02					0.0006	0.0056	0.0044	0.0104	0.0201	0.0085	0.0104	0.0038	0.0032	0.0087	0.0066	**0.0427**	**0.0406**	**0.0347**	**0.0427**	**0.0296**	**0.0356**	**0.0433**	**0.0253**	**0.0361**	0.0182
Ca2_08						0.0126	0.0099	-0.0002	0.0180	0.0097	-0.0003	-0.0016	0.0026	0.0038	0.0108	**0.0267**	**0.0181**	**0.0229**	**0.0349**	**0.0232**	**0.0297**	**0.0239**	**0.0227**	**0.0290**	0.0158
Ca3_10							0.0045	0.0118	0.0148	0.0050	-0.0011	0.0019	-0.0033	0.0012	-0.0016	**0.0479**	**0.0399**	**0.0374**	**0.0438**	**0.0303**	**0.0343**	**0.0430**	**0.0329**	**0.0408**	0.0090
Ca3_12								0.0037	0.0102	0.0067	0.0069	-0.0007	-0.0003	0.0039	0.0104	**0.0425**	**0.0275**	**0.0307**	**0.0265**	**0.0161**	**0.0226**	**0.0337**	0.0223	**0.0345**	0.0172
Cb1_02									0.0116	0.0048	0.0030	0.0049	0.0023	0.0005	0.0178	**0.0381**	**0.0259**	**0.0438**	**0.0497**	**0.0278**	**0.0312**	**0.0325**	**0.0355**	**0.0421**	**0.0199**
Cb1_08										0.0089	0.0105	0.0137	0.0037	0.0128	0.0290	**0.0577**	**0.0276**	**0.0409**	**0.0316**	**0.0395**	**0.0202**	**0.0378**	**0.0238**	**0.0575**	**0.0289**
Cb1_10											-0.0003	0.0020	0.0014	0.0020	0.0061	**0.0670**	**0.0507**	**0.0577**	**0.0484**	**0.0445**	**0.0512**	**0.0543**	**0.0396**	**0.0603**	0.0012
Cb2_02												-0.0023	0.0007	-0.0028	0.0028	**0.0399**	**0.0241**	**0.0345**	**0.0362**	**0.0261**	**0.0302**	**0.0307**	**0.0254**	**0.0368**	0.0019
Cb2_08													-0.0074	-0.0037	0.0019	**0.0347**	**0.0235**	**0.0277**	**0.0372**	0.0147	**0.0317**	**0.0291**	0.0204	**0.0319**	0.0018
Cb3_08														-0.0023	0.0057	**0.0369**	**0.0240**	**0.0243**	**0.0288**	0.0138	**0.0205**	0.0215	0.0144	**0.0295**	0.0071
Cb3_10															-0.0036	**0.0443**	**0.0326**	**0.0351**	**0.0429**	**0.0265**	**0.0372**	**0.0376**	**0.0281**	**0.0415**	0.0102
Cb3_12																**0.0512**	**0.0489**	**0.0392**	**0.0501**	**0.0291**	**0.0482**	**0.0457**	**0.0363**	**0.0385**	0.0117
V1_06																	**0.0069**	**0.0086**	**0.0261**	0.0081	**0.0157**	0.0069	0.0168	-0.0005	**0.0635**
V1_08																		0.0016	0.0083	0.0034	0.0034	0.0058	0.0089	0.0181	**0.0516**
V1_10																			0.0018	0.0000	0.0041	0.0015	-0.0003	0.0092	**0.0548**
V1_12																				0.0130	0.0060	0.0139	0.0008	0.0206	**0.0559**
V2_08																					0.0061	-0.0017	0.0047	-0.0011	**0.0418**
V2_10																						-0.0029	-0.0028	0.0115	**0.0617**
V2_12																							-0.0012	-0.0004	**0.0584**
V3_08																								0.0119	**0.0472**
V3_12																									**0.0604**

Significant F_{ST} are represented in bold. Details on sample codes are given in Table 1.

Discussion générale

Les objectifs de cette thèse étaient (i) d'apporter des connaissances sur l'histoire évolutive et démographique de l'espèce *C. imicola*, espèce vectrice de virus d'intérêt vétérinaire et (ii) de déterminer et comprendre les facteurs sous-jacents au succès de son expansion géographique dans le bassin méditerranéen. Pour répondre à ces objectifs, nous avons retracé les routes de colonisation de *C. imicola* depuis l'Afrique subsaharienne. L'origine des populations installées au Maghreb, au Moyen-Orient, dans le sud de l'Europe et les voies empruntées par les individus fondateurs et la chronologie de ces évènements ont ainsi été caractérisées (chapitre I, article 1). Nous avons complété l'histoire de cette colonisation par l'étude des populations installées à l'échelle du bassin méditerranéen afin de définir les caractéristiques démographiques, évolutives et temporelles de la colonisation du sud de l'Europe (chapitre II, article 2). Enfin, l'étude des populations à la limite septentrionale de la distribution géographique a permis d'identifier les principaux facteurs expliquant le succès d'expansion géographique des populations installées et donc de mieux comprendre la dynamique d'expansion actuelle de *C. imicola* (chapitre III, article 3 et 4). Les principaux résultats de cette thèse décrivent les évènements majeurs, historiques et démographiques de l'histoire de la colonisation de *C. imicola* et permettent de proposer des hypothèses pour expliquer le succès de son installation dans le bassin méditerranéen. Ce travail permet également d'approfondir les connaissances fondamentales sur les processus de colonisations naturelles et de fournir des prérequis pour l'élaboration de stratégies de contrôle et de gestion des populations de cette espèce.

I. De l'Afrique au sud de l'Europe : l'histoire de la colonisation de *C. imicola*

Bien que l'implication de *C. imicola* dans la transmission d'*Orbivirus* en région afrotropicale soit connue depuis les années 40 (Du Toit 1944), le regain d'intérêt pour cette espèce est récent, en lien avec les épizooties massives de fièvre catarrhale ovine dans le bassin méditerranéen. La découverte de populations de *C. imicola* dans la majorité des régions nouvellement affectées par la FCO a conféré un statut d'espèce envahissante à ce moucheron pour ces régions. Toutefois, l'histoire de la colonisation du bassin méditerranéen par les populations de *C. imicola* était jusqu'alors incomplète.

L'ensemble de nos résultats permet de dresser un scenario global retraçant l'histoire de la colonisation de *C. imicola* depuis son aire native (Figure 1). Ce scenario, présenté comme une hypothèse, permet de mieux comprendre les évènements historiques et démographiques qui ont modelé la distribution de l'espèce, et ceux qui régissent aujourd'hui la dynamique actuelle des populations installées dans le bassin méditerranéen.

En Afrique subsaharienne, l'aire d'origine de *C. imicola*, les marqueurs mitochondriaux montrent une différentiation génétique des populations bien marquée entre l'ouest et l'est du continent, mais moins apparente lors de l'analyse des marqueurs microsatellites qui suggère un continuum de différenciation génétique entre l'ouest et le sud. Durant le Pléistocène supérieur (126 000 - 11 700 ans), une période marquée par des variations climatiques importantes, ces populations auraient subi une expansion démographique. A la même époque, ou plus tard au début de l'Holocène (aux alentours de 10 000 ans), l'humidification du Sahara, caractérisée par la présence de végétation et de lacs (Castaneda *et al.* 2009) aurait permis le passage d'individus fondateurs depuis la limite nord de la distribution en Afrique subsaharienne vers le Maghreb et le Moyen Orient. La différenciation génétique nucléaire observée entre les populations natives et les populations méditerranéennes traduirait ce premier évènement d'introduction. Lors d'une période aride, le Sahara aurait progressivement évolué pour former le désert que nous connaissons aujourd'hui. Cette barrière géographique restreindrait les flux géniques entre les populations situées au nord (populations du Maghreb) et au sud du Sahara, expliquant l'absence d'haplotypes partagés entre elles. La présence de couloirs présentant des habitats favorables à *C. imicola*, comme la vallée du Nil, aurait pu permettre le passage d'émigrants de la zone subsaharienne vers le Moyen-Orient, comme le suggère le partage d'haplotypes mitochondriaux entre ces deux zones.

Plus tard, les populations installées sur la côte sud méditerranéenne (du Maroc à Israël) auraient divergé en restant suffisamment isolées pour former deux groupes génétiques - un groupe à l'ouest et un groupe à l'est du bassin méditerranéen – structuration qui est marquée sur l'ensemble des marqueurs génétiques utilisés. Les processus à l'origine de cette divergence restent encore à élucider. Des facteurs environnementaux tels que les variations climatiques, l'apparition de barrières physiques ou la fragmentation des habitats pourraient expliquer cette subdivision génétique. Une question reste en suspens, les populations des deux groupes sont-elles totalement isolées ou existe-t-il une connexion géographique sur une zone non étudiée dans nos travaux ? Une partie du Maghreb (le sud de l'Algérie, la Libye et l'Egypte) est difficilement accessible pour des collectes entomologiques du fait de la situation

politique de ces pays. L'inclusion de ces populations dans nos jeux de données permettrait probablement de répondre à cette question.

A partir de ces deux groupes, des individus auraient colonisé ensuite le nord du bassin méditerranéen. A l'est, l'origine des populations grecques n'est pas totalement confirmée, mais le lien généalogique avec les populations turques suggère que ces dernières sont des sources potentielles. A l'ouest du bassin méditerranéen, les moucherons auraient emprunté deux routes pour coloniser le sud-ouest de l'Europe. Il y a au moins 200 ans, le Maroc aurait servi de source aux populations espagnoles et portugaises, tandis que l'Algérie aurait fourni des émigrants à la France et à l'Italie. La différenciation génétique faible, attesterait d'une dynamique de colonisation avec des flux géniques récurrents entre les populations. La forte capacité de *C. imicola* à se disperser par les vents apparait fondamentale dans le succès de la colonisation des régions méditerranéennes. Le rôle du réchauffement climatique contemporain n'est pas à exclure et pourrait avoir contribué à l'expansion géographique de *C. imicola* en augmentant la taille des populations (Guis *et al.* 2012; Purse *et al.* 2005), en créant de nouveaux habitats favorables l'installation de l'espèce, ou en augmentant sa dispersion par le vent. D'autre part, les activités humaines pourraient avoir facilité l'établissement des populations en modifiant les habitats. Notamment, l'augmentation de la disponibilité en hôtes via l'intensification des élevages au cours des siècles derniers pourrait avoir contribué à l'accroissement local des populations. Aujourd'hui, la dynamique d'expansion des populations installées semble, au moins partiellement, régie par les abondances locales et leurs capacités de dispersion, les conditions environnementales à la limite septentrionale contribuant aussi à limiter ces expansions.

Les routes de colonisation décrites dans ce scénario ont été inférées à partir du polymorphisme des marqueurs microsatellites. Néanmoins, les liens généalogiques déduits du polymorphisme mitochondrial suggèrent l'existence de deux populations sources et de deux routes de colonisation depuis l'aire native. Les résultats pris dans leur ensemble tendent à soutenir le scénario présenté (Figure 1A), mais on ne peut totalement exclure l'hypothèse d'une colonisation impliquant deux sources et deux voies d'introduction indépendantes (Figure 1B). Dans ce scénario alternatif, les individus fondateurs de *C. imicola* seraient probablement remontés vers le bassin méditerranéen via des corridors connectant d'une part l'Afrique de l'Ouest à la région ouest du bassin méditerranéen, et d'autre part, l'Afrique de l'Est à la région est du bassin méditerranéen. Les populations méditerranéennes ayant colonisé indépendamment la région, les populations installées à l'ouest et l'est ont pu diverger

et n'être jamais rentrées en contact. Des travaux supplémentaires pourraient clarifier des points d'étape des scénarii proposés notamment sur la colonisation du sud-est de l'Europe.

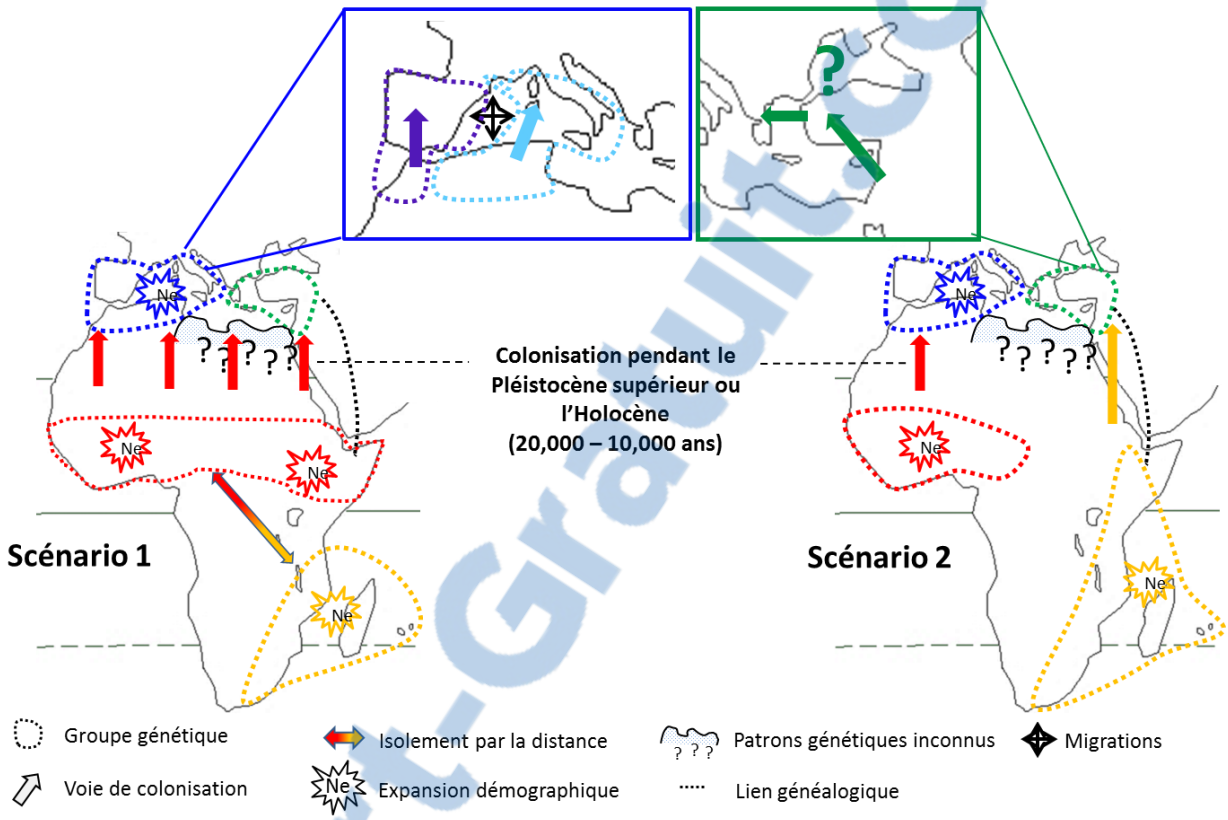


Figure 1. Représentation schématique des hypothèses de scénarii de la colonisation du bassin méditerranéen par *C. imicola*

II. Approche multi-marqueurs : congruences, discordances et perspectives

L'utilisation des marqueurs moléculaires pour étudier les aspects écologiques et évolutifs des populations, ont permis de reconstruire l'histoire évolutive des invasions biologiques avec parfois une grande précision (Handley *et al.* 2011). Notamment, la combinaison multi-loci peut réduire la variance des paramètres estimés résultant des effets aléatoires des échantillonnages et du tri incomplet des lignées (Brito 2007), et apporter de nouvelles connaissances qui n'auraient pu être obtenues avec l'utilisation d'un marqueur unique (Prugnolle & De Meeûs 2002).

Nous avons combiné l'information génétique portée par neuf marqueurs microsatellites et deux gènes mitochondriaux (COI et CytB) pour retracer l'histoire évolutive de *C. imicola* et déterminer les mécanismes à l'origine de la distribution actuelle de l'espèce. D'un point de vue pratique, ces marqueurs sont simples d'utilisation et leur génotypage peu coûteux, permettant ainsi d'extraire l'information génétique d'un grand nombre d'individus. Hautement polymorphes et considérés comme neutre et co-dominants, les microsatellites sont des marqueurs de choix pour étudier la variabilité génétique intra et inter-populationnelle *in natura*, la structure contemporaine des populations à fine échelle et les évènements démographiques relativement récents (Hewitt 2004). L'ADN mitochondrial a une héritabilité maternelle, évolue rapidement et présente peu ou pas de recombinaison, ce qui en fait un bon marqueur pour définir les processus historiques et reconstruire la généalogie intra-spécifique (Avise 2000).

La combinaison de ces marqueurs s'est révélée pertinente pour décrire les liens généalogiques et la structure des populations de *C. imicola* à différentes échelles spatiales (globale, régionale et locale). De plus, elle nous a permis de caractériser des processus démographiques et évolutifs ayant opéré à des échelles temporelles distinctes qui s'étendent des temps géologiques au contemporain. Ainsi, les patrons génétiques congruents entre marqueurs ont apporté une perspective relativement robuste de l'histoire évolutive de *C. imicola.* Toutefois, il est à noter que des résultats discordants ont également été obtenus.

Des discordances mito-nucléaires ont été reportées dans de nombreuses études (e.g. Johnson *et al.* 2003; Kolleck *et al.* 2013; Pavlova *et al.* 2013; Rašić *et al.* 2015) et sont fréquemment associées à la différenciation génétique entre populations. Ces conflits peuvent s'expliquer par différents facteurs incluant les taux d'évolution et les modes d'héritabilité des marqueurs, le balayage sélectif « selective sweeps », les phénomènes d'introgression et certains traits d'histoire de vie (dispersion biaisée en fonction le sexe) (Toews & Brelsford 2012; Zink & Barrowclough 2008).

II.1. Discordances mito-nucléaires : structure des populations natives de *C. imicola*

Dans notre cas d'étude, une première discordance apparait dans la structuration des populations natives. Les séquences mitochondriales indiquent la présence de deux groupes génétiques distincts opposant les populations de l'Afrique de l'Ouest à celles de l'Afrique de l'Est et de l'Océan Indien. En revanche, les marqueurs microsatellites suggèrent un

continuum ouest/sud avec un patron d'isolement par la distance, où les populations situées à l'extrémité géographique du continuum (Afrique de l'Ouest et Océan Indien) sont très différenciées.

Dans la littérature, l'hypothèse d'un biais de dispersion sexe-spécifique (l'un des sexes étant phylopatrique) est souvent avancée pour expliquer ce type de conflit mito-nucléaire (Zink & Barrowclough 2008). Cette hypothèse semble peu probable dans notre cas, puisque la grande capacité de dispersion passive par les vents s'applique aussi bien aux mâles et qu'aux femelles de *Culicoides* (Braverman 1991). Une seconde hypothèse pouvant expliquer les discordances mito-nucléaires est la sélection naturelle d'un ou plusieurs haplotypes mitochondriaux dans une zone donnée. En effet, en dépit du fait que l'ADN mitochondrial soit considéré comme un marqueur neutre, plusieurs études ont identifié des variations intra- et inter-spécifiques dans des protéines codées par les gènes mitochondriaux, suggérant le rôle de la sélection naturelle (Ballard & Melvin 2010; Bazin *et al.* 2006; Edwards & Bensch 2009). Si la sélection de variants mitochondriaux varie géographiquement, alors des patrons discordants sont attendus entre l'ADN mitochondrial et les gènes nucléaires (Irwin 2012). Nous avons testé si les gènes mitochondriaux utilisés au cours de notre étude étaient sous sélection. Le test n'a révélé aucun signal de sélection positive. Bien que de la sélection négative ait été détectée, celle-ci n'était pas géographique-dépendante et était présente dans l'ensemble du jeu de données (dans toutes les populations et sur tous les marqueurs), donc écartant un rôle de la sélection dans les patrons génétiques mito-nucléaires observés.

Un autre facteur pouvant expliquer la discordance entre marqueurs nucléaires et mitochondriaux est le mode d'héritabilité. La transmission de l'ADN mitochondrial est maternelle contrairement à celle de l'ADN nucléaire qui est bi-parentale. Ainsi, la taille efficace liée aux marqueurs mitochondriaux est théoriquement égale à un quart de celle d'un gène nucléaire. Par conséquent, le tri des lignées mitochondriales sera plus rapide et le taux d'extinction des allèles ancestraux sera élevé (Zhang & Hewitt 2003). Si les populations subsahariennes de *C. imicola* étaient isolées, notamment sous l'effet des variations climatiques du Pléistocène comme le suggèrent nos données, il est possible que la divergence des populations ait eu le temps d'être détectée dans l'ADN mitochondrial, mais que ce temps eut été trop court pour qu'elle soit perceptible sur les marqueurs microsatellites. De plus, les microsatellites étant des marqueurs hypervariables, une accumulation plus importante de mutations homoplasiques pourrait également conduire à une réduction de la différentiation génétique entre les populations.

Alternativement, dans le cas d'une divergence ancienne, la discordance mito-nucléaire pourrait s'expliquer par l'existence de flux géniques qui homogénéiserait la variabilité génétique intra- et inter-populationnelle des microsatellites (Edwards & Bensch 2009). Cette dernière hypothèse est compatible avec le patron d'isolement des populations par la distance observé dans l'aire native.

II.2. Discordance mito-nucléaire : liens généalogiques entre populations

De ce conflit mito-nucléaire concernant la structure des populations natives de *C. imicola,* découle une seconde discordance au niveau des liens généalogiques entre les populations méditerranéennes et subsahariennes. Les marqueurs microsatellites supportent une seule voie de colonisation impliquant une remontée des populations de *C. imicola* depuis la limite septentrionale de la distribution en l'Afrique subsaharienne vers la côte sud du bassin méditerranéen. A l'opposé, les gènes mitochondriaux suggèrent l'existence de deux voies de colonisation : une voie connectant l'Afrique de l'Ouest à la région ouest du bassin méditerranéen, et une route reliant l'Afrique de l'Est à la région est du bassin méditerranéen. Compte tenu des points discutés dans le paragraphe précédent, si le patron génétique observé à partir des séquences mitochondriales résulte d'une divergence récente entre les populations natives de l'ouest et du sud-est, il est possible que des populations fondatrices de *C. imicola* aient colonisé le bassin méditerranéen via deux corridors distincts depuis l'Afrique de l'Ouest et l'Afrique de l'Est. Toutefois, il est à noter que l'utilisation de l'ADN mitochondrial présente des limites méthodologiques. Du fait de son héritabilité uni-parentale, les inférences basées sur ce marqueur reflètent l'histoire de la lignée maternelle et donc représente une fraction incomplète de l'histoire de l'espèce (Edwards & Bensch 2009; Zhang & Hewitt 2003). Les relations évolutives inférées à l'aide de ce type de marqueur peuvent donc paraître plus simples qu'elles ne le sont en réalité (Zhang & Hewitt 2003). De plus, bien que notre échantillonnage couvre une grande partie de l'aire de distribution de *C. imicola,* celui-ci est limité dans la partie est du bassin méditerranéen, avec notamment des zones non échantillonnées (Lybie, Egypte). Il est donc possible que les liens généalogiques entre les populations méditerranéennes et les populations natives ne soient pas totalement caractérisés.

Si le patron génétique observé dans l'aire native résulte de flux géniques qui ne sont pas détectables avec les marqueurs mitochondriaux, une remontée de la limite septentrionale de la distribution en Afrique subsaharienne n'est pas totalement incompatible avec les résultats mitochondriaux. Il est à rappeler que nos inférences des scenarii de colonisation avec

les méthodes ABC sont basées sur les clusters génétiques définis à partir des analyses d'assignements d'individus. Par conséquent, une population source correspond à un cluster génétique. Si les populations colonisatrices sont remontées depuis la limite septentrionale de la distribution en Afrique subsaharienne, il est probable que les populations à l'ouest de l'Afrique subsaharienne aient emprunté des voies à l'ouest, et réciproquement à l'est, mais que ce signal ne soit pas suffisamment fort pour être détecté à l'aide de la méthode ABC. Les liens généalogiques inférés à l'aide des marqueurs mitochondriaux pourraient refléter dans ce cas, la distribution de la variabilité ancestrale dans l'aire native.

II.3. Approche multi-marqueurs : perspectives

Ces résultats montrent la nécessité de confronter l'information génétique de multiples marqueurs pour décrire et comprendre les processus à l'origine de la distribution actuelle d'une espèce. Des études complémentaires pourraient être mises en place avec un échantillonnage plus extensif, en Afrique Centrale et australe et dans la région est du bassin méditerranéen, et par l'emploi d'autres marqueurs génétiques indépendants. Ces études sont nécessaires pour clarifier les discordances mito-nucléaires observées et repréciser les routes de colonisation du bassin méditerranéen par *C. imicola*.

Des marqueurs génétiques hautement polymorphes répartis le long du génome, tels que les SNPs (Single Nucleotide Polymorphism) seraient d'excellents candidats pour élucider les points non clarifiés de l'histoire évolutive de *C. imicola*. La structure génomique des populations envahissantes est influencée par une variété de processus, incluant les évènements de goulot d'étranglement, les introductions multiples, l'expansion démographique des populations, les flux géniques entre populations et la sélection (Lee 2002). Par conséquent, l'analyse d'un grand nombre de polymorphismes génétiques distribué le long du génome, comme les marqueurs SNPs, pourrait faciliter l'identification de ces processus (Puzey & Vallejo-Marín 2014). Ce type de marqueurs, bien que bi-alléliques permettrait d'augmenter considérablement le nombre de loci analysés compensant ainsi cette perte d'information (Peery *et al.* 2012). L'accessibilité croissante aux technologies de séquençage à haut débit, à des coûts raisonnables, a permis l'étude des patrons génétiques dans des populations naturelles, incluant les relations génétiques entre les populations natives et envahissantes et les voies de colonisation qui lient ces populations (Dlugosch *et al.* 2013; Puzey & Vallejo-Marín 2014). Lorsqu'un génome de référence est aussi disponible pour l'espèce étudiée, le reséquençage complet du génome permet de génotyper un nombre considérable de

marqueurs SNPs (Davey *et al.* 2011). Il n'existe pas à l'heure actuelle de génome complet et annoté pour le genre *Culicoides*. Cependant des approches bioinformatiques existent pour pallier à l'absence de séquence de référence (Iqbal *et al.* 2012; Peterlongo *et al.* 2010). Des marqueurs SNPs, microsatellites et mitochondriaux combinés aux méthodes statistiques ABC permettraient ainsi de confirmer et/ou révéler d'autres aspects de l'histoire évolutive et démographique des populations de *C. imicola*.

III. *Culicoides imicola* : expansion géographique versus invasion biologique ?

Jusqu'à récemment, *C. imicola* était considérée comme une espèce envahissante sur la seule base des observations historiques de présence dans le bassin méditerranéen. En étudiant l'histoire évolutive de cette espèce, ce travail de thèse a mis en lumière les caractéristiques de la colonisation dans cette zone. Plus spécifiquement, le modèle *C. imicola* apporte des connaissances fondamentales sur les processus éco-évolutifs des phénomènes de colonisation et d'invasions biologiques naturelles suivant deux scénarii distincts : (i) la colonisation d'un nouveau milieu suite à la suppression d'une barrière physique qui séparait initialement les aires native et colonisée à une échelle de temps importante, et (ii) la colonisation de nouveaux milieux géographiquement déconnectés de l'aire originelle à une échelle temporelle plus récente (quelques siècles).

Lorsqu'une espèce est détectée dans une zone éloignée de son aire d'origine et que son apparition est récente et en lien avec les activités humaines, le statut d'espèce envahissante peut être établi de façon quasi-certaine. Cependant, pour des populations ayant colonisé un nouveau milieu sans l'action directe de l'homme et sans discontinuité géographique marquée entre l'aire d'origine et colonisée, le manque de consensus sur la définition d'une espèce envahissante rend cette qualification difficile. En effet, certains auteurs se focaliseront sur le potentiel de l'espèce à être introduite dans un nouvel environnement, à former une population viable, à proliférer et à se répandre géographiquement pour qualifier une espèce d'envahissante (Facon *et al.* 2006). D'autres auteurs considèrent aussi les caractéristiques et processus des invasions contemporaines engendrées par les activités humaines (introductions multiples, plusieurs populations sources, géographie large, échelle de temps relativement courte) pour différencier les invasions biologiques de ce qu'ils appellent la « colonisation naturelle » (Cassey *et al.* 2005; Estoup & Guillemaud 2010). Un débat est donc nécessaire

pour établir des critères universellement admis afin de définir la notion d'invasion biologique. Il est tout d'abord nécessaire de définir une espèce native pour savoir comment qualifier une espèce qui est transportée en dehors de son aire originelle et qui parvient à s'établir dans des zones où elle était initialement absente.

III.1. Un statut à clarifier pour la côte sud du bassin méditerranéen

Nous avons montré que la remontée de *C. imicola* depuis son aire native vers le bassin méditerranéen avait été influencée par des variations climatiques importantes, et facilitée par la disparition d'une barrière géographique, le désert du Sahara, qui limitait son aire de distribution. Si certains auteurs parlent d'invasions naturelles pour décrire ces scénarios (Cristescu 2015; Facon *et al.* 2006; Williamson 1996), d'autres considèrent ces phénomènes comme des expansions géographiques locales naturelles (e.g. Estoup & Guillemaud 2010) où les individus en bordure de répartition migrent graduellement et colonisent les milieux adjacents en fonction de leurs exigences écologiques et de la disponibilité des habitats favorables. Toutefois, toutes les populations n'auront pas nécessairement le potentiel adaptatif et/ou les traits phénotypiques pour s'acclimater au nouvel environnement. De même, si certaines populations parviennent à s'installer, elles ne seront pas toutes capables de proliférer et de se répandre géographiquement pour coloniser d'autres milieux.

Les populations de *C. imicola* sont largement répandues et présentent des abondances importantes voire élevées (e.g. Acevedo *et al.* 2010; Ippoliti *et al.* 2013; Venail *et al.* 2012), suggérant qu'une fois arrivées dans le nouvel environnement, les populations auraient proliféré et se seraient répandues avec succès. Ainsi, ces populations auraient réussi à faire face aux contraintes démographiques, écologiques et évolutives qu'impose la colonisation d'un nouveau milieu. Notamment, si les conditions abiotiques et biotiques du nouvel environnement sont différentes de l'aire native – ce qui est fortement probable en cas d'une séparation ancienne – des contraintes adaptatives fortes ont pu être présentes (Colautti & Lau 2015). Les travaux menés au cours de cette thèse ne visaient pas à caractériser les traits adaptatifs de *C. imicola,* néanmoins compte tenu des différences de climats entre la région méditerranéenne et la région subsaharienne, on peut faire l'hypothèse que les populations colonisatrices ont subi des adaptations évolutives ou des traits plastiques qui leur ont permis de s'installer durablement. D'autre part, des évènements fondateurs peuvent conduire à de fortes réductions de la diversité génétique limitant la capacité des populations à s'établir (Sakai *et al.* 2001). La signature d'un goulot d'étranglement est détectée dans les populations

de *C. imicola* installées en Afrique du Nord, qui présentent une forte réduction de la diversité par rapport aux populations subsahariennes et des signatures d'expansion démographique. Cependant, il est à noter que cette réduction de la diversité génétique a pu être amplifiée par les variations climatiques de la fin du Pléistocène ou de l'Holocène.

Toutes ces contraintes sont communes aux invasions biologiques et aux expansions d'aire de distribution à l'échelle régionale. La remontée des populations de *C. imicola* vers la côte sud du bassin méditerranéen a probablement impliqué la migration graduelle et diffuse d'individus à la limite de leur distribution suite à la disparition d'une barrière physique. Toutefois, l'espèce est aussi capable de se disperser passivement sur de très longues distances. Par conséquent, cette colonisation pourrait également avoir impliqué la dispersion d'individus depuis des populations sources centrales. Ce dernier point couplé à la large répartition et l'abondance des populations de l'espèce dans les régions colonisées, soutiennent une grande capacité des populations de *C. imicola* à coloniser de nouveaux milieux. Néanmoins, son statut invasif reste à clarifier selon la définition donnée à une espèce native et à une espèce envahissante. Par exemple, si on se réfère à la définition de « natif » et « non-natif » de Manchester et Bullock (2000) - une espèce native étant un taxon indigène, endémique ou ayant connu une invasion préhistorique (avant la période Néolithique) - *C. imicola* pourrait-être considérée comme native en Afrique subsaharienne et sur la côte sud du bassin méditerranéen.

III.2. Une invasion naturelle facilitée par la dispersion naturelle à longue distance

Si le statut invasif des populations de *C. imicola* dans les régions de la côte sud du bassin méditerranéen reste encore à définir, la colonisation massive des régions déconnectées de l'aire d'origine sur une échelle de temps estimée à quelques siècles, justifie le statut invasif de l'espèce au sud de l'Europe. Cette invasion se caractérise par la présence étendue et abondante de populations dans la quasi-totalité des régions méditerranéennes (Acevedo *et al.* 2010; Goffredo *et al.* 2003; Ippoliti *et al.* 2013; Venail *et al.* 2012), avec des flux de gènes récurrents entre les populations.

Les invasions biologiques des zones géographiquement distantes et déconnectées de l'aire native (dispersion inter- et transcontinentales ou transocéaniques) sont souvent associées à une dispersion anthropique (Ricciardi 2007; Wilson *et al.* 2009). Bien qu'on reconnaisse le rôle de la dispersion naturelle à longue distance, ce mode de dispersion est

considéré comme rare (Ricciardi 2007; Wilson *et al.* 2009) et impliquant l'introduction d'une faible quantité de variabilité génétique sur des échelles de temps très importantes (Cristescu 2015; Estoup & Guillemaud 2010; Wilson *et al.* 2009). Par conséquent, la probabilité du succès d'établissement d'une population envahissante introduite naturellement est considérée comme plus faible que celle associée à une dispersion anthropique (Cassey *et al.* 2005).

Cependant, nos résultats montrent que lorsqu'un environnement est favorable à l'installation d'une espèce (suite à un changement environnemental ou à un changement évolutif de l'espèce) (Facon *et al.* 2006) et que la capacité de dispersion de cette espèce est importante, la dispersion naturelle à longue distance peut accélérer l'invasion biologique la rendre possible sur une échelle de temps relativement courte (moins d'un millénaire selon nos inférences ABC), par rapport à ce qui est attendu lors des invasions naturelles (plusieurs millénaires) (Cassey *et al.* 2005; Estoup & Guillemaud 2010; Wilson *et al.* 2009). Notre étude suggère par ailleurs, que *C. imicola* se disperse sur de longues distances et que des flux géniques sont possibles entre les populations sources et les populations envahissantes. Ceci induit l'introduction répétée de migrants dans les zones colonisées, ce qui augmente la probabilité de migration de génotypes différents et garantit un maintien du niveau de diversité au sein des populations envahissantes. Ce processus peut permettre aux populations introduites d'être moins impactées par la stochasticité environnementale (par exemple les variations climatiques) (Lockwood *et al.* 2005) et donc, peut augmenter la probabilité du succès d'établissement d'une espèce introduite naturellement dans un nouvel environnement. Ainsi, contrairement à ce qui est généralement admis, notre étude révèle que la dispersion naturelle à longue distance n'est pas un évènement rare et peut permettre la colonisation de milieux géographiquement éloignés de l'aire d'origine. Un exemple similaire est présenté par Alsos *et al.* (2007) qui montrent que des évènements fréquents de dispersion naturelle passive ont permis le maintien de la diversité génétique des populations envahissantes de neuf espèces de plantes de l'archipel Arctique.

III.3. Les facteurs clefs du succès d'invasion de *C. imicola* : quelles perspectives ?

La forte capacité de *C. imicola* à se disperser par les vents sur de longues distances apparaît comme un trait fondamental du succès de son invasion dans le bassin méditerranéen. Reynolds *et al.* (2006) suggèrent que ce mode de dispersion est un processus semi-passif, dans lequel les individus transportés participent activement à la dispersion en initiant et en maintenant leur transport par du vol actif. Une étude plus poussée de ce mécanisme visant à

caractériser ce mode de dispersion, à quantifier sa fréquence et à estimer la vitesse de dispersion permettrait de (i) comprendre le rôle de la dispersion dans les phénomènes d'expansion géographique et (ii) d'anticiper l'introduction d'individus dans de nouvelles zones.

Outre la capacité de dispersion, d'autres traits d'histoire de vie et facteurs écologiques, évolutifs et démographiques ont certainement été à l'œuvre lors de la colonisation du bassin méditerranéen.

La bio-écologie de *C. imicola* sur ses aires natives et colonisées est encore mal connue. Les habitats larvaires préférentiels de cette espèce sont décrits comme des zones humides, chaudes (entre 12.6 et 32 °C, avec un optimum entre 18 et 29 °C), et riches en matière organique (Braverman *et al.* 1974; Mellor & Prrzous 1979). L'étude des traits d'histoire de vie des adultes *C. imicola* au laboratoire suggère que les températures comprises entre 21 et 24 °C sont optimales à son cycle de développement (Veronesi *et al.* 2009). Ainsi, la distribution de cette espèce dépend grandement des facteurs environnementaux dont la température, l'humidité, le type de sol et la végétation.

Lors de la colonisation du bassin méditerranéen, les populations envahissantes de *C. imicola* ont certainement été confrontées à un climat différent, notamment avec des saisons hivernales marquées, où les températures extérieures pour les adultes sont largement en-dessous des valeurs optimales, et avec des périodes de gels des sols. Par conséquent, la question d'une adaptation génétique et/ou d'une modification phénotypique de l'espèce se pose. Il serait donc intéressant d'approfondir les connaissances apportées par cette thèse en étudiant les mécanismes d'adaptation qui ont pu être à l'œuvre lors de l'invasion des populations de *C. imicola*. Dans l'aire native, il est démontré que la dynamique des adultes est continue au cours de l'année (Diarra *et al.* 2014), alors que cette dynamique est interrompue en Corse lors des périodes froides entre décembre et mars (Venail *et al.* 2012). Durant ces périodes, les adultes disparaissent et seuls les stades immatures persistent jusqu'à ce que les températures redeviennent clémentes, suggérant une adaptation des populations aux périodes hivernales. Il est également intéressant de noter que des différences morphologiques des patrons alaires ont été observées entre les populations afrotropicales et méditerranéennes lors de l'identification des échantillons de *C. imicola* au cours de cette thèse. Ceci renforce l'intérêt d'explorer le rôle potentiel de la sélection naturelle ou de la plasticité phénotypique dans le succès invasif des populations de *C. imicola*.

Une approche comparative intra-spécifique, entre populations natives et envahissantes, pourrait permettre de mieux comprendre la contribution des traits d'histoire de vie à la valeur sélective. Ainsi, des études expérimentales comparatives de génétique quantitative entre les populations méditerranéennes et les populations sources dont elles dérivent généalogiquement, permettraient de tester et mesurer les facteurs clefs qui ont pu évoluer en réponse de la sélection naturelle durant l'invasion (Facon *et al.* 2006).

IV. Gestion et lutte : une issue possible ?

Définir des stratégies de lutte anti-vectorielle adéquates et efficaces nécessite de bien connaître la biologie et l'écologie des espèces impliquées dans la transmission. Dans un contexte d'invasion biologique, les méthodes préventives visant à limiter l'introduction des espèces envahissantes sont décrites comme le moyen le plus efficace et le moins coûteux (Simberloff *et al.* 2013). L'identification des populations sources et la description des routes de colonisation sont alors des informations fondamentales à la mise en place de stratégie de prévention (Estoup & Guillemaud 2010). Lorsque la prévention échoue, des programmes de gestion des populations, basés sur la régulation de l'abondance des populations envahissantes via des moyens de lutte chimiques ou biologiques sont une dernière option (Simberloff *et al.* 2013). Une connaissance de la structuration des populations et des mécanismes de dispersion des espèces est alors nécessaire. Ces connaissances sont très souvent difficiles à recueillir par des observations directes (capture-marquage-recapture) et nécessitent bien souvent une approche indirecte par l'étude de la génétique des populations. Par exemple, des préconisations de lutte contre les glossines, mouches piqueuses responsables de la transmission de trypanosomes humains ou animaux ont pu être émises suite à des études de la structuration génétique des populations (e.g. Bouyer *et al.* 2009; Solano *et al.* 2010)

L'ensemble de nos résultats montre une faible différentiation génétique des populations aussi bien dans l'aire native que dans l'aire envahie de *C. imicola*. Ce niveau de différenciation est dû à la forte capacité de l'espèce à se disperser sur de longues distances via les vents, à l'échelle régionale et au-dessus de masses d'eau. Par ailleurs, la simulation mathématique de cette capacité de dispersion par les vents suggère que ces évènements peuvent être fréquents dans certaines zones et qu'ils contribuent à l'expansion régionale de l'espèce dans le bassin méditerranéen et à la dynamique des populations installées. Ce potentiel de dispersion constitue ainsi un risque majeur pour les pays déclarés indemnes des

virus transmis par *C. imicola,* où l'arrivée d'individus infectés provenant de zones avec des foyers importants pourrait conduire à l'introduction d'un nouveau sérotype ou d'un virus exotique et conduire à des émergences.

Ainsi, dans un contexte où la dispersion de l'espèce n'est pas limitée, des actions ciblées de lutte anti-vectorielle au niveau local ne permettraient pas un contrôle efficace des populations. Si l'idée d'une élimination complète des populations de *Culicoides* d'un environnement est improbable (Sperlova & Zendulkova, 2011), il est cependant possible de diminuer le contact hôte-vecteur en protégeant les hôtes des piqûres des moucherons (barrière physique comme les moustiquaires) et de réduire très localement les populations avec l'utilisation d'insecticides agréés tels que pyréthrinoïdes synthétiques (Carpenter *et al.* 2008, Venail *et al*. 2011). Les contraintes logistiques et financières de telles mesures dans le contexte d'élevage en Europe du Nord, limitent leurs utilisations aux animaux de très haute valeur économique et en cas de situation d'épizootique sans vaccination. De plus, l'utilisation continue d'insecticides peut constituer une pression de sélection sur les populations de vecteurs et peut conduire à l'apparition de génotypes résistants aux molécules insecticides.

Conclusion générale

Les phénomènes de colonisation et d'invasions biologiques concernent de nombreuses espèces animales et végétales dont le nombre a considérablement augmenté ces dernières années. Parmi les espèces concernées, le moucheron hématophage *C. imicola* a largement étendu son aire de répartition et s'est installé avec succès dans des régions géographiquement éloignées et déconnectées de son aire d'origine. L'ensemble de ce travail de thèse a permis une caractérisation des populations sources et des routes de colonisation empruntées par les populations colonisatrices pour étendre l'aire de répartition de l'espèce. Alors que la présence de cette espèce dans le bassin méditerranéen était décrite comme récente et liée au réchauffement climatique contemporain, nous avons montré que les populations méditerranéennes de l'espèce étaient installées depuis des décennies voire des siècles avant leur découverte. L'histoire de la colonisation de *C. imicola* est une excellente illustration des facteurs démographiques et évolutifs impliqués dans les processus de colonisation. Parmi ces facteurs, le rôle fondamental de la dispersion naturelle dans les phénomènes d'invasions est mis en lumière dans cette thèse. Toutefois, les résultats obtenus soulèvent de nombreuses autres questions. Notamment, quels sont les mécanismes adaptatifs à l'origine du succès invasif des populations de l'espèce ? Quel est le rôle de la plasticité phénotypique ? Quels sont les effets de cette installation dans les écosystèmes méditerranéens ?

Des études complémentaires adressant ces questions permettront une meilleure compréhension des facteurs sous-jacents au succès invasif de *C. imicola*, et apporteront des connaissances fondamentales sur les processus d'invasions naturelles. De plus, une étude de l'écologie du paysage permettra d'appréhender les facteurs environnementaux et le rôle du paysage dans la distribution, la dynamique et la diversité des populations de *C. imicola* dans le bassin méditerranéen.

Références

Abbott RJ, James JK, Milne RI, Gillies ACM (2003) Plant introductions, hybridization and gene flow. *Philosophical Transactions of the Royal Society B: Biological Sciences* **358**, 1123-1132.

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.

Alsos IG, Eidesen PB, Ehrich D, *et al.* (2007) Frequent long-distance plant colonization in the changing Arctic. *Science* **316**, 1606-1609.

Anderson E, Stebbins G (1954) Hybridization as an evolutionary stimulus. *Evolution* **8**, 378-388.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Havard university press, Cambridge, United Kingdom.

Ayres DR, Garcia-Rossi D, Davis HG, Strong DR (1999) Extent and degree of hybridization between exotic (*Spartina alterniflora*) and native (*S. foliosa*) cordgrass (Poaceae) in California, USA determined by random amplified polymorphic DNA (RAPDs). *Molecular Ecology* **8**, 1179-1186.

Bailly-Choumara H, Kremer M (1970) Epizooties de 1989, 1990 et 1991. Données épidemiologiques et stratégie de lutte. *Report of the Direction de l'Elevage. Ministère de l'Agriculture et de la Réforme Agraire, Rabat, Maroc.*

Ballard J, Melvin R (2010) Linking the mitochondrial genotype to the organismal phenotype. *Molecular Ecology* **19**, 1523-1539.

Baylis M, Githeko AK (2006) The effects of climate change on infectious diseases of animals. *Report for the Foresight Project on Detection of Infectious Diseases, Department of Trade and Industry, UK Government*.

Bazin E, Glémin S, Galtier N (2006) Population size does not influence mitochondrial genetic diversity in animals. *Science* **312**, 570-572.

Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics* **41**, 379-406.

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.

Berkvens N, Moens J, Berkvens D, *et al.* (2010) Dinocampus coccinellae as a parasitoid of the invasive ladybird Harmonia axyridis in Europe. *Biological control* **53**, 92-99.

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.

Blackburn TM, Lockwood JL, Cassey P (2015) The influence of numbers on invasion success. *Molecular Ecology* **24**, 1942-1953.

Blair AC, Hufbauer RA (2010) Hybridization and invasion: one of North America's most devastating invasive plants shows evidence for a history of interspecific hybridization. *Evolutionary Applications* **3**, 40-51.

Blum MG, François O (2010) Non-linear regression models for Approximate Bayesian Computation. *Statistics and Computing* **20**, 63-73.

Bock DG, Caseys C, Cousens RD, *et al.* (2015) What we still don't know about invasion genetics. *Molecular Ecology* **24**, 2277-2297.

Bohrer G, Nathan R, Volis S (2005) Effects of long-distance dispersal for metapopulation survival and genetic structure at ecological time and spatial scales. *Journal of Ecology* **93**, 1029-1040.

Bonizzoni M, Gasperi G, Chen X, James AA (2013) The invasive mosquito species *Aedes albopictus*: current knowledge and future perspectives. *Trends in Parasitology* **29**, 460-468.

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.

Born W, Rauschmayer F, Bräuer I (2005) Economic evaluation of biological invasions—a survey. *Ecological Economics* **55**, 321-336.

Bouyer J, Balenghien T, Ravel S, *et al.* (2009) Population sizes and dispersal pattern of tsetse flies: rolling on the river? *Molecular Ecology* **18**, 2787-2797.

Bowman DD (2014) *Georgis' parasitology for veterinarians.*Saunders Elsevier Health Sciences, Missouri, United States of America.

Braverman Y (1991) The possible introduction to Israel of *Culicoides* (Diptera, Ceratopogonidae) borne animal diseases by wind. In: *Bluetongue, African horse sickness and related orbiviruses* (TE Walton & BI Osburn, eds). pp. 17-21. Proc. Second International Symposium, Paris.

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, Galun R, Ziv M (1974) Breeding sites of some *Culicoides* species (Diptera, Ceratopogonidae) in Israel. *Mosquito News* **34**, 303-308.

Brenner RJ, Wargo MJ, Stains GS, Mulla MS (1984) The dispersal of *Culicoides mohave* (Diptera: Ceratopogonidae) in the desert of Southern California. *Mosquito News* **44**, 343-350.

Brito PH (2007) Contrasting patterns of mitochondrial and microsatellite genetic structure among Western European populations of tawny owls (*Strix aluco*). *Molecular Ecology* **16**, 3423-3437.

Brouat C, Tollenaere C, Estoup A, *et al.* (2014) Invasion genetics of a human commensal rodent: the black rat *Rattus rattus* in Madagascar. *Molecular Ecology* **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. *Bulletin of Entomological Research* **99**, 583-591.

Campano Lopez A, Sanchez Botija C (1958) L'épizootie de fièvre catarrhale ovine en Espagne (bluetongue). *Bulletin de l'Office International des Epizooties* **50**, 65-93.

Carlson SM, Cunningham CJ, Westley PA (2014) Evolutionary rescue in a changing world. *Trends in Ecology & Evolution* **29**, 521-530.

Carlton JT (1996) Biological invasions and cryptogenic species. *Ecology* **77**, 1653-1655.

Carpenter S, Mellor P, Torr S (2008) Control techniques for *Culicoides* biting midges and their application in the UK and northwestern Palaearctic. *Medical and Veterinary Entomology* **22**, 175-187.

Cassey P, Blackburn TM, Duncan RP, Chown SL (2005) Concerning invasive species: reply to Brown and Sax. *Austral Ecology* **30**, 475-480.

Castaneda IS, Mulitza S, Schefuss E, *et al.* (2009) Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. *Proceedings of the National Academy of Sciences of the United Statesof America* **106**, 20159-20163.

Caswell H, Lensink R, Neubert MG (2003) Demography and dispersal: life table response experiments for invasion speed. *Ecology* **84**, 1968-1978.

Chen I-C, Hill JK, Ohlemüller R, Roy DB, Thomas CD (2011) Rapid range shifts of species associated with high levels of climate warming. *Science* **333**, 1024-1026.

Chernin E (1983) Sir Patrick Manson's studies on the transmission and biology of filariasis. *Review of Infectious Diseases* **5**, 148-166.

Cohen AN, Carlton JT (1998) Accelerating invasion rate in a highly invaded estuary. *Science* **279**, 555-558.

Colautti RI, Lau JA (2015) Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Molecular Ecology* **24**, 1999-2017.

Colautti RI, MacIsaac HJ (2004) A neutral terminology to define 'invasive' species. *Diversity and Distributions* **10**, 135-141.

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.

Courchamp F, Chapuis J-L, Pascal M (2003) Mammal invaders on islands: impact, control and control impact. *Biological Reviews* **78**, 347-383.

Courchamp F, Clutton-Brock T, Grenfell B (1999) Inverse density dependence and the Allee effect. *Trends in Ecology & Evolution* **14**, 405-410.

Cristescu ME (2015) Genetic reconstructions of invasion history. *Molecular Ecology* **24**, 2212-2225.

Csillery K, Blum MG, Gaggiotti OE, Francois O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**, 410-418.

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. *Medical and Veterinary Entomology* **17**, 379-387.

Davey JW, Hohenlohe PA, Etter PD, *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**, 499-510.

Davis MA, Thompson K (2000) Eight ways to be a colonizer; two ways to be an invader: a proposed nomenclature scheme for invasion ecology. *Bulletin of the ecological society of America* **81**, 226-230.

Delécolle 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.

Diarra M, Fall M, Fall AG, *et al.* (2014) Seasonal dynamics of *Culicoides* (Diptera: Ceratopogonidae) biting midges, potential vectors of African horse sickness and bluetongue viruses in the Niayes area of Senegal. *Parasites & Vectors* **7**, 147.

Diaz Montilla R, Panos Marti P (1967) Epizootologia de la peste equina en Espana. *Bulletin Office International des Epizooties* **86**, 705-714.

Diaz Montilla R, Panos Marti P (1968) La peste equina. *Bulletin Office International des Epizooties* **70**, 647-662.

Diouf N, Etter E, Lo M, Lo M, Akakpo A (2013) Outbreaks of African horse sickness in Senegal, and methods of control of the 2007 epidemic. *Veterinary Record* **172**, 152-152.

Dlugosch KM, Lai Z, Bonin A, Hierro J, Rieseberg LH (2013) Allele identification for transcriptome-based population genomics in the invasive plant *Centaurea solstitialis*. *G3: Genes Genomes Genetics* **3**, 359-367.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* **17**, 431-449.

Du Toit R (1944) The transmission of bluetongue and horsesickness by *Culicoides*. *Onderstepoort Journal of veterinary Science and animal Industry* **19**, 7-16.

Duncan RP, Blackburn TM, Rossinelli S, Bacher S (2014) Quantifying invasion risk: the relationship between establishment probability and founding population size. *Methods in Ecology and Evolution* **5**, 1255-1263.

Dyce A, Wirth W (1983) Reappraisal of some Indian *Culicoides* species in the subgenus Avaritia (Diptera: Ceratopogonidae). *International Journal of Entomological Research* **25**, 221-225.

Edwards S, Bensch S (2009) Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology* **18**, 2930-2933.

Ehrenfeld JG (2010) Ecosystem consequences of biological invasions. *Annual Review of Ecology Evolution and Systematics* **41**, 59-80.

Eilenberg J, Hajek A, Lomer C (2001) Suggestions for unifying the terminology in biological control. *BioControl* **46**, 387-400.

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology* **19**, 4113-4130.

Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. *Annual Review of Ecology Evolution and Systematics* **40**, 481-501.

Facon B, Genton BJ, Shykoff J, *et al.* (2006) A general eco-evolutionary framework for understanding bioinvasions. *Trends in Ecology & Evolution* **21**, 130-135.

Fontenille D, Lagneau C, Lecollinet S, *et al.* (2009) *La lutte antivectorielle en France*. IRD éditions, collection Expertise collégiale Marseille, France.

François O, Blum MG, Jakobsson M, Rosenberg NA (2008) Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genetics* **4**, e1000075.

Gage KL, Burkot TR, Eisen RJ, Hayes EB (2008) Climate and vectorborne diseases. *American Journal of Preventice Medicine* **35**, 436-450.

Galil B (2007) Loss or gain? Invasive aliens and biodiversity in the Mediterranean Sea. *Marine Pollution Bulletin* **55**, 314-322.

Gambles RM (1949) Bluetongue of Sheep in Cyprus. *Journal of Comparative Pathology and Therapeutics* **59**, 176-190.

García-Marín J-L, Utter FM, Pla C (1999) Postglacial colonization of brown trout in Europe based on distribution of allozyme variants. *Heredity* **82**, 46-56.

Gillespie RG, Baldwin BG, Waters JM, *et al.* (2012) Long-distance dispersal: a framework for hypothesis testing. *Trends in Ecology & Evolution* **27**, 47-56.

Goffredo M, Conte A, Cocciolito R, Meiswinkel R (2003) The distribution and abundance of Culicoides imicola in Italy. *Veterinaria Italiana* **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. *Veterinaria Italiana* **40**, 270-273.

Goffredo M, Meiswinkel R (2004) Entomological surveillance of bluetongue in Italy: methods of capture, catch analysis and identification of *Culicoides* biting midges. *Veterinaria Italiana* **40**, 260-265.

Gray ME, Sappington TW, Miller NJ, Moeser J, Bohn MO (2009) Adaptation and invasiveness of western corn rootworm: intensifying research on a worsening pest. *Annual Review of Entomology* **54**, 303-321.

Green AJ, Figuerola J (2005) Recent advances in the study of long-distance dispersal of aquatic invertebrates via birds. *Diversity and Distributions* **11**, 149-156.

Guichard S, Guis H, Tran A, *et al.* (2014) Worldwide niche and future potential distribution of *Culicoides imicola*, a major vector of bluetongue and African horse sickness viruses. *PLoS One* **9**, e112491.

Guillemaud T, Ciosi M, Lombaert E, Estoup A (2011) Biological invasions in agricultural settings: Insights from evolutionary biology and population genetics. *Comptes rendus biologies* **334**, 237-246.

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. *Journal of the Royal Society Interface* **9**, 339-350.

Handley L-JL, Estoup A, Evans D, *et al.* (2011) Ecological genetics of invasive alien species. *BioControl* **56**, 409-428.

Hendry G, Godwin G (1988) Biting midges in Scottish forestry: a costly irritant or a trivial nuisance. *Scottish Forestry* **42**, 113-119.

Hewitt GM (2004) The structure of biodiversity - insights from molecular phylogeography. *Frontiers in Zoology* **1**, 4.

Hill DS (2012) *The economic importance of insects*. Springer Science & Business Media, London, United Kingdom.

Hoddle MS (2004) Restoring balance: using exotic species to control invasive exotic species. *Conservation Biology* **18**, 38-49.

Holmes TP, Aukema JE, Von Holle B, Liebhold A, Sills E (2009) Economic impacts of invasive species in forests. *Annals of the New York Academy of Science* **1162**, 18-38.

Hufbauer RA, Facon B, Ravigne V, *et al.* (2012) Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evolutionary Applications* **5**, 89-101.

Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology* **46**, 10-18.

Hulme PE (2014) Invasive species challenge the global response to emerging diseases. *Trends in Parasitology* **30**, 267-270.

Hulme PE, Bacher S, Kenis M, *et al.* (2008) Grasping at the routes of biological invasions: a framework for integrating pathways into policy. *Journal of Applied Ecology* **45**, 403-414.

Iqbal Z, Caccamo M, Turner I, Flicek P, McVean G (2012) De novo assembly and genotyping of variants using colored de Bruijn graphs. *Nature genetics* **44**, 226-232.

Ippoliti C, Gilbert M, Vanhuysse S, *et al.* (2013) Can landscape metrics help determine the *Culicoides imicola* distribution in Italy? *Geospatial Health* **8**, 267-277.

Irwin DE (2012) Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. *American Naturalist* **180**, 35-49.

Johnson JA, Toepfer JE, Dunn PO (2003) Contrasting patterns of mitochondrial and microsatellite population structure in fragmented populations of greater prairie-chickens. *Molecular Ecology* **12**, 3335-3347.

Jones AR, Thomson DJ, Hort M, Devenish B (2007) The U.K. Met Office's next-generation atmospheric dispersion model, NAME III. In: *Air Pollution Modeling and its Application XVII* (Borrego C. and Norman A.-L eds). pp. 580-589. Springer, New York, United States of America.

Joyce P, Marjoram P (2008) Approximately sufficient statistics and Bayesian computation. *Statistical Applications in Genetics and Molecular Biology* **7,** 1.

Kampen H, Werner D (2011) Arthropod vectors and their growing importance in Europe. In: *Progress in Parasitology* (H. Mehlhorn ed). pp. 259-282. Springer-Verlag, Berlin, Germany.

Keane RM, Crawley MJ (2002) Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology & Evolution* **17**, 164-170.

Kilpatrick AM, Randolph SE (2012) Drivers, dynamics, and control of emerging vector-borne zoonotic diseases. *The Lancet* **380**, 1946-1955.

Kirkeby C, Bodker R, Stockmarr A, Lind P, Heegaard PM (2013) Quantifying dispersal of European *Culicoides* (Diptera: Ceratopogonidae) vectors between farms using a novel mark-release-recapture technique. *PLoS One* **8**, e61269.

Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution* **23**, 482-490.

Kluiters G, Swales H, Baylis M (2015) Local dispersal of Palaearctic *Culicoides* biting midges estimated by mark-release-recapture. *Parasites & Vectors* **8**, 54.

Knop E, Reusser N (2012) Jack-of-all-trades: phenotypic plasticity facilitates the invasion of an alien slug species. *Proceedings of the Royal Society of London B: Biological Sciences*, **279**, 4668-4676.

Knowles LL (2009) Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics* **40**, 593-612.

Kokko H, López-Sepulcre A (2006) From individual dispersal to species ranges: perspectives for a changing world. *Science* **313**, 789-791.

Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution* **16**, 199-204.

Kolleck J, Yang M, Zinner D, Roos C (2013) Genetic diversity in endangered guizhou snub-nosed monkeys (*Rhinopithecus brelichi*): Contrasting results from microsatellite and mitochondrial DNA data. *Plos One* **8**, e73647

Kovats R, Campbell-Lendrum D, McMichel A, Woodward A, Cox JSH (2001) Early effects of climate change: do they include changes in vector-borne disease? *Philosophical Transactions of the Royal Society B: Biological Sciences* **356**, 1057-1068.

Kulhanek SA, Ricciardi A, Leung B (2011) Is invasion history a useful tool for predicting the impacts of the world's worst aquatic invasive species? *Ecological Applications* **21**, 189-202.

Laaberki A (1969) Évolution d'une épizootie de peste équine africaine au Maroc. *Bulletin Office International des Epizooties* **71**, 921-936.

Lee CE (2002) Evolutionary genetics of invasive species. *Trends in Ecology & Evolution* **17**, 386-391.

Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution* **20**, 223-228.

Lockwood JL, Cassey P, Blackburn TM (2009) The more you introduce the more you get: the role of colonization pressure and propagule pressure in invasion ecology. *Diversity and Distributions* **15**, 904-910.

Lockwood JL, Hoopes MF, Marchetti MP (2013) *Invasion ecology*. John Wiley & Sons, Malden, United States of America.

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, 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*. *Molecular Ecology* **20**, 4654-4670.

Lomer C, Bateman R, Johnson D, Langewald J, Thomas M (2001) Biological control of locusts and grasshoppers. *Annual Review of Entomology* **46**, 667-702.

Lopes J, Beaumont M (2010) ABC: a useful Bayesian tool for the analysis of population data. *Infection, Genetics and Evolution* **10**, 825-832.

Lubroth J (1988) African Horsesickness and the epizootic in Spain, 1987. *Equine Practice* **10**, 26-31.

Mack RNS, D.; Lonsdale, W.M.; Evans, H.; Clout, M.; Bazzaz, F.A (2000) Biotic invasions: causes, epidemiology, global consequences and control. *Ecological Applications* **10**, 689-710.

Manchester SJ, Bullock JM (2000) The impacts of non-native species on UK biodiversity and the effectiveness of control. *Journal of Applied Ecology* **37**, 845-864.

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. *Molecular Ecology* **22**, 2456-2466.

Matisoo-Smith E, Roberts R, Irwin GJ, *et al.* (1998) Patterns of prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat. *Proceedings of the National Academy of Sciences of the United State of America* **95**, 15145-15150.

Mazza G, Tricarico E, Genovesi P, Gherardi F (2014) Biological invaders are threats to human health: an overview. *Ethology Ecology & Evolution* **26**, 112-129.

McMichael AJ, Bouma MJ, Mooney H (2000) Global changes, invasive species, and human health. In: *Invasive Species in a Changing World.* pp. 191-210. Island Press, Washington, United States of America.

McNeely JA (2006) As the world gets smaller, the chances of invasion grow. *Euphytica* **148**, 5-15.

Medlock JM, Hansford K M, Schaffne F, *et al.* (2012). A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. *Vector-borne and zoonotic diseases* **12**, 435-447.

Mellor P, Boned J, Hamblin C, Graham S (1990) Isolations of African horse sickness virus from vector insects made during the 1988 epizootic in Spain. *Epidemiology & Infection* **105**, 447-454.

Mellor P, Prrzous G (1979) Observations on breeding sites and light-trap collections of *Culicoides* during an outbreak of bluetongue in Cyprus. *Bulletin of Entomological Research* **69**, 229-234.

Mellor PS, Boorman J, Baylis M (2000) *Culicoides* biting midges: their role as arbovirus vectors. *Annual Review of Entomology* **45**, 307-340.

Mellor PS, Boorman J, Wilkinson PJ, Martinez-Gomez F (1983) Potential vectors of bluetongue and African horse sickness in Spain. *Veterinary Record* **112**, 229-230.

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. *The Veterinary Journal* **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. *Medical and Veterinary Entomology* **17**, 52-54.

Monecke S, Monecke H, Monecke J (2009) Modelling the black death. A historical case study and implications for the epidemiology of bubonic plague. *International Journal of Medical Microbiology* **299**, 582-593.

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**, 415-417.

Morand S, Bordes F, Chen HW, *et al.* (2015) Global parasite and *Rattus* rodent invasions: The consequences for rodent-borne diseases. *Integrative Zoology* **10**, 409-423.

Nathan R, Perry G, Cronin JT, Strand AE, Cain ML (2003) Methods for estimating long-distance dispersal. *Oikos* **103**, 261-273.

Nathan R, Schurr FM, Spiegel O, *et al.* (2008) Mechanisms of long-distance seed dispersal. *Trends in Ecology & Evolution* **23**, 638-647.

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1-10.

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. *Medical and Veterinary Entomology* **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.

Olden JD, Poff NL, Douglas MR, Douglas ME, Fausch KD (2004) Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology & Evolution* **19**, 18-24.

Pandit MK, White SM, Pocock MJ (2014) The contrasting effects of genome size, chromosome number and ploidy level on plant invasiveness: a global analysis. *New Phytologist* **203**, 697-703.

Panov VE, Piotr I, Riccardi N (2004) Role of diapause in dispersal and invasion success by aquatic invertebrates. *Journal of Limnology* **63**, 56-69.

Parry HR, Eagles D, Kriticos DJ, Venette R (2015) Simulation modelling of long-distance windborne dispersal for invasion ecology. In: *Pest Risk Modelling and Mapping for Invasive Alien Species* (R.C Venette ed). pp. 49-64. CABI, Croydon, United Kingdom.

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. *Molecular Ecology* **16**, 3069-3083.

Patakakis MJ (2004) *Culicoides imicola* in Greece. *Veterinaria Italiana* **40**, 232-234.

Patakakis MJ, Papazahariadou M, Wilson A, *et al.* (2009) Distribution of *Culicoides* in Greece. *Journal of Vector Ecology* **34**, 243-251.

Pavlova A, Amos JN, Joseph L, *et al.* (2013) Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an australian bird. *Evolution* **67**, 3412-3428.

Peery MZ, Kirby R, Reid BN, *et al.* (2012) Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology* **21**, 3403-3418.

Peischl S, Dupanloup I, Kirkpatrick M, Excoffier L (2013) On the accumulation of deleterious mutations during range expansions. *Molecular Ecology* **22**, 5972-5982.

Peterlongo P, Schnel N, Pisanti N, Sagot M-F, Lacroix V (2010) Identifying SNPs without a reference genome by comparing raw reads. In: *String Processing and Information Retrieval* (E. Chavez, S. Lonardi ed). pp.147-158. Springer Berlin Heidelberg, Germany.

Peterson AT, Vieglais DA (2001) Predicting species invasions using ecological niche modeling: new approaches from bioinformatics attack a pressing problem. *BioScience* **51**, 363-371.

Phillips BL, Brown GP, Greenlees M, Webb JK, Shine R (2007) Rapid expansion of the cane toad (*Bufo marinus*) invasion front in tropical Australia. *Austral Ecology* **32**, 169-176.

Pimentel D (2011) *Biological invasions: economic and environmental costs of alien plant, animal, and microbe species.* CRC Press, USA.

Pimentel D, McNair S, Janecka J, *et al.* (2001) Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture, Ecosystems & Environment* **84**, 1-20.

Prentis PJ, Wilson JR, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in invasive species. *Trends in Plant Science* **13**, 288-294.

Prugnolle F, De Meeûs T (2002) Inferring sex-biased dispersal from population genetic tools: a review. *Heredity* **88**, 161-165.

Purse B, Carpenter S, Venter G, Bellis G, Mullens B (2015) Bionomics of Temperate and Tropical *Culicoides* Midges: Knowledge Gaps and Consequences for Transmission of Culicoides-Borne Viruses. *Annual Review of Entomology* **60**, 373-392.

Purse BV, Mellor PS, Rogers DJ, *et al.* (2005) Climate change and the recent emergence of bluetongue in Europe. *Nature Reviews Microbiology* **3**, 171-181.

Puzey J, Vallejo-Marín M (2014) Genomics of invasion: diversity and selection in introduced populations of monkeyflowers (Mimulus guttatus). *Molecular Ecology* **23**, 4472-4485.

Rašić G, Schama R, Powell R, *et al.* (2015) Contrasting genetic structure between mitochondrial and nuclear markers in the dengue fever mosquito from Rio de Janeiro: implications for vector control. *Evolutionary Applications* **8**, 901-915.

Reynolds C, Miranda NA, Cumming GS (2015) The role of waterbirds in the dispersal of aquatic alien and invasive species. *Diversity and Distributions* **7**, 744-754.

Reynolds D, Chapman J, Harrington R (2006a) The migration of insect vectors of plant and animal viruses. *Advances in Virus Research* **67**, 453-517.

Rhodain F, Perez C (1985) *Précis d'entomologie médicale et vétérinaire : notions d'épidémiologie des maladies à vecteurs*. Maloine, Paris, France.

Ricciardi A (2007) Are modern biological invasions an unprecedented form of global change? *Conservation Biology* **21**, 329-336.

Richards CL, Bossdorf O, Muth NZ, Gurevitch J, Pigliucci M (2006) Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. *Ecology Letters* **9**, 981-993.

Richardson DM, Pyšek P, Carlton JT (2011) A compendium of essential concepts and terminology in invasion ecology. In: *Fifty years of invasion ecology: The legacy of Charles Elton* (DM Richardson ed). pp. 409-420. *Wiley-Blackwell*, Chichester, United Kingdom.

Richardson DM, Pyšek P, Rejmánek M, *et al.* (2000) Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distributions* **6**, 93-107.

Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology & Evolution* **22**, 454-464.

Roy HE, De Clercq P, Handley L-JL, *et al.* (2011) Alien arthropod predators and parasitoids: an ecological approach. *BioControl* **56**, 375-382.

Ruiz GM, Rawlings TK, Dobbs FC, *et al.* (2000) Global spread of microorganisms by ships. *Nature* **408**, 49-50.

Sakai AK, Allendorf FW, Holt JS, *et al.* (2001) The population biology of invasive species. *Annual Review of Ecology and Systematics* **32**, 305-332.

Sanders CJ, Carpenter S (2014) Assessment of an immunomarking technique for the study of dispersal of *Culicoides* biting midges. *Infection, Genetics and Evolution* **28**, 583-587.

Sappington TW, Siegfried BD, Guillemau T (2006) Coordinated *Diabrotica* genetics research: accelerating progress on an urgent insect pest problem. *American Entomologist* **52**, 90-97.

Sax DF, Stachowicz JJ, Brown JH, *et al.* (2007) Ecological and evolutionary insights from species invasions. *Trends in Ecology & Evolution* **22**, 465-471.

Sax DF, Stachowicz JJ, Gaines SD (2005) *Species invasions: insights into ecology, evolution and biogeography*. Sinauer Associates Incorporated, Sunderland, United States of America.

Scott JD, Fernando K, Banerjee SN, *et al.* (2001) Birds disperse ixodid (Acari: Ixodidae) and *Borrelia burgdorferi*-infected ticks in Canada. *Journal of Medical Entomology* **38**, 493-500.

Sellers R (1975) Bluetongue in Cyprus. *Australian veterinary journal* **51**, 198-203.

Sellers R (1992) Weather, *Culicoides*, and the distribution and spread of bluetongue and African horse sickness viruses. In: *Bluetongue, African horse sickness, and related orbiviruses. Proceedings of the Second International Symposium,* pp. 284-290. CRC press, Florida, United States of America.

Sellers R, Pedgley D, Tucker M (1977) Possible spread of African horse sickness on the wind. *Journal of Hygiene* **79**, 279-298.

Sellers RF, Pedgley DE, Tucker MR (1978) Possible windborne spread of bluetongue to Portugal, June-July 1956. *Journal of Hygiene* **81**, 189-196.

Sexton JP, McIntyre PJ, Angert AL, Rice KJ (2009) Evolution and ecology of species range limits. *Annual Review of Ecology Evolution and Systematics* **40**, 415-436.

Shine R, Brown GP, Phillips BL (2011) An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 5708-5711.

Simberloff D (2009) The role of propagule pressure in biological invasions. *Annual Review of Ecology, Evolution, and Systematics* **40**, 81-102.

Simberloff D, Martin JL, Genovesi P, *et al.* (2013) Impacts of biological invasions: what's what and the way forward. *Trends in Ecology & Evolution* **28**, 58-66.

Solano P, Kaba D, Ravel S, *et al.* (2010) Population genetics as a tool to select tsetse control strategies: suppression or eradication of *Glossina palpalis gambiensis* in the Niayes of Senegal. *PLoS Neglected Tropical Diseases* **4**, e692.

Sperlova A, Zendulkova D (2011) Bluetongue: a review. *Veterinary Medicine-Czech* **56**, 430-452.

Spooner DM, McLean K, Ramsay G, Waugh R, Bryan GJ (2005) A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14694-14699.

Tatem AJ, Hay SI (2007) Climatic similarity and biological exchange in the worldwide airline transportation network. *Proceedings of the Royal Society of London B: Biological Sciences* **274**, 1489-1496.

Tatem AJ, Rogers DJ, Hay S (2006) Global transport networks and infectious disease spread. *Advances in Parasitology* **62**, 293-343.

Taylor WP, Sellers RF, Gumm ID, Herniman KA, Owen L (1985) Bluetongue epidemiology in the Middle East. *Progress in Clinical Biological Research* **178**, 527-530.

Templeton AR (2010) Coherent and incoherent inference in phylogeography and human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 6376-6381.

Toews DP, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology* **21**, 3907-3930.

Velthuis A, Saatkamp H, Mourits M, De Koeijer A, Elbers A (2010) Financial consequences of the Dutch bluetongue serotype 8 epidemics of 2006 and 2007. *Preventive Veterinary Medicine* **93**, 294-304.

Velthuis AG, Mourits MC, Saatkamp HW, De Koeijer AA, Elbers AR (2011) Financial evaluation of different vaccination strategies for controlling the bluetongue virus serotype 8 epidemic in The Netherlands in 2008. *PLoS One* **6**, e19612.

Venail R. (2014) Sensibilité aux insecticides et évaluation des méthodes de lutte anti-vectorielle disponibles contre les *Culicoides* (Diptera : Ceratopoginidae) paléarctiques, vecteurs de virus émergents d'intérêt en santé animale. Thèse Cirad, Montpellier.

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. In: *Parasitology Research Monographs* (ed. Mehlhorn H), pp. 77-103. Springer-Verlag, Berlin, Germany.

Venail R, Mathieu B, Setier-Rio M-L, *et al.* (2011) Laboratory and field-based tests of deltamethrin insecticides against adult *Culicoides* biting midges. *Journal of Medical Entomology* **48**, 351-357.

Venter GJ, Wright IM, Del Rio R, Lucientes J, Miranda MA (2011) The susceptibility of *Culicoides imicola* and other South African livestock-associated *Culicoides* species to infection with bluetongue virus serotype 8. *Medical and Veterinary Entomology* **25**, 320-326.

Veronesi E, Venter G, Labuschagne K, Mellor P, Carpenter S (2009) Life-history parameters of *Culicoides* (*Avaritia*) *imicola* Kieffer in the laboratory at different rearing temperatures. *Veterinary Parasitology* **163**, 370-373.

Villemant C, Muller F, Haubois S, *et al.* (2011) Bilan des travaux (MNHN et IRBI) sur l'invasion en France de *Vespa velutina*, le frelon asiatique prédateur d'abeilles. *Proceedings of the Journée Scientifique Apicole–11 February*, 3-12.

Vitousek PM, D'antonio CM, Loope LL, Rejmanek M, Westbrooks R (1997) Introduced species: a significant component of human-caused global change. *New Zealand Journal of Ecology* **21**, 1-16.

Wares J, Hughes A, Grosberg R (2005) Mechanisms that drive evolutionary change: insights from species introductions and invasions. In: *Species invasions: insights into ecology, evolution, and biogeography*, pp. 229-257. Sinauer Associates Inc, Sunderland, United States of America.

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

Williamson M (1996) *Biological invasions.* Chapman & Hall, London, United Kingdom.

Williamson M, Fitter A (1996) The varying success of invaders. *Ecology* **77**, 1661-1666.

Wilson JR, Dormontt EE, Prentis PJ, Lowe AJ, Richardson DM (2009) Something in the way you move: dispersal pathways affect invasion success. *Trends in Ecology & Evolution* **24**, 136-144.

Wittmann EJ, Baylis M (2000) Climate change: effects on *Culicoides*-transmitted viruses and implications for the UK. *The Veterinary Journal* **160**, 107-117.

Yü I (2005) *Ceratopogonidae of China:(Insecta: Diptera)*. Military Medical Science Press, Beijing, China.

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12**, 563-584.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17**, 2107-2121.

Résumé

Les invasions biologiques constituent une source de préoccupation majeure du fait des conséquences écologiques, économiques et sanitaires dont elles sont responsables. Déterminer et comprendre les facteurs sous-jacents au succès invasif des espèces envahissantes permet de prédire de nouvelles invasions et de mettre en place des stratégies de contrôle. *Culicoides imicola* est un vecteur majeur d'*Orbivirus* d'intérêt vétérinaire incluant le virus de la fièvre catarrhale ovine (FCO). Suite à l'émergence de la FCO dans le bassin méditerranéen, les populations de *C. imicola* ont été découvertes dans des territoires où elles étaient considérées comme absentes, caractérisant alors cette présence comme la résultante d'une expansion récente de l'espèce. Cette thèse décrit un ensemble de travaux visant à comprendre l'histoire de la colonisation du bassin méditerranéen par *C. imicola*. L'utilisation d'une approche multi-marqueurs combinant des analyses de génétique de populations, des inférences basées sur la méthode *Approximate Bayesian Computation* (ABC) et la simulation mathématique de la dispersion atmosphérique de l'espèce, a permis (i) de déterminer l'origine des populations installées au Maghreb et au Moyen Orient et de décrire les routes de colonisation et la chronologie de ces évènements, (ii) de définir les caractéristiques démographiques, évolutives et temporelles de la colonisation du sud de l'Europe et (iii) de caractériser les principaux facteurs expliquant le succès d'expansion géographique des populations installées. Les principaux résultats de cette thèse permettent de proposer des hypothèses pour expliquer le succès de l'installation des populations de *C. imicola* dans le bassin méditerranéen.

Mots-clés : *Culicoides*, routes d'invasion, dispersion, génétique des populations, méthode ABC, microsatellites, gènes mitochondriaux.

Biological invasions and emerging infectious diseases: expansion and colonization of the Mediterranean basin by *Culicoides imicola* (Diptera: Ceratopogonidae), a biting midge vector species of Orbiviruses

Biological invasions are of major concern because of their environmental, economic and health consequences. Determining and understanding the factors underlying the invasion success of species allow predicting potential other biological invasions, and developing vector control strategies. *Culicoides imicola* is a major vector species of *Orbivirus*, including the bluetongue virus (BTV) which affects domestic ruminants. Following BT emergence in the Mediterranean basin, *C. imicola* populations were recorded in territories where the species was considered to be absent, and consequently was described as expanding its range expansion on a short period. This Phd work describes a set of studies aiming at understanding the colonization history of the Mediterranean basin by *C. imicola*. The use of a multi-loci approach combining population genetics analyses, *Approximate Bayesian Computation* (ABC) methods and mathematical simulations of the atmospheric dispersion of the species enabled to (i) determine the origin of the established populations in the Maghreb and the Middle-East and describe the routes of colonization and the chronology of such events, (ii) define the demographic, evolutionary and temporal characteristics of south-western Europe colonization and (iii) characterize the main factors explaining the successful range expansion of the established populations. The main results of this thesis allow suggesting hypotheses to explain the successful establishment of *C. imicola* populations in the Mediterranean basin.

Keywords : *Culicoides*, invasion routes, dispersal, population genetics, ABC methods, microsatellites, mitochondrial genes.

Discipline: Écologie, Évolution, Ressources Génétique, Paléontologie
Cirad, UMR Contrôle des Maladies Animales Exotiques et Emergentes
Campus International de Baillarguet, 34398 Montpellier Cedex 5, France