

Institut National de la Recherche Agronomique

THÈSE présentée par :

Julien HARAN

pour obtenir le grade de : **Docteur de l'université d'Orléans**

Discipline : Physiologie et biologie des organismes et des populations - interactions

**Dispersion des espèces impliquées dans une
association phorétique vecteur - pathogène
nouvellement formée: Le cas de *Monochamus
galloprovincialis*, vecteur natif d'un nématode invasif
en Europe (*Bursaphelenchus xylophilus*)**

Thèse soutenue à Orléans le 4 Décembre 2015

devant le jury composé de :

Jean-François ARNAUD *Rapporteur*

Professeur, Centre National de la Recherche Scientifique, Université de Lille 1

Thomas GUILLEMAUD *Rapporteur*

Directeur de Recherche, Institut National de la Recherche Agronomique (ISA)

François LIEUTIER *Examineur-président du jury*

Professeur, Université d'Orléans, Laboratoire de Biologie des Ligneux et des Grandes Cultures (LBLGC)

Lionel ROQUES *Examineur*

Chargé de Recherche, Institut National de la Recherche Agronomique (BioSp)

Edmundo SOUSA *Examineur*

Directeur de Recherche, Instituto Nacional de Investigação Agrária e Veterinária, Portugal (INIAV)

Géraldine ROUX *Co-encadrante de thèse*

Maître de conférences, Université d'Orléans / Institut National de la Recherche Agronomique (URZF)

Christelle ROBINET *Co-encadrante de thèse*

Chargé de Recherche, Institut National de la Recherche Agronomique (URZF)

Alain ROQUES *Directeur de thèse*

Directeur de Recherche, Institut National de la Recherche Agronomique (URZF)

Dispersion des espèces impliquées dans une association phorétique vecteur - pathogène nouvellement formée:

Le cas de *Monochamus galloprovincialis*, vecteur natif d'un nématode invasif en Europe
(*Bursaphelenchus xylophilus*)

Dispersal of species involved in a novel vector-pathogen phoretic association:

The case of *Monochamus galloprovincialis*, native vector of an invasive nematode in Europe
(*Bursaphelenchus xylophilus*)

*Vis comme si tu devais mourir demain,
Apprends comme si tu devais vivre toujours*

Ghandi

Cette thèse a été réalisée de 2012 à 2015 au sein de l'Unité de Recherche en Zoologie Forestière (URZF, UR633) de l'Institut National de la Recherche Agronomique (INRA, centre Val de Loire). Les travaux présentés dans ce manuscrit sont issus de financements du Ministère de l'Enseignement Supérieur et de la Recherche (bourse doctorale MESR), du projet Européen REPHRAME (KBBE.2010.1.4-09) et du programme COST STSM (FP1002-Pathway Evaluation and pest Risk Management In Transport, PERMIT).

RESUME

Les invasions biologiques se sont intensifiées au cours des dernières décennies en raison d'une accélération des échanges commerciaux. Ces invasions représentent une menace pour les écosystèmes et de nombreuses activités économiques, il est donc crucial de comprendre les mécanismes qui les sous-tendent afin de mieux prévoir et limiter leurs impacts. Les espèces introduites dans un nouvel environnement interagissent avec les espèces de l'écosystème receveur et les nouvelles interactions induites (prédation, parasitisme, mutualisme...) peuvent significativement contribuer au succès de l'invasion et à l'émergence de dégâts. C'est le cas des associations phorétiques nouvellement formées, qui peuvent apparaître entre un vecteur et un phoronte. De telles associations peuvent générer des dégâts considérables, en particulier lorsqu'elles combinent la virulence d'un pathogène et les aptitudes dispersives d'une espèce vectrice. Les associations phorétiques vecteurs-pathogènes ont montré plusieurs exemples d'impacts spectaculaires sur les écosystèmes forestiers avec de lourdes conséquences écologiques et économiques. L'étude des associations phorétiques nouvellement formées est complexe car elle requiert à la fois une connaissance des espèces endémiques et des espèces invasives impliquées, et surtout de l'effet synergique de leur nouvelle association. De ce fait, et malgré l'importance de ce phénomène, ces associations restent relativement mal connues, en particulier en ce qui concerne le potentiel dispersif des espèces qu'elles impliquent.

Dans cette thèse, j'aborde la question du potentiel dispersif d'espèces impliquées dans une association phorétique nouvellement formée. En particulier, je me focalise sur le cas de l'association entre un nématode ravageur des pins, le nématode du pin (*Bursaphelenchus xylophilus*), et son insecte vecteur endémique (*Monochamus galloprovincialis*). Cette association nouvellement formée est à l'origine de dégâts considérables en péninsule ibérique depuis sa découverte au Portugal en 1999, et son expansion spatiale représente un risque conséquent pour les pinèdes européennes. Dans un premier temps j'ai caractérisé la structure génétique du vecteur endémique à l'échelle de l'Europe et exploré les facteurs historiques et environnementaux responsables des patrons génétiques observés. L'objectif de cette première étude était d'identifier les facteurs pouvant affecter la dispersion du vecteur à large échelle. Dans un second temps, et à l'échelle de la péninsule ibérique, j'ai évalué la dispersion du vecteur dans un environnement hétérogène, en testant l'existence de corridors et de barrières aux flux de gènes par une approche de génétique spatialisée. Enfin, je me suis intéressé à la dispersion du vecteur et du nématode à travers la barrière pyrénéenne, en couplant à la fois une étude des flux de gènes et un modèle d'expansion intégrant l'effet de synergie de cette association nouvellement formée.

Les résultats obtenus au travers de ces différentes études ont permis de mettre en évidence chez *M. galloprovincialis* une structure génétique caractéristique des recolonisations postglaciaires des organismes rencontrés en Europe tempérée, et une distribution des lignées évolutives liée principalement au régime de précipitations au-delà des facteurs historiques. Ces résultats suggèrent également que ce coléoptère ne montre pas de spécialisation pour les essences hôtes de pins auxquelles il est associé. L'étude de génétique spatialisée réalisée au sein de la péninsule ibérique a permis de mettre en évidence l'existence de barrières à la dispersion de cette espèce, principalement au niveau des reliefs, où les températures sont les plus basses. Les résultats de cette étude montrent également que les fortes densités en pins représentent des barrières à la dispersion de cette espèce, et inversement, que les zones ouvertes avec des pins épars peuvent favoriser cette dispersion. Enfin, les études menées sur la chaîne pyrénéenne suggèrent que ce massif pourrait jouer un rôle de barrière, mais seulement partielle, à l'expansion du nématode du pin, celui-ci pouvant contourner le relief dans les zones de plaines à ses extrémités Est et Ouest. Les simulations montrent également un effet potentiellement fort des températures basses sur l'expansion du nématode du pin, et par conséquent un risque élevé de voir ces barrières être réduites ou disparaître sous l'effet du réchauffement climatique.

D'un point de vue fondamental, cette thèse a apporté une contribution significative à la connaissance de l'histoire évolutive du vecteur et aux paramètres qui influencent sa dispersion. Cette thèse a également conduit au développement d'une nouvelle méthode dans l'élaboration des dispositifs expérimentaux en analyses de génétique spatialisée. Cette approche permet de généraliser l'inférence faite dans ce type d'étude en se basant sur des réplifications d'échelles et de sites d'étude, et s'adapte

particulièrement aux espèces à fort pouvoir dispersif, tels que les insectes vecteurs. Le modèle d'expansion développé dans cette thèse a permis de simuler la dispersion des populations de vecteur infestées par le nématode tout en intégrant l'effet de synergie entre les deux organismes. Ces approches peuvent être adaptées à d'autres cas d'associations phorétiques et, par extension, contribuer à la compréhension de la dispersion des espèces au sein de ces systèmes complexes et peu étudiés. D'un point de vue appliqué, les résultats obtenus au cours de cette thèse ont permis de mettre en évidence plusieurs barrières à la dispersion de *M. galloprovincialis*, qui constituent des informations essentielles dans l'élaboration de stratégies de contrôle de l'expansion du nématode du pin en Europe.

Mots clés : invasion biologique, association phorétique, dispersion, phylogéographie, génétique spatialisée, modèle de dispersion.

ABSTRACT

Biological invasions dramatically increased over the last decades due to the intensification of international trade. These invasions constitute a threat for ecosystems and anthropic activities, therefore it is crucial to understand underlying processes in order to better predict and manage their impacts. Species introduced in a new environment interact with species of the receiver ecosystem and the novel interactions induced (predation, parasitism, mutualism...) can significantly contribute to invasion success and associated damages. This case is encountered for novel phoretic associations that may appear between a vector species and a phoronte. Such associations may generate important damages, in particular when they combine the virulence of a pathogen and the important dispersal abilities of a vector species. Several examples of novel phoretic associations including vector-pathogen relationships have shown considerable impacts on forests ecosystems with dramatic ecological and economic consequences. The study of novel phoretic associations is complex because it requires knowledge on native and introduced species together with the synergetic effect of their association. Therefore, despite the importance of this phenomenon, these associations remain poorly known, in particular regarding the dispersal ability of species involved.

In this PhD thesis, I have explored the dispersal potential of species involved in a novel phoretic association. I have focused on the case of the association between a pest for pine forests, the pinewood nematode (*Bursaphelenchus xylophilus*) introduced in Europe, and its endemic insect vector (*Monochamus galloprovincialis*). This novel phoretic association has generated huge damages in the Iberian Peninsula since its detection in Portugal in 1999, and its spatial expansion constitutes a major threat for European pine forests. First, I have characterized the genetic structure of the endemic vector across its European range and I have explored the historical and environmental drivers of observed genetic patterns. The objective of this first study was to detect the barriers and factors possibly affecting the dispersal of the vector at wide scale. Secondly, and at the scale of the Iberian Peninsula, I have evaluated dispersal of the vector in a heterogeneous environment by testing the existence of barriers and corridors to gene flow using a spatial genetics approach. Finally, I have explored the spread of the vector and of the nematode through the Pyrenean barrier by combining a study of gene flow and a spread model accounting for the synergistic effect of this novel association.

The results obtained through these studies have shown that *M. galloprovincialis* exhibits a genetic structure characteristic of postglacial recolonisation patterns of European temperate biota, and that the precipitations are the main parameters driving the distribution of evolutionary lineages above historic factors. These results also suggest that this beetle is not specialized according host taxa of *Pinus*. The spatial genetics study performed in Iberian Peninsula has shown the existence of barriers to dispersal, mainly where elevated area occurs, and where temperatures are the lowest. The results also show that high pine densities constitute barriers to dispersal of *M. galloprovincialis* and conversely that open areas with scattered pines could enhance dispersal. Finally, the study on the Pyrenean chain suggests that this mountain could represent a barrier, but only partial, to the spread of the pinewood nematode, this nematode being able to surround this chain in lowland valley located in eastern and western hillsides. Simulations have also shown a potentially strong effect of low temperatures on the spread of the pinewood nematode, and therefore a high risk to observe a reduction or a disappearance of these barriers due to climate warming.

From a fundamental point of view, this work provides a significant contribution to knowledge on evolutionary history of the vector and on the parameter driving its dispersal. This PhD thesis has also led to the development of a new method in the design of spatial genetics analysis. This approach allows generalizing inference in such approach using replications of scales and localities of study area, and is particularly suitable for species with good dispersal abilities, such as insect vector species. The refined spread model developed in the course of this thesis has allowed simulating the spread of nematode-infested populations while accounting for the synergistic effect of the phoretic association between the two species. These approaches may be transferred to other cases of phoretic associations and, by extension, may contribute to the understanding of the dispersal of species involved in those complex and poorly known systems. From an applied perspective, the results obtained during this

thesis have highlighted several barriers to dispersal of *M. galloprovincialis* that constitute substantial information for the development of strategies for the management of the spread of the pinewood nematode in Europe.

Key-words: biological invasion, phoretic association, dispersal, phylogeography, spatial genetics, spread model.

REMERCIEMENTS

Le chemin parcouru pour atteindre un objectif importe autant que l'objectif lui-même : c'est là un des enseignements de plusieurs philosophies orientales. Cette thèse est l'aboutissement d'un sentier fait de travail, d'échanges et de rencontres, jalonné de difficultés, de surprises et de joies, que j'ai eu la chance de toujours pouvoir partager. Je souhaite donc remercier ici toutes les personnes qui ont été à mes côtés au cours de ces trois années, et qui ont contribué, de près ou de loin, à la richesse et à l'effervescence de cet épisode de vie.

Je remercie en particulier mes encadrants, Christelle Robinet, Géraldine Roux et Alain Roques. J'ai particulièrement apprécié la confiance que vous m'avez accordée et la liberté avec laquelle j'ai pu évoluer dans mes travaux, tout en pouvant compter sur votre soutien. Vous avez joué un rôle déterminant dans la construction de ma jeune carrière de chercheur, et je vous en suis reconnaissant. Je souhaite également saluer l'ensemble des membres du laboratoire « zoo », pour l'ambiance conviviale, pour votre aide et pour la complémentarité de votre regard sur mes travaux. Un merci particulier à Jérôme Rousselet, Claudine Courtin et à tous ceux avec qui j'ai pu avoir tant de discussions passionnantes, autour d'un repas, devant un verre ou entre deux portes. J'ai beaucoup appris à votre contact. Merci à Natalia Kirichenko pour tes cours de russe, qui m'ont beaucoup servi sur le terrain. Je remercie aussi les collègues des unités voisines d'« amélio » et de « carto », pour votre aide et votre sympathie. Merci à Vincent Segura, Leo Sanchez, Joël Daroussin et en particulier à Sébastien Lehmann, pour cette initiation à la cartographie sous Arcgis autant que pour tes conseils avisés sur la construction en paille.

Mes remerciements vont également aux collaborateurs du programme REPHRAME : ce travail de thèse a été irrigué par le bouillon d'idées que j'ai récoltées dans les nombreux meetings et interactions que j'ai pu avoir avec ce réseau international. Merci aussi aux membres des sociétés entomologiques (SEF & SoMOS), qui ont toujours montré pour mes recherches un intérêt qui dépasse notre passion commune pour l'entomologie. Je remercie en particulier Jean-David Chapelin-Viscardi et Daniel Rougon auprès desquels j'ai pu recueillir de précieuses observations naturalistes sur la biologie et la distribution de *M. galloprovincialis*.

Je souhaite aussi adresser quelques mots de gratitude à ma famille, qui a toujours été attentive et a encouragé mon investissement en entomologie dès le plus jeune âge, malgré le caractère ésotérique que cette discipline pouvait avoir à leur yeux. Que mes parents et grands-parents trouvent dans ces lignes ma reconnaissance pour m'avoir soutenu et ouvert à la beauté et à la diversité du monde qui nous entoure. Mon épanouissement professionnel d'aujourd'hui repose sur leur accompagnement au cours de ces années de scolarité où chasser les insectes était à mes yeux nettement plus intéressant que de travailler mes leçons avec assiduité.

Un merci spécial à tous les membres du Groupe : ces lignes sont trop courtes pour décrire ce que vous représentez à mes yeux. Merci de m'avoir permis de mettre mon cerveau de côté de temps en temps, pour mieux le reprendre ensuite.

Enfin, je remercie celle avec qui j'ai la chance de partager ma vie et ma passion. Ces thèses sont un début, une vie de voyages et de découvertes nous attend...

AVANT PROPOS

Cette thèse s'est déroulée dans le cadre du projet Européen REPHRAME¹, ayant pour objectif de contribuer à la connaissance et au contrôle du nématode du pin² en Europe. Ce projet vise à fédérer l'action de plusieurs laboratoires pour mieux cerner les différents compartiments de la problématique de cette invasion (nématode du pin, maladie causée par le nématode, sensibilité des pins, insectes vecteurs, etc.). REPHRAME a permis la constitution d'un important corpus de connaissances sur ce modèle d'invasion, en particulier à travers le développement de deux thèses portant sur l'invasion du nématode (Mallez 2014) et sur la dispersion de son vecteur natif, *Monochamus galloprovincialis* (David 2014). Cet ensemble de connaissances a représenté un contexte favorable à une étude intégrative du potentiel dispersif du nématode du pin au sein de son association phorétique avec son insecte vecteur.

J'ai choisi de présenter ce manuscrit sous la forme d'une thèse sur articles, avec pour objectif que cela représente un ensemble cohérent et contribue de ce fait à la compréhension de ma démarche et des apports relatifs au modèle d'étude. Chaque article publié ou en préparation est présenté en anglais et précédé d'un résumé en français, à l'exception des articles méthodologiques qui sont brièvement présentés et joints en annexes du manuscrit. Enfin, une discussion générale me permettra de discuter de manière synthétique les résultats obtenus au cours des différentes études.

¹ REPHRAME: Analysis of the potential of the pinewood nematode (*Bursaphelenchus xylophilus*) to spread, survive and cause pine wilt disease in European coniferous forests in support of EU plant health policy.

²Nématode du pin: *Bursaphelenchus xylophilus* (Steiner & Buhrer, 1934; Nickle, 1970; Nematoda: Aphelenchoididae), appelé "pinewood nematode" (PWN) en anglais.

SOMMAIRE

– Chapitre I : Introduction	3
I.1- Les invasions biologiques	4
I.1.1- L’ampleur du phénomène	4
I.1.2- Le processus d’invasion	7
I.1.3- L’importance de la dispersion	9
I.1.4- Les mesures de la dispersion	10
A- Les méthodes directes	12
B- Les méthodes indirectes	13
I.2- Les associations phorétiques nouvellement formées (APNF)	16
I.2.1- Les associations phorétiques hors contexte des invasions biologiques	16
I.2.2- L’émergence de nouvelles interactions	17
I.2.3- Dispersion des invasifs dans le cadre des associations phorétiques nouvellement formées	18
I.2.4- Etude des flux de gènes et modélisation de la dispersion : deux outils complémentaires pour étudier la dispersion d’une espèce invasive impliquée dans une APNF	20
I.3- Modèle biologique : l’association entre le nématode du pin et le genre <i>Monochamus</i>	22
I.3.1- Le nématode du pin et la maladie du flétrissement du pin	22
I.3.2- Une invasion à l’Est et à l’Ouest de l’Eurasie	24
I.3.3- Les vecteurs et la dispersion du nématode du pin	25
I.3.4- La dispersion des vecteurs	30
I.3.5- Méthodes de lutte et réglementation	31
I.4 – Objectifs de la thèse	33
– Chapitre II : Diversité et structure génétique de <i>M. galloprovincialis</i> dans son aire native	35
II.1- Pertinence des marqueurs génétiques pour l’étude de la structure génétique de <i>M. galloprovincialis</i>	36
II.2- Phylogéographie de <i>M. galloprovincialis</i> : quels facteurs peuvent expliquer la structure génétique du vecteur à large échelle ?	40

– Chapitre III : Effet des paramètres environnementaux sur la dispersion de <i>M. galloprovincialis</i> : l'importance des reliefs	75
– Chapitre IV: Synergie potentielle de la dispersion du nématode et de <i>M. galloprovincialis</i> dans un milieu hétérogène : le cas de la barrière Pyrénéenne	119
– Chapitre V: Discussion générale et perspectives	153
V.1 – Potentiel dispersif du nématode à travers son association avec <i>M. galloprovincialis</i>	154
V.1.1 - Phylogéographie de <i>M. galloprovincialis</i> : une espèce structurée par les glaciations et les facteurs climatiques	154
V.1.2 - Les barrières et les corridors à la dispersion naturelle de <i>M. galloprovincialis</i>	161
V.1.3 - Synergie de l'association phorétique et dispersion du nématode	170
V.1.4 - Perspectives de recherche sur le modèle nématode- <i>Monochamus</i>	175
V.2 - Synthèse et implications pour les stratégies de lutte contre le nématode du pin	178
V.3 - Contribution du modèle à la connaissance des associations phorétiques	180
Références	185
Annexes	199

Chapitre I

Introduction

Chapitre I : Introduction

I.1- Les invasions biologiques

I.1.1- L'ampleur du phénomène

La distribution spatiale des organismes à la surface de la terre est conditionnée par leur histoire évolutive, leurs exigences écologiques et leurs capacités dispersives. Les ruptures spatiales dans les continuums écologiques délimitent des régions au sein desquelles sont définies les espèces dites natives et desquelles sont exclues les non-natives (Richardson et al. 2011). Au cours des siècles passés, les activités anthropiques ont entraîné le déplacement d'organismes de leur aire d'origine vers de nouvelles aires. Ces activités ont favorisé l'introduction, l'établissement et la prolifération spatiale d'espèces transférées, conduisant au phénomène d'invasion biologique (Mack et al. 2000).

Fruit d'un monde multi-connecté où le nombre et la diversité des échanges augmentent et où les temps de transports diminuent, le nombre d'invasions biologiques s'est accru de manière exponentielle au cours des dernières décennies en Europe comme à travers le monde (Pysek et al. 2003, Figure 1A : Roques et al. 2010). Les conséquences de ce phénomène se font à plusieurs niveaux. Tout d'abord, les invasions biologiques représentent dans leur globalité une menace pour la stabilité des écosystèmes et pour le maintien de la biodiversité (McKinney & Lockwood 1999, Courchamp *et al.* 2003, Olden *et al.* 2004). Les espèces introduites peuvent altérer les écosystèmes envahis par compétition, prédation ou hybridation avec les espèces natives, conduisant à une réduction de la diversité biologique et à une homogénéisation biotique globale (Rhymer & Simberloff 1996, Mooney & Cleland 2001, Sakai et al. 2001). De ce fait, les invasions biologiques sont considérées comme la seconde menace pour la biodiversité après la perte d'habitats et la fragmentation du paysage (Walker & Steffen 1997). Les invasions biologiques peuvent également affecter la santé humaine, par le biais d'introductions d'espèces vectrices de maladies. C'est le cas par exemple de l'invasion mondiale d'*Aedes albopictus* (Diptera, Culicidae) vecteur potentiel de 22 arbovirus et en particulier du chikungunya (Paupy et al. 2009). Enfin, l'effet le plus manifeste des invasions biologiques est

probablement leur impact sur les activités économiques. Ces impacts sont à la fois directs, lorsqu'ils affectent directement la production de biens et de denrées ou les services (l'agriculture, la foresterie, la pêche, le tourisme etc.), et indirects en raison des embargos et restriction sur les exportations que ces invasions impliquent. On estime leur impact au niveau mondial à 1400 milliards \$US par an (Pimentel et al. 2001).

La compréhension des mécanismes sous-jacents au succès ou à l'échec des invasions biologiques peut contribuer significativement à limiter la fréquence et les impacts de ce phénomène. Cependant, l'étude des invasions biologiques ne se résume pas à des objectifs de lutte ou de prévention. En effet, au-delà des considérations appliquées, les invasions biologiques émergent en tant que modèle pour explorer des questions plus fondamentales. La diversité des situations générées constitue une importante source d'expérimentation en milieu naturel pour tester des hypothèses relatives aux champs de la biologie évolutive (phénomènes micro et macro-évolutifs) tels que l'adaptation à un nouvel environnement, la spéciation, l'hybridation (Allendorf & Lundquist 2003). Les cas d'introductions multiples dans différentes zones peuvent même constituer des répliquats permettant de comparer de tels phénomènes évolutifs selon les environnements (Ciosi et al. 2008, Lombaert et al. 2010). On observe actuellement une explosion du nombre d'études portant sur les invasions biologiques (Figure 1B, Gurevitch et al. 2011) - due, certes - à l'intensification du phénomène et aux moyens accordés à cette thématique de recherche, mais témoignant également de l'intérêt fondamental qu'offrent ces situations inédites.

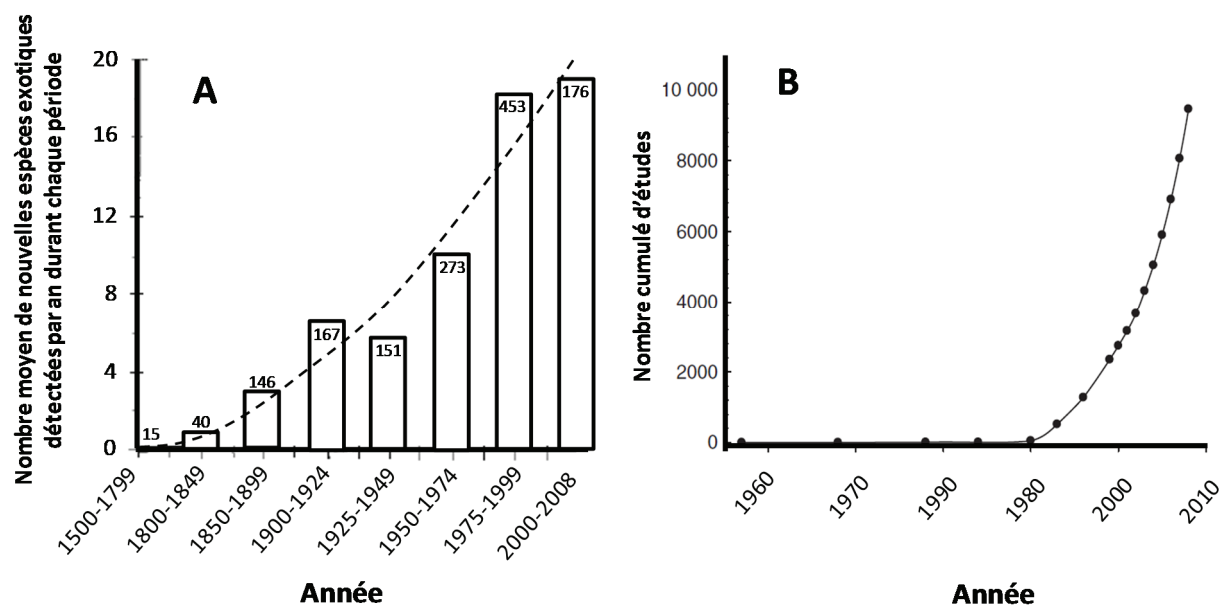


Figure 1 : **A** ; Nombre moyen d'espèces d'invertébrés exotiques détectées par an et par intervalle de temps en Europe (adapté à partir de Roques et al. 2010). **B** ; Nombre cumulé d'études relatives aux invasions biologiques (Gurevitch et al. 2011).

I.1.2- Le processus d'invasion

Une invasion biologique est un processus pouvant se résumer en quatre étapes: le transport, l'introduction, l'établissement et l'expansion (Figure 2A, Duncan et al. 2003). Le transport correspond à la translocation de propagules (graines, femelle fécondée ou population) de leur aire native vers une nouvelle zone. Cette étape peut conduire à la seconde étape : l'introduction (ou relâchement), qui correspond au moment où les propagules atteignent une nouvelle zone déconnectée de leur aire de distribution native. Ces deux premières étapes sont réalisées par l'intermédiaire des activités anthropiques. Une fois l'introduction réalisée, vient l'étape d'établissement, qui correspond à la survie, l'installation et le maintien de la population introduite dans le nouvel environnement. Cette étape est critique car la population introduite doit faire face et s'adapter à de nouvelles conditions avec un pool d'individus en général restreint (Berec et al. 2007, Suckling et al. 2012). Enfin, l'étape d'expansion correspond à la progression spatiale de l'espèce dans l'aire envahie, accompagnée d'une croissance démographique. La réussite de cette dernière étape conditionne le statut envahissant ou non des populations établies.

La réussite du processus d'invasion n'est pas systématique, et seule une faible proportion des espèces qui émigrent deviennent envahissantes dans une nouvelle aire. Ce constat a été théorisé pour les plantes par Williamson et Fitter (1996) en tant que règle des « dixièmes » qui stipule que seulement 10% des espèces parviennent à franchir chaque étape du processus d'invasion. Concrètement, cela signifie qu'une espèce introduite sur mille devient invasive. Cette règle traduit un fonctionnement général plus qu'une règle stricte et peut varier selon les groupes taxonomiques ou fonctionnels considérés (Williamson & Fitter 1996). En revanche, elle illustre la manière dont les barrières physiques et biologiques peuvent affecter chaque étape du processus d'invasion. Déterminer les facteurs permettant aux espèces de se maintenir et de s'étendre reste un enjeu majeur en biologie de l'invasion, en particulier pour la prédiction du potentiel invasif de certaines espèces. Cependant, l'établissement de lois générales est confronté à la singularité de chaque événement d'invasion et aux exigences particulières de l'espèce impliquée.

Au cours du processus d'invasion, la pression propagulaire – qui correspond à la quantité de propagules arrivant dans un nouvel environnement et à la fréquence de leur arrivée – joue un rôle prépondérant dans le franchissement des différentes étapes (Figure 2B, Allendorf & Lundquist 2003). En amont de l'établissement, cette pression propagulaire est principalement déterminée par les transports d'origine anthropique. En aval de l'établissement, la pression peut être également déterminée par les transports humains, mais aussi par les moyens propres à l'espèce, qui correspondent à ses capacités dispersives.

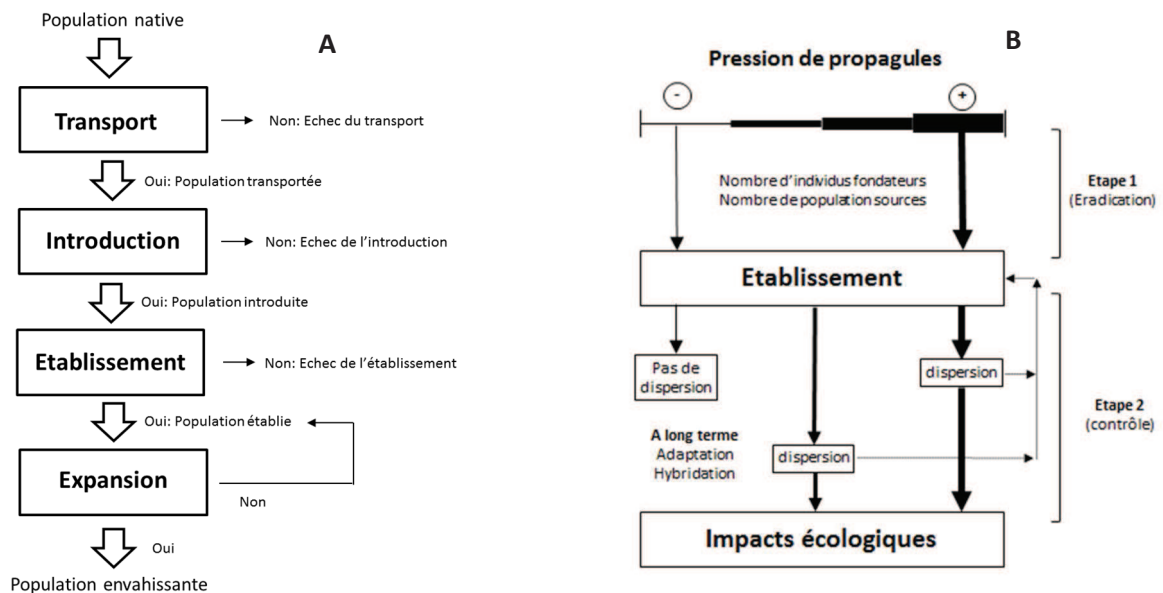


Figure 2 : A ; Schéma du processus d'invasion biologique (adapté à partir de Duncan et al. 2003). Les cadres représentent les étapes successives requises pour qu'une population native devienne envahissante. **B** ; Représentation schématique de l'effet de la pression de propagule sur le succès de l'invasion et les types de contrôles associés (adapté à partir de Allendorf & Lundquist 2003). Une pression de propagule importante augmente la probabilité d'établissement et de dispersion d'une espèce dans un laps temps plus court.

I.1.3 - Importance de la dispersion

Le comportement de dispersion est un trait adaptatif des espèces, leur permettant de trouver leur ressource, un partenaire et un substrat pour se reproduire, tout en évitant la consanguinité (Hanks 1999), c'est également un des paramètres qui conditionne la capacité d'une population à s'adapter à de nouveaux environnements (Ellstrand & Elam 1993). Dans le contexte des invasions biologiques, la dispersion joue un rôle central car elle détermine la capacité de la population établie à progresser spatialement et donc à franchir l'étape d'expansion, ce qui correspond – comme nous l'avons vu précédemment – à la définition même d'une espèce invasive. Si une population établie mais restreinte géographiquement peut être éradiquée, une population en phase d'expansion donne en général un tournant irréversible au processus d'invasion, et peut être au mieux contenue ou ralentie mais rarement éradiquée (Figure 2B, Pluess et al. 2012a, 2012b). Dans la majorité des cas, une population qui disperse et qui n'est pas soumise à un effet Allee trop fort s'étend spatialement et démographiquement (Roques et al. 2008). Cette phase d'expansion correspond en général au moment où elle est détectée au sein de son aire d'introduction (Roques et al. 2010, Eschen et al. 2015) et où ses effets sur les écosystèmes et les activités anthropiques deviennent perceptibles (Allendorf & Lundquist 2003).

La dispersion des organismes peut se faire de plusieurs manières. On distingue deux grands types de dispersion : active et passive. La dispersion active concerne la dispersion réalisée par les moyens de locomotion propres à l'espèce (autopropulsion : marche, vol, nage, etc.), et se rencontre principalement chez les animaux. La dispersion passive, quant à elle, concerne les cas où l'espèce est dispersée par un élément physique tel que le vent (anémochorie) ou la gravité (barochorie), ou un autre organisme (zoochorie). Ce mode de dispersion concerne principalement les organismes sessiles et les microorganismes. On parle également de dispersion passive dans les cas de transports d'individus par l'homme.

La dispersion (active ou passive) d'une espèce est un phénomène complexe, influencé par une multitude de paramètres (variabilité entre individus (poids, taille), variabilité au sein d'un individu (âge), climat, structure du paysage, intensité des transports par l'homme ou autres vecteurs, etc.). La

dispersion peut-être modélisée par un noyau de dispersion (Klein et al. 2006). Ce noyau donne la probabilité de se disperser en fonction de la distance et cette probabilité peut diminuer plus ou moins vite selon les paramètres considérés (Figure 3). En fonction des espèces et des conditions environnementales la forme du noyau de dispersion varie. La partie asymptotique de la courbe (queue de distribution, ou « tail » en anglais) est importante dans les phénomènes d'invasion car les rares individus qui dispersent sur de longues distances peuvent significativement contribuer à l'expansion spatiale de l'espèce et déterminer la vitesse d'avancée du front de colonisation (Klein 2006 et al. 2006, Robinet et al. 2012). Une queue de distribution fine correspond à une dispersion de type diffusive, ou « de proche en proche », formant une vague de colonisation avançant à vitesse constante. A l'inverse, l'existence d'individus dispersant à longue distance (par dispersion naturelle ou transport humain) favorise une dispersion de type stratifiée (Liebhold & Tobin 2008), où des patchs de colonisation à longue distance peuvent s'établir. Les patrons génétiques formés par ces types de dispersion (diffusive ou stratifiée) peuvent permettre d'estimer les distances de dispersion lorsque la différenciation génétique le permet (Manel et al. 2003, voir chapitre suivant pour ces méthodes).

I.1.4 - Les mesures de la dispersion

Estimer la dispersion d'un organisme est un problème constant en écologie car ce paramètre est complexe et influencé par de nombreux facteurs. La dispersion peut être quantifiée en termes de taux de dispersion ou de distance de dispersion. Il existe plusieurs méthodes pour estimer ce paramètre, que l'on peut regrouper en méthodes directes et indirectes. Une brève présentation de ces méthodes, de leurs avantages et de leurs limites est exposée ci-dessous.

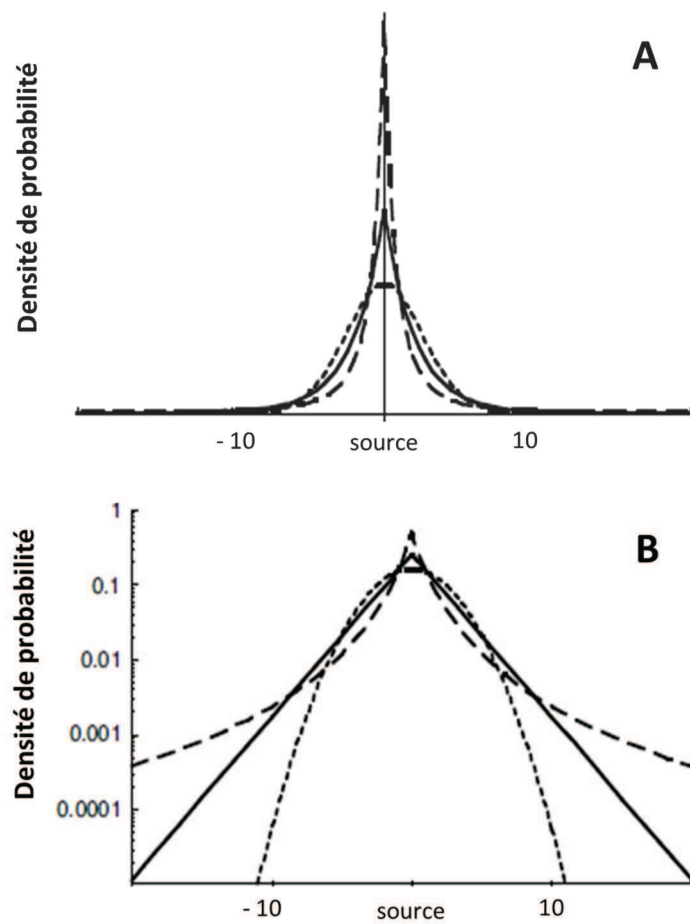


Figure 3. Représentation schématique de 3 noyaux de dispersion appliqués à des propagules se dispersant à partir d'une source. La courbe pleine représente le noyau exponentiel. La courbe en pointillés représente le cas des espèces avec peu d'individus dispersant sur de longues distances (noyau gaussien, queue de distribution fine). La courbe formée de tirets représente le cas des espèces avec un nombre important d'individus dispersant sur de longues distances (noyau puissance négative, queue de distribution lourde). **A** ; Densité de probabilité représentée sur une échelle linéaire. **B** ; Densité de probabilité représentée sur une échelle logarithmique faisant apparaître clairement les queues de distribution (d'après Klein et al. 2006).

A- Les méthodes directes

Ces méthodes regroupent les estimations de dispersion faites directement en milieu naturel. Elles peuvent résulter de la compilation de données historiques d'une invasion ou d'une expansion géographique. C'est le cas par exemple de l'expansion vers le nord de l'aire de distribution de la processionnaire du pin (*Thaumetopoea pityocampa*, Denis & Schiffermüller). Le suivi de cette expansion sur plusieurs années a permis d'évaluer l'avancée du front à 5,5 km par an, ce qui donne une estimation des distances de dispersion des adultes (Battisti et al. 2005). Cependant, ce type d'estimation doit être considéré avec précaution car la vitesse d'un front d'expansion d'une espèce peut résulter des capacités de dispersion intrinsèques à l'insecte, mais aussi des transports longue distance résultant des activités humaines, ou encore de la densité des populations. Une seconde méthode est celle de marquage d'individus, de leur lâcher et de leur recapture (« mark-release-recapture » ; Togashi 1990). Cette technique présente l'avantage de pouvoir être conduite en milieu naturel, ce qui recrée les conditions réelles de dispersion. En revanche, leur utilisation est limitée par le faible pourcentage d'insectes généralement re-capturés, par le nombre de pièges pouvant être mis en place (en particulier pour re-capturer les individus à des distances lointaines du point de lâcher), mais également en raison du fait que les individus sont souvent capturés à des distances plus courtes que ce que leurs capacités leur permettent (Yamamura et al. 2003). En conséquence, cette méthode sous-estime en général les capacités de dispersion (Ranius 2006). Le radio-tracking permet de suivre efficacement la dispersion des organismes (Millspaugh & Marzluff 2001, Hedin & Ranius 2002), mais cette méthode est difficile à adapter aux organismes de petite taille, en particulier aux insectes et aux pathogènes, qui constituent pourtant une part significative des invasions biologiques (Roques et al. 2010).

En général, les méthodes directes présentent l'avantage d'apporter des informations relativement précises sur les distances de dispersion, de surcroît mesurées directement en milieu naturel. En revanche, leur application nécessite fréquemment une logistique lourde, ce qui limite leur application dans l'espace et le temps (Broquet & Petit 2014).

B- Les méthodes indirectes

Les méthodes indirectes correspondent aux estimations de capacités dispersives faites hors milieu naturel. Une première méthode, adaptée aux espèces voilières, correspond à la mesure de dispersion en laboratoire à l'aide de manèges de vol (Vogt et al. 2000, Yamanaka et al. 2001, David et al. 2014). Cette méthode présente l'avantage de mesurer des distances en conditions standardisées, ce qui permet d'évaluer l'effet de paramètres tels que l'âge (Hughes & Dorn 2002), la taille (Boggs & Freeman 2005) ou le sexe (Senger et al. 2007) sur les distances de vol. Cependant, cette méthode a tendance à surestimer les distances de vol, car les individus ne portent pas leur propre poids, et car l'absence de contact avec le sol a tendance à pousser les individus au vol en raison d'une absence de contact des tarses avec un substrat (Edwards 2006).

La majorité des méthodes indirectes reposent sur l'inférence faite à partir des patrons génétiques (Broquet & Petit 2014). La dispersion et les flux de gènes qui lui sont associés laissent une signature dans la composition et la fréquence des allèles rencontrées dans les populations. Un premier ensemble de méthodes se base sur la fréquence des génotypes alléliques (obtenu par génotypage microsatellites ou SNP) dans une population et entre populations. Une mesure des flux de gènes associés à la dispersion d'un organisme (ou d'absence de dispersion) peut être obtenue avec des mesures telles que l'indice de différenciation (ou de fixation : F_{ST} ; Weir & Cockerham 1984), le coefficient d'admixture entre populations (Langella et al. 2001, Choisy et al. 2004), ou la distance génétique interindividuelle basée sur une approche ACP (Shirk et al. 2010). Ces mesures peuvent être corrélées aux distances géographiques dans les approches de génétique spatialisée (ou « landscape genetics ») pour estimer le différentiel de dispersion dans un environnement hétérogène en réaction aux diverses caractéristiques du paysage (Manel & Holderegger 2013). Ce type de méthode indirecte permet d'estimer un taux de migration relatif, mais ne permettent pas de calculer une distance de dispersion précise. Une situation particulière : l'hybridation intraspécifique, permet d'estimer les distances de dispersion sur la base de fréquences alléliques dans certaines situations. C'est le cas par exemple des zones de contact et d'hybridation entre deux foyers d'introduction d'espèces envahissantes (Bermond et al. 2013).

Un second ensemble de méthodes correspond à l'assignation des individus à des regroupements génétiques ou « clusters » (Pritchard et al. 2000, Corander et al. 2008), à des populations (Rannala & Mountain 1997, Paetkau et al. 2004, Manel et al. 2005) ou directement à des individus parents (Jones & Ardren 2003). Ces méthodes se basent sur des modèles (approche bayésienne) ou sur des calculs de vraisemblance. Les méthodes qui permettent d'assigner un individu à ses parents ou à sa population d'origine sont d'ailleurs considérées comme des méthodes directes par certains auteurs (Manel et al. 2005) quand les parents ne dispersent pas et que les mesures de dispersion parents-propagules peuvent être estimées directement (Telfer et al. 2003).

L'ensemble des méthodes indirectes basées sur l'étude des fréquences alléliques présentent l'avantage de mesurer les taux de migration et les distances réalisées en conditions naturelles et peuvent permettre d'identifier des phénomènes de dispersion à longue distance. En revanche, les invasions biologiques se font en général à partir d'un pool restreint d'individus (ou « bottleneck ») associé à une très faible diversité génétique. Par conséquent, l'estimation de ces distances peut être parfois difficile à calculer par manque de différenciation entre populations établies. Il faut aussi préciser que les méthodes indirectes basées sur l'analyse des flux de gènes offrent un aperçu de ces flux sur un grand nombre de générations et un grand nombre de mouvements dans l'espace et le temps. Cela se fait au détriment d'une connaissance précise de la distance dispersion en un lieu et moment donné.

La principale différence entre les méthodes directes et indirectes réside dans ce qui est réellement mesuré. Les méthodes directes, dans leur globalité, mesurent soit une fraction de la dispersion (expériences de marquage-lâcher-recapture) soit un ensemble de processus qui englobe la dispersion (vitesse d'expansion). Les méthodes indirectes quant à elles (hors manège de vol) donnent une estimation de la dispersion efficace, c'est-à-dire quand cette dispersion implique un flux de gènes. Il s'agit donc d'une mesure de distance entre le lieu de naissance et le lieu de reproduction des individus (Broquet & Petit 2014). Cette distinction est importante en fonction du type d'invasion biologique étudié et de ce que l'on cherche à mesurer. On ne s'intéressera pas aux mêmes distances selon que l'on considère l'invasion d'une espèce seule, ou du parasite (ou autre phoronte) qu'elle

transporte. En effet, il peut y avoir transport et transmission d'un phoronte sur des distances plus longues (si le phoronte est transmis lors de la prise alimentaire par exemple) ou plus courtes (s'il est transporté entre les prises alimentaires) que celle à laquelle l'espèce vectrice se reproduit. En raison des intérêts et des limites relatives à chacune des méthodes présentées, ces approches doivent être considérées comme complémentaires (Koenig et al. 1996, Broquet et al. 2006) pour estimer la dispersion des organismes.

Les méthodes exposées dans ce chapitre sont en général utilisées pour mesurer la dispersion d'une seule espèce. Dans certains cas d'invasion en revanche, la dispersion des espèces introduites peut se faire par l'intervention d'une ou plusieurs espèces endémiques. C'est le cas par exemple des associations phorétiques nouvellement formées.

I.2- Les associations phorétiques nouvellement formées (APNF)

I.2.1- Les associations phorétiques hors contexte des invasions biologiques

Les symbioses sont très largement répandues sur Terre et impliquent un large éventail d'organismes (Klepzig et al. 2009). Le terme de symbiose employé ici fait référence à la définition qu'en donne Klepzig et al. (2009): « l'acquisition et le maintien d'un ou plusieurs organisme par un autre, qui peut résulter en l'apparition de nouvelles structures et/ou métabolismes ». Les symbioses peuvent conduire à plusieurs types d'interactions tels que le bénéfice réciproque (mutualisme), le bénéfice unilatéral (commensalisme) ou encore le parasitisme. Les associations phorétiques constituent un exemple de symbiose qui correspond au transport d'un organisme par un autre. Elles peuvent être parasitaires ou commensales, mais également mutualistes lorsque l'organisme transporté apporte un bénéfice à son vecteur. Les associations phorétiques mutualistes sont particulièrement répandues chez les insectes. Ceux-ci ont formé des associations avec des microorganismes (phorontes) appartenant à plusieurs embranchements, tels que les champignons, les bactéries, les virus, les algues ou encore les nématodes, leur permettant la colonisation de multiples milieux et substrats (Klepzig et al. 2009). L'exemple le plus probant pour illustrer ce propos est le cas de l'association entre certaines espèces de scolytes (Coleoptera, Curculionidae) et de champignons phytopathogènes. Les adultes véhiculent les champignons qu'ils utilisent pour affaiblir les arbres aux-dépens desquels ils se développent, dépendant parfois également du développement du mycélium dans le bois pour s'en nourrir (Francke-Grosmann 1967). Dans certains cas, les adultes transportent même une communauté de champignons leur permettant ainsi faire face aux variations climatiques, en jouant sur la complémentarité des optimum de développement des phorontes (Hofstetter et al. 2006, Six & Bentz 2007). Plusieurs espèces (Genre *Scolytus*, *Platypus*) véhiculent les propagules des champignons qui leur sont associés dans des structures cuticulaires spéciales plus ou moins complexes (ou mycangium), témoignant d'une véritable coévolution entre les partenaires de l'association (Klepzig et al. 2009).

I.2.2- L'émergence de nouvelles interactions

L'introduction d'une espèce dans un nouvel environnement peut engendrer un ensemble de nouvelles interactions plus ou moins directes avec les espèces de l'écosystème receveur (Gandhi & Herms 2010, Alsan et al. 2015). Dans certains cas, l'expansion d'un invasif peut être fortement encouragée par la formation d'une relation phorétique avec une ou plusieurs espèces natives, qui sera appelée association phorétique nouvellement formée (APNF) dans la suite de cette thèse. L'apparition de telles associations est favorisée par l'existence de similitudes structurelles et phylogénétiques entre écosystèmes d'un même biome au sein duquel les espèces sont transférées. Suite à une introduction, il peut y avoir compétition entre espèces natives et invasives proches phylogénétiquement pour la même association phorétique et remplacement au sein de cette niche. Les APNF de type mutualiste sont peu fréquentes au regard du nombre d'invasions biologiques, mais sont spectaculaires par l'envergure des dommages qu'elles engendrent au sein des écosystèmes receveurs. Dans de telles situations, le phoronte bénéficie du transport vers un hôte compatible, et de son inoculation dans des tissus susceptibles, tandis que l'insecte bénéficie de la virulence du microorganisme pour accéder à grand volume de ressource (Lu et al. 2010).

Ces situations sont surtout documentées pour les insectes forestiers car les associations insectes - pathogènes y sont naturellement très répandues et les dommages associés facilement observables. Les nouvelles associations phorétiques mutualistes existent dans deux sens. Elles peuvent impliquer un pathogène natif et un vecteur introduit (i), c'est le cas par exemple de l'invasion de *Dendroctonus valens* LeConte en Asie et de son association avec plusieurs champignons endémiques de la famille des Ophiostomatoïdes, qui est supposée être responsable de la mort de plus de 3 millions de *Pinus tabulaeformis* en Chine (Lu et al. 2010). Ces associations peuvent également se faire entre un vecteur natif et un pathogène introduit (ii) (Potter et al. 2011). Nous nous focaliserons sur ce dernier cas dans la suite de ce manuscrit. L'exemple le plus documenté est celui de la graphiose de l'Orme (ou Dutch Elm Disease, DED, en anglais) qui a dévasté les ormes d'Europe à partir de 1918 (Karnosky 1979). Cette maladie résulte de l'association entre un coléoptère natif associé aux ormes (*Scolytus multistriatus*, Marsham) et de plusieurs souches de champignons ascomycètes dérivées d'*Ophiostoma*

Ulmi (Buisman) melin & Nannf supposées être originaires d'Orient (Karnosky 1979). De telles associations sont rendues possibles car il existe en général une cohérence phylogénétique dans la distribution des hôtes possibles d'un pathogène. Celui-ci est donc susceptible d'infecter plusieurs espèces au sein d'un genre ou d'une famille, pouvant donc se trouver en dehors de son aire d'origine (Farret et al. 2004). En revanche, les hôtes concernés peuvent montrer des degrés très élevés de susceptibilité en raison d'un manque de coévolution avec le pathogène. Cette sensibilité est parfois également due à une valeur sélective (fitness) supérieure du pathogène dans sa zone d'introduction par rapport à sa zone d'origine (Garbelotto et al. 2012). Les vecteurs natifs impliqués dans de nouvelles associations ne causent en général pas de dommages dans leur aire native avant l'arrivée du pathogène. Ce sont souvent de simples phytophages en équilibre dans leur milieu et sont tout au plus des ravageurs secondaires ou dits « de faiblesse ». En présence du pathogène introduit, ceux-ci peuvent atteindre des densités de populations très élevées en raison de la densité de ressource (arbres morts) mise à disposition, favorisant ainsi la dispersion du pathogène. Il y a alors effet de synergie entre les deux organismes, contribuant fortement à l'expansion spatiale et au succès de l'invasion du pathogène introduit.

Les impacts des APNF tant sur les écosystèmes que sur les activités économiques peuvent être considérables. Ce constat est en faveur d'une meilleure compréhension de leur fonctionnement et en particulier des mécanismes qui permettent au partenaire non-natif de l'association de s'étendre spatialement une fois établi.

I.2.3- Dispersion des invasifs dans le cadre des associations phorétiques nouvellement formées

Si l'effet des symbioses en général sur les invasions biologiques est bien documenté (McLeod et al. 2005, Lu et al. 2010, Himler et al. 2011, Zaho 2013), le fonctionnement des APNF est cependant mal connu, en particulier pour ce qui concerne le potentiel à la dispersion des espèces qu'elles impliquent. Les études portent généralement sur les pullulations de pathogènes forestiers, en se

focalisant sur la structure génétique du pathogène, sur la distribution des dépérissements et sur les facteurs associés à l'expression de la maladie. Peu d'attention est portée sur les modes de vexion des pathogènes et de leur distance de dispersion (Garbelotto et al. 2012). Ceci est dû en partie au cloisonnement entre les domaines de l'entomologie forestière et de la pathologie forestière (Klepzig et al. 2009).

Comme nous l'avons évoqué précédemment, l'estimation de la dispersion d'une espèce est rendue difficile par la multitude des facteurs pouvant l'influencer en conditions naturelles, et par les biais relatifs aux différentes méthodes de mesure (Chapitre I.1.4). L'estimation du potentiel de dispersion des espèces impliquées dans une APNF se trouve à un degré supérieur de complexité car elle doit intégrer les exigences et les aptitudes de chaque espèce et les différents aspects de leurs interactions. L'existence de vecteurs natifs peut permettre de faire une première estimation du potentiel de dispersion du partenaire non-natif d'une APNF, sur la base de la distribution et de la dispersion des vecteurs en conditions naturelles (Battisti et al. 2005, David et al. 2014). Ces connaissances sont essentielles pour identifier les zones favorables ou défavorables à l'expansion de l'espèce introduite. En revanche, elles n'intègrent pas les effets positifs ou négatifs de cette nouvelle association, qui peuvent pourtant affecter fortement la dispersion des partenaires impliqués. De nombreuses associations mènent par exemple à une élévation de la densité de population du vecteur (Jiu et al. 2007), et par conséquent à une augmentation de la proportion d'événements de dispersion à longue distance. Certains cas d'associations nouvellement formées peuvent favoriser la dispersion du vecteur, par une élévation de l'attractivité des arbres infestés (McLeod et al. 2005). C'est le cas par exemple de la manipulation des ormes nord-américains par *Ophiostoma novo-ulmi* qui entraîne une émission de substances sémiochimiques plus attractives pour le vecteur (*Hylurpoinus rufipes*). A l'inverse, les phorontes peuvent réduire les distances de dispersion du vecteur. Ce cas est rencontré chez *Monochamus carolinensis*, qui montre des distances de vol plus faibles lorsque la charge en nématodes du pin (*Bursaphelenchus xylophilus*) dans ses spiracles est importante (Akbulut & Linit 1999). Ces associations peuvent également faire intervenir des manipulations comportementales du vecteur par le pathogène, qui sont susceptibles d'impacter sa dispersion (Mayer et al. 2008).

I.2.4- Etude des flux de gènes et modélisation de la dispersion: deux approches complémentaires pour étudier la dispersion d'une espèce invasive impliquée dans une APNF

Pour estimer au mieux la dispersion des espèces impliquées dans une APNF, il est donc essentiel d'estimer à la fois la dispersion du vecteur et de prédire l'effet synergique de l'association. Pour cela, l'utilisation simultanée d'approches de génétique (estimation des flux de gènes chez le vecteur) et de modélisation de la dispersion est particulièrement adaptée car elle apporte des informations complémentaires.

Nous l'avons vu précédemment, les méthodes d'estimation de la dispersion basées sur les flux de gènes permettent de décrire la dispersion naturelle du vecteur (barrières et couloirs de dispersion), mais également d'identifier les sauts à longue distance liés aux activités anthropiques (cf. section I.1.4). Cumulées, ces mesures contribuent à donner une image complète des phénomènes de dispersion chez le vecteur, et permettent de mieux identifier les voies potentielles d'invasion du phoronte. Ceci représente d'ailleurs une particularité des APNF qui contraste avec les invasions impliquant seulement une espèce et pour lesquelles une telle prévision n'est pas possible. Cependant, ces estimations présentent plusieurs limites. La première a été développée dans le chapitre précédent et correspond au fait que l'estimation basée sur les flux de gènes du vecteur ne prend pas en compte l'effet de synergie de l'APNF. La seconde, est que cette dispersion est inférée à partir de patrons génétiques qui se sont établis sur une échelle de temps inconnue. Il est donc parfois difficile de déterminer si ces patrons proviennent d'une dispersion contemporaine dans l'environnement présent, ou s'ils sont issus de la dispersion et de variations démographiques des individus au sein de paysages anciens (Broquet & Petit 2009).

Pour contourner ces difficultés, les approches de modélisation représentent une approche complémentaire. L'expansion des espèces invasives est en général modélisée à l'aide de modèles mathématiques de type réaction-diffusion (Shigesada & Kawasaki 1997). En tant qu'outils prédictifs, ces modèles offrent la possibilité d'intégrer non seulement le processus de dispersion par le vecteur

mais aussi les effets de synergie au sein de l'APNF (*i.e.* croissance des populations de vecteurs). De plus, la dispersion de l'invasif peut être simulée au sein de l'environnement présent, mais également être replacée dans un contexte de climat changeant ou de connectivité du paysage modifiée sous l'effet des activités anthropiques (Robinet et al. 2014). En revanche, l'application de ces modèles dépend de l'existence de connaissances relatives à la biologie et à la dispersion de l'espèce, qui sont parfois rares ou inexistantes. La modélisation est donc un complément intéressant à l'estimation des flux de gènes, mais elle dépend des données disponibles pour être appliquée.

En raison de la faible prévalence des APNF, ces deux outils ont très peu été utilisés pour les cas d'APNF décrits et il n'existe pas – à notre connaissance – d'étude les intégrant simultanément. Plusieurs limites subsistent donc dans leur application, et celles-ci doivent être prises en compte lors de la prédiction de la dispersion des espèces impliquées dans ces associations. En particulier, les approches de génétique spatialisées, qui permettent de tester l'existence de barrières à la dispersion du vecteur (Manel & Holderegger 2013), montrent des limites lorsqu'elles sont utilisées chez des espèces à fort pouvoir dispersif (Dreier et al. 2014). En effet, ces approches ont été développées dans un but de conservation sur des espèces non voilières et à pouvoir dispersif relativement restreint (Cushman et al. 2006). Cependant, les espèces vectrices impliquées dans les APNF sont généralement des insectes voiliers pouvant disperser sur de grandes distances, et il existe très peu de travaux sur la dispersion de ces insectes (Zeller et al. 2012). En parallèle, il existe peu d'études modélisant la dispersion des espèces impliquées dans une APNF et ces travaux n'intègrent en général pas ou peu les multiples composants de l'interaction. Potter et al. (2011) ont par exemple modélisé la dispersion de la graphiose de l'orme en Angleterre en considérant le pathogène et le vecteur comme un seul organisme. Cette approche exclut donc les potentielles variations dans les optimums écologiques, la distribution ainsi que la densité de vecteurs qui sont pourtant déterminants dans les processus de dispersion. Des modèles de type réaction-diffusion ou noyaux de dispersion ont été développés pour simuler la dispersion du nématode du pin en Asie et en Europe (Togashi & Shigesada 2006, Robinet et al. 2009, Robinet et al. 2011). Ces modèles considèrent soit la dispersion de la maladie associée au nématode (pine wilt disease, Robinet et al. 2009, 2011), soit la dispersion des insectes vecteurs (supposés être

tous porteurs du nématode) et le nombre d'arbres infestés à l'échelle d'une parcelle (Togashi & Shigesada 2006). Ces modèles n'intègrent donc pas les effets de synergie au sein de l'APNF étudiée. En conclusion, les approches visant à évaluer les flux de gènes du vecteur d'une APNF et les modèles mathématiques sont des outils complémentaires à adapter à la complexité des associations phorétiques nouvellement formées.

I.3 - Modèle biologique : l'association entre le nématode du pin et le genre *Monochamus*

Cette thèse se focalise sur le cas de l'invasion du nématode du pin (*Bursaphelenchus xylophilus* (Steiner & Buhner 1934, Nickle 1970, Nematoda: Aphelenchoididae) en Europe et de son association avec un vecteur endémique : *Monochamus galloprovincialis* (Olivier 1795 ; Coleoptera, Cerambycidae). Cette section fait état des connaissances que nous avons de ces deux organismes et de leur association.

I.3.1- Le nématode du pin et la maladie du flétrissement du pin

Le nématode du pin (pinewood nematode, PWN), *Bursaphelenchus xylophilus* (Steiner & Buhner 1934, Nickle 1970), est un nématode pathogène nord-américain responsable de la maladie du flétrissement du pin, aussi appelée « pine wilt disease » (PWD, Figure 4A). Les symptômes de cette maladie se manifestent sous forme d'un jaunissement des frondes (« flagging »), d'un bleuissement du bois et d'un dépérissement de l'arbre qui peut survenir dans les deux mois consécutifs à l'infestation de l'arbre par le nématode (Figure 4B/C, Mamiya 1983, Naves et al. 2007a, Kuroda 2008). Ces symptômes sont les conséquences de l'activité du nématode qui se nourrit des cellules épithéliales des canaux résinifères, entraînant un blocage du flux de sève, ainsi qu'une diminution de la circulation de l'eau par cavitation des trachéides. Le PWN étant phytophage et mycophage, il commence son développement aux dépens des tissus vivants de l'arbre et le poursuit sur les champignons qui envahissent l'arbre lorsque celui-ci dépérit. Le développement du PWN est fortement influencé par la

température, et une augmentation de la température entraîne une réduction du temps de cycle, jusqu'à un optimum de 33°C (Mamiya 1975). De même, les conditions climatiques affectent le développement de la PWD et la vitesse d'apparition des symptômes (Mamiya 1983). Par exemple, cette maladie n'est pas observée dans les régions où les moyennes du mois de juillet sont inférieures à 20°C (Rutherford & Webster 1987, Rutherford et al. 1990). De manière générale, tout paramètre environnemental induisant le stress de l'arbre favorise l'apparition de la maladie (Evans et al. 1996, Evans 2014).

La PWD affecte principalement le genre *Pinus*, bien que d'autres espèces de conifères puissent être touchées par cette maladie (Evans et al. 1996). Les espèces de pins montrent une variation dans leur susceptibilité à la maladie. Les essences nord-américaines n'expriment généralement pas la maladie associée à la présence du nématode dans leur bois, en raison d'une longue coévolution avec cet organisme, et les dépérissements massifs d'arbres adultes ne sont donc pas observés (Evans et al. 1996). En revanche, les essences eurasiatiques et d'Amérique centrale montrent des degrés variés de susceptibilité (Evans et al. 1996). *P. sylvestris*, *P. pinaster*, *P. nigra* en Europe et *P. thunbergii*, *P. luchuensis*, *P. densiflora* en Asie représentent les essences les plus sensibles et des mortalités massives sont observées dans les zones où le PWN s'est établi (Evans et al. 1996, Rodrigues 2008).

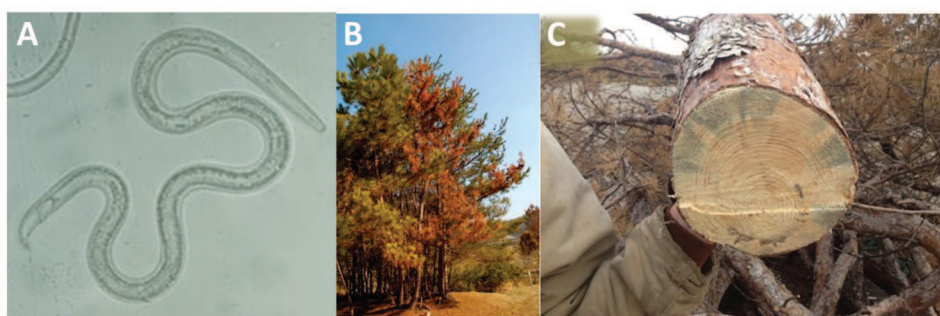


Figure 4. Nématode du pin (*Bursaphelenchus xylophilus* (Steiner & Buhrer 1934, Nickle 1970) ; **A**) et symptômes de la maladie du flétrissement du pin (« pine wilt disease », PWD). **B** ; dépérissement des frondes et (**C**) bleuissement du bois associé à l'expression de la pine wilt disease.

I.3.2- Une invasion à l'Est et à l'Ouest de l'Eurasie

Les routes d'invasion du PWN ne sont pas encore clairement établies en raison de difficultés techniques relatives aux l'analyses génétiques sur le PWN (Mallez et al. 2015). En revanche, la documentation des dépérissements liés à l'expression de la PWD permet de dresser un historique de l'invasion de ce nématode au cours du siècle dernier. Les premières observations de la PWD hors de son aire native ont été faites au Japon en 1905 sur l'île de Kyushu (Mamiya 1988). Le PWN a ensuite été détecté en Chine en 1982, puis à Taïwan en 1985 et enfin en Corée du sud en 1988 (Futai 2013). Les dommages aux forêts consécutifs à ces introductions sont considérables. En 2000 au Japon, la PWD s'étendait sur 28% des 2,1 millions d'hectares de forêts de pins que comptait le pays (Mamiya 2004). En Chine, cette maladie a infesté 1 million d'hectares de forêts sur la période 1982-2007 (Yu et al. 2011).

En Europe, le PWN a été détecté pour la première fois en 1999 au Portugal (Mota et al. 1999), en provenance, semble-t-il, de populations américaines (Mallez 2015). En 2005, la zone infestée a atteint une superficie de 617000 ha (Naves et al. 2007a). Malgré d'importantes mesures d'éradication, de nouveaux foyers ont été signalés dès 2008 dans le centre du pays (Rodriguez 2008), à Madère en 2009 (Fonseca et al. 2012), puis dans plusieurs régions limitrophes d'Espagne, en Extrémadure en 2008 (Abelleira et al. 2011), en Galice en 2010 (Robertson et al. 2011) et en Castille (EPPO 2014). La progression du PWN en Europe représente une menace sérieuse pour les pinèdes naturelles et pour les activités économiques associées à l'exploitation du pin. Son impact économique sur une période de 22 ans a été estimé à 22 milliards d'Euros de pertes si aucune mesure de contrôle n'était appliquée (Soliman et al. 2012). L'expansion géographique de ce ravageur et les dommages associés mettent en lumière l'importance de comprendre son potentiel de dispersion et les mécanismes sous-jacents.

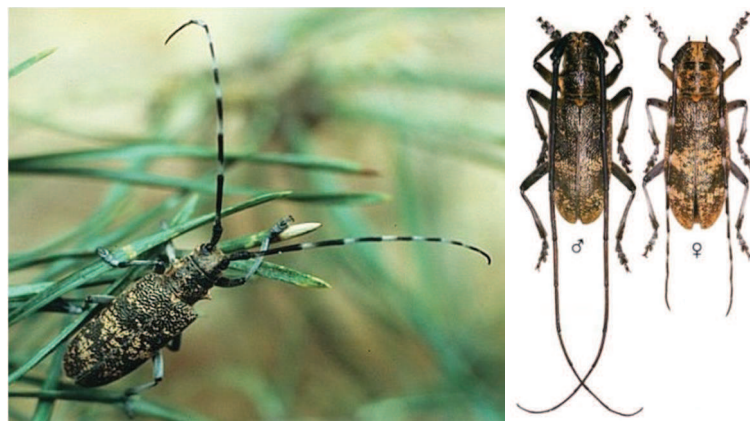


Figure 5. *Monochamus galloprovincialis* (Olivier 1795, Coleoptera, Cerambycidae), vecteur du nématode du pin en Europe.

I.3.3- Les vecteurs et la dispersion du nématode du pin

En condition naturelle, le PWN assure sa dispersion par le biais de son association phorétique avec plusieurs coléoptères longicornes du genre *Monochamus* (Coleoptera, Cerambycidae) associés aux pins et à d'autres conifères (Akbulut & Stamps 2012). Dans sa zone native, ce nématode est principalement dispersé par *Monochamus carolinensis* Olivier (Linit 1988), mais on le retrouve également associé à trois autres vecteurs : *M. mutator* (LeConte in Agassiz, 1850), *M. scutellatus* (Say, 1824) et *M. titillator* (Fabricius). En raison de la similarité entre les écosystèmes néarctiques et paléarctiques, le PWN a trouvé dans les zones où il a été introduit des systèmes pins-*Monochamus* préétablis compatibles, favorables à la formation de nouvelles associations phorétiques. De plus, les espèces du genre *Monochamus* sont fréquemment associées à des espèces natives de nématodes du genre *Bursaphelenchus*, les prédisposant naturellement à de telles relations de phorésie (Mamiya & Eda 1972, Vincent et al. 2008). En particulier, *M. galloprovincialis* (Olivier 1795), espèce vectrice en Europe, est déjà associée à l'état naturel avec l'espèce native *Bursaphelenchus mucronatus* (Vincent et al. 2008). Dans les zones d'introduction du PWN en Asie de l'est, *M. alternatus* (Hope, 1842) est le principal vecteur, bien qu'on le retrouve également associé à *M. saltuarius* (Gebler, 1830), *M. nitens* (Bates, 1884) et *M. rosenmuelleri* (Cederjelm 1798) (= *M. urussovi* (Fischer, 1806), Jong et al. 2014), (Akbulut & Stamps 2012). En Europe, *M. galloprovincialis* (Figure 5, Olivier, 1795) est à ce jour la

seule espèce vectrice connue (Sousa et al. 2001). Cette espèce présente deux formes mélaniques décrites : *M. galloprovincialis galloprovincialis* (aux téguments roux, distribuée principalement en plaine) et *M. galloprovincialis pistior* (aux téguments sombres, retrouvée principalement en altitude, et dans les régions au climat frais en général) dont le statut taxonomique et la validité génétique font débat (Koutroumpa et al. 2013, Wallin et al. 2013). Il existe en Europe cinq autres espèces endémiques de *Monochamus* potentiellement vectrices (*i.e.* *M. sutor* (Linnaeus, 1758), *M. sartor* (Fabricius, 1787), *M. rosenmuelleri*, *M. saltuarius*, *M. impluviatus* (Motschulsky 1859)) (Mamiya & Enda 1972, Mamiya 1988, Evans et al. 1996, Akbulut & Linit 1999). Toutes ces espèces peuvent se développer sur les pins, au moins ponctuellement (Wallin et al. 2013). Les *Monochamus* holarctiques sont des xylophages secondaires, cependant, ils ne se développent que dans un bois dépérissant ou fraîchement mort. Les femelles déposent leurs œufs à l'aide d'encoches qu'elles forment dans l'écorce avec leurs mandibules. Les larves se développent aux dépens du phloème dans les premiers stades et s'enfoncent dans l'aubier pour finir leur développement et former leur loge nymphale. Le développement larvaire démarre à la fin de l'été et se poursuit durant l'automne (Mamiya 1972, Evans et al. 1996). Les larves de dernier stade passent l'hiver en loge pour se nymphoser au printemps suivant. Les espèces de ce genre réalisent une génération par an en général, cependant ce nombre peut varier de 0.5 à 3 selon les conditions climatiques (Akbulut & Stamps 2012). Suite à leur émergence, les jeunes adultes ont une phase obligatoire de maturation durant laquelle ils se nourrissent de l'écorce des rameaux des arbres avoisinants. Cette phase est indispensable au stockage d'énergie et à la mise en place de leurs organes de reproduction (Katsuyama et al. 1989).

La dispersion du PWN par les *Monochamus* résulte du contact des deux organismes dans le bois attaqué et d'une concordance de leur cycle biologique (Figure 6). Le PWN effectue quatre stades larvaires qui précèdent le stade adulte sexué (Futai 2013). Le cycle biologique présente deux modes alternatifs, l'un de reproduction/multiplication (qui assure la colonisation de l'arbre) et l'autre dispersif/résistant (qui permet la dispersion via le coléoptère, Figure 7). Dans ce dernier mode, les deux derniers stades juvéniles (appelés « juvéniles Dauer ») possèdent des lipides de réserve qui leur permettent de résister à des conditions défavorables. Au printemps, les J3 Dauer migrent autour des

chambres nymphales des *Monochamus* par chémotaxie (Zaho et al. 2013) et atteignent le stade J4 Dauer. Ce stade ne se nourrit pas et va infester le système trachéen du vecteur après sa mue imaginale. Les nématodes sont alors dispersés par les adultes infestés suite à leur émergence (Figure 6), en particulier lorsque ces derniers volent et les transmettent aux arbres avoisinants.

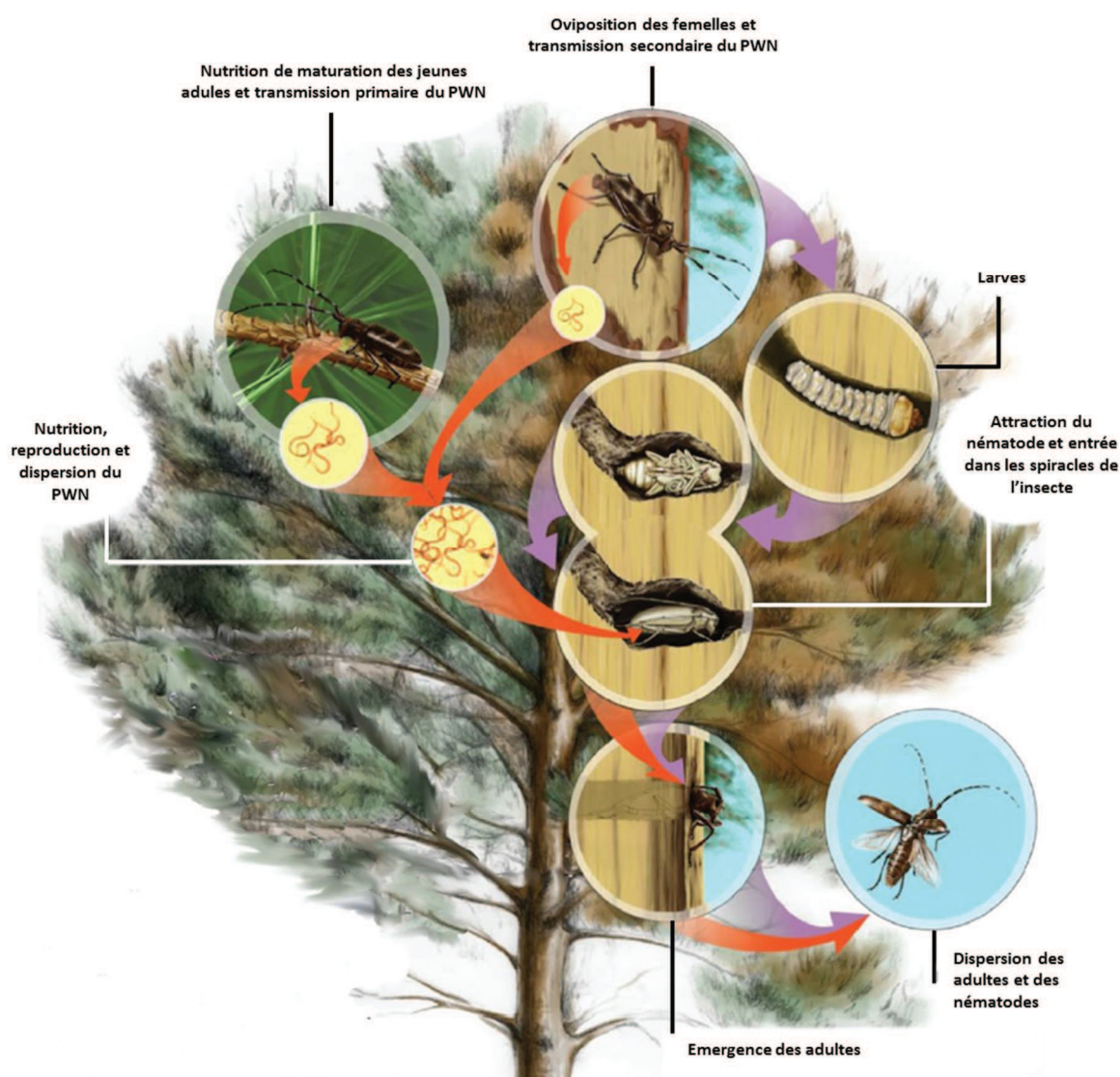


Figure 6. Développement conjoint du nématode du pin et de son vecteur (*Monochamus* sp.) sur les pins (Adapté à partir de Nascomento et al. 2015).

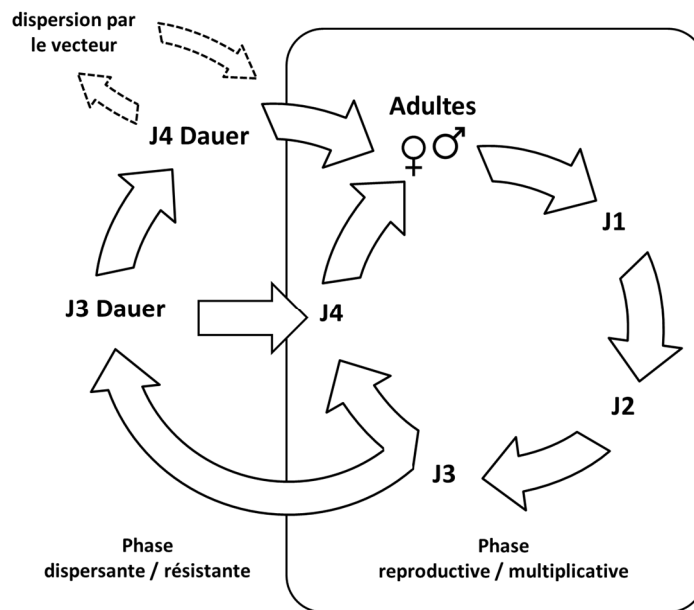


Figure 7. Cycle biologique du nématode du pin (Adapté à partir de Mallez 2014).

On distingue deux types de transmission du nématode par les *Monochamus* infestés : la transmission primaire et la transmission secondaire. La transmission primaire est réalisée lorsque les jeunes imagos blessent l'écorce des tiges de pins pour effectuer leur nutrition de maturation. Lors de cette prise alimentaire, les nématodes quittent les replis cuticulaires des coléoptères et s'introduisent dans les tissus des rameaux attaqués (Linit 1988, Vicente et al. 2012). Ce type de transmission semble majoritaire et peut conduire au développement de la maladie (PWD) sur des arbres sains, sous certaines conditions environnementales. La transmission primaire joue un rôle essentiel dans le processus d'expansion du nématode du pin et dans la dynamique des populations de vecteurs. L'affaiblissement et le dépérissement des arbres qu'elle implique conduit à la mise à disposition d'une ressource importante pour le vecteur. Cette ressource abondante a un effet attractif pour les femelles cherchant un site de ponte. Ceci conduit à une explosion démographique des coléoptères se développant dans les arbres infestés par le nématode. Les adultes émergeant de ces arbres sont donc porteurs du nématode et pourront eux-même le transmettre aux arbres avoisinants lors de leur nutrition

de maturation. Il y a donc effet de synergie entre les deux partenaires de cette APNF ; le vecteur bénéficie de la mise à disposition d'une ressource abondante et PWN bénéficie du transport de l'inoculation dans des arbres sains. Ce mécanisme conduit à la formation de foyers infectieux à forte densité de nématodes et de vecteurs, propices à la dispersion *de novo* du PWN.

La transmission secondaire est réalisée par les femelles lors de l'oviposition sur les arbres dépérissant (Linit 1988, Vicente et al. 2012). Il semblerait que ce second mode de transmission implique un nombre nettement plus faible de nématodes (Naves et al. 2007a). La transmission secondaire se fait généralement sur des arbres déjà morts ou dépérissants. Elle joue donc un rôle secondaire dans les mortalités liées à la PWD et n'entraîne pas d'effet de synergie. On notera cependant, que des vecteurs tels que *M. galloprovincialis* peuvent se développer sur les branches mortes du houppier (élagage naturel des pins) et ont donc le potentiel d'infester l'arbre par transmission secondaire via le contact entre la branche morte et le reste de l'arbre sain. Bien que le nématode puisse être transmis entre bois mort et vivant en conditions contrôlées (copeaux de bois et jeunes plants, Hopf & Schroeder 2013), ce potentiel n'a pas été observé *in natura*. L'importance relative des transmissions primaires et secondaires dans la dispersion du PWN est difficile à quantifier et reste donc globalement mal connue.

I.3.4- La dispersion des vecteurs

La dispersion des *Monochamus* est un paramètre critique dans l'invasion du PWN. Elargir nos connaissances sur l'aptitude à la dispersion de ces vecteurs est un prérequis essentiel pour estimer le potentiel du nématode à être dispersé et à coloniser de nouvelles aires. C'est également une donnée critique pour l'élaboration de mesures de lutte contre ce ravageur. La distance de dispersion des principaux vecteurs aux USA et en Asie (*M. carolinensis*, *M. alternatus*) est estimée à environ 2-3 km (Takasu et al. 2000, Linit & Akbulut 2001) sur la base d'observations en milieu naturel et de mesures en conditions contrôlées (manège de vol). Le vecteur européen, *M. galloprovincialis*, montre une distance de vol moyenne de 16 km cumulés sur sa durée de vie (David et al. 2014) en manège de vol. Les études de marquage-recapture réalisées en Espagne montrent que les adultes en conditions naturelles peuvent disperser sur des distances allant de 3 à 13 km selon les cas (Hernandez et al. 2011, Gallego et al. 2012, Mas et al. 2013). Un événement de dispersion de 22 km a été documenté (Mas et al. 2013), mais celui-ci pourrait être due à un transport par l'homme. Inversement, plusieurs études font état qu'une large proportion d'individus dispersent très peu et se maintiennent dans un périmètre inférieur à 100 mètres au cours de leur vie (Etxebeste et al. 2013, Torres-vila et al. 2015), en particulier lorsque les essences hôte sont disponibles (Bonifacio, 2009). Cependant, ces résultats doivent être considérés avec précaution, en raison des biais relatifs à ces méthodes (évoqués dans le chapitre I.1.4). Il a aussi été montré que les immatures de *M. galloprovincialis* répondent mal aux phéromones d'attraction utilisés dans les pièges, ce qui fausse les estimations de dispersion des individus impliquant ces stades (Ibeas et al. 2007). La dispersion de *M. galloprovincialis* est également variable selon le poids et l'âge des individus. David et al. (2015) ont en effet montré par manège de vol, que la distance de dispersion d'un individu est corrélée positivement à son poids à l'émergence et à son âge. L'hétérogénéité du milieu sur la dispersion peut également jouer un rôle important dans la dispersion de *M. galloprovincialis*. Cette hétérogénéité a été considérée dans une seule étude menée par Torres-vila et al. (2015) sur la base de marquages et de recaptures d'individus en milieu naturel. Cependant, cette étude n'a intégré qu'une seule variable (la topographie), et a été conduite en milieu ouvert, ce qui a tendance à encourager la dispersion, comme l'ont précisé les

auteurs. De manière générale, ces résultats descriptifs et dépendant du dispositif expérimental ne permettent pas d'extrapoler les conclusions à l'ensemble des milieux dans lesquels se maintient l'espèce et n'incluent donc pas toutes les barrières potentielles à la dispersion. Par conséquent, la dispersion des *Monochamus* réellement réalisée sous l'effet du paysage et des paramètres environnementaux reste méconnue.

En dehors de la dispersion « naturelle » réalisée par les vecteurs, le PWN peut également être dispersé par le transport sur de longues distances de bois, d'adultes ou de larves de coléoptères infectés par le biais des activités anthropiques. Le transport de coléoptères infectés est un mode de dispersion à longue distance important pour cette espèce, car la transmission à des arbres sains peut se faire directement par les coléoptères transportés. De tels événements sont difficiles à intercepter car très sporadiques, mais ils peuvent significativement contribuer à l'expansion de la PWD sur des distances supérieures à 100 km (Robinet et al. 2009). Un adulte de *M. galloprovincialis* a par exemple été intercepté dans du bois de palette à destination de la Nouvelle-Calédonie (Mille et al. 2015). La forte longévité des adultes (2-3 mois ; Hellrigl 1971, Naves et al. 2006, Koutroumpa et al. 2008, Torres-vila et al. 2015) favorise de tels événements de dispersion, car les individus sont susceptibles de survivre à des transports de longue durée, par containers par exemple. Il n'existe en revanche pas de travaux mettant en évidence l'existence de corridors de transport de *Monochamus sp.* sur de longues distances.

I.3.5 - Méthodes de lutte et réglementation

Le PWN est classé organisme de quarantaine par l'Organisation Européenne et méditerranéenne pour la Protection des Plantes depuis 1985 (OEPP/EPPO, 2014). Des mesures phytosanitaires ont été imposées aux états membres pour limiter de manière préventive l'expansion du PWN par le biais du commerce du bois (décisions d'exécution de la Commission Européenne 2006, 2009, 2012). Une fois détecté, le PWN est contrôlé par une coupe rase (abattage et destruction systématique des pins) dans une zone variant de 500 m à 3 km de rayon autour de la zone infestée (OEPP/EPPO 2012) et par la surveillance accrue au sein d'une zone tampon d'un rayon allant de 6 à

20 km. Cependant, l'application de ces méthodes est rendue difficile par l'ampleur des zones à couper et à surveiller, par les coûts impliqués ainsi que par la difficulté de leur mise en place dans certaines zones (réserves naturelles en particulier). Il faut également noter que ce protocole a été développé en l'absence de littérature sur les modes et des distances de dispersion du PWN, il correspond donc à une vision relativement subjective de ce potentiel. L'efficacité des coupes rases elles-mêmes fait d'ailleurs l'objet de remises en cause (Jactel et al. 2015). Enfin, le développement tardif et la non-spécificité des symptômes de la PWD sur les pins rendent la détection visuelle précoce de nouveaux foyers difficile, et impliquent souvent une certaine inertie dans la mise en place de ces mesures (Naves et al. 2007a).

Des tentatives de lutte biologique contre le PWN existent et se basent sur l'utilisation de champignons (Wang et al. 2011) ou d'huiles essentielles (Faria et al. 2013). La lutte chimique par épandage d'insecticide (Emamectin benzoate) s'est révélée efficace pour contrôler la PWD aux USA dans les plantations de *P. sylvestris* (James et al. 2006) et au Japon (Takai et al. 2003). L'injection de Emamectin benzoate dans les troncs permet également de protéger les arbres de la maladie au Portugal (Sousa et al. 2011, 2013). Ces méthodes montrent une certaine efficacité dans la lutte contre la PWD. Cependant, la viabilité de leur application à large échelle et les dommages environnementaux associés à la non-spécificité de cet insecticide restent des freins importants à leur utilisation. Les méthodes de contrôle des populations de vecteurs par lutte biologique et piégeage de masse sont peu prometteuses vu l'absence d'ennemis naturels spécifiques et efficaces (Petersen-Silva et al. 2014) et les très fortes densités de ces insectes en milieu naturel (Torres-vila et al. 2015). L'utilisation de filets imprégnés de alpha-cyperméthrine pour éviter l'émergence des imagos infectés lors du transport de bois rond est en cours d'évaluation (Naves et al. 2013). Enfin, un programme de sélection de pins résistants à la PWD a été initié (Ribeiro et al. 2012), mais le différentiel entre la vitesse d'expansion du PWN et l'inertie des programmes de sélection laisse peu d'espoir qu'une telle approche limite les dégâts associés à l'expansion de la PWD.

L'ensemble de ces méthodes de lutte et la difficulté de leur application montre un réel besoin de mieux comprendre les mécanismes de dispersion de *M. galloprovincialis* et du PWN afin d'optimiser le contrôle de son expansion.

I.4 - Objectifs de la thèse

L'objectif général de cette thèse est d'estimer le potentiel dispersif du nématode du pin au travers de son association phorétique nouvellement formée avec *Monochamus galloprovincialis* en Europe. Cette estimation requière une étude approfondie de l'histoire évolutive et des capacités de dispersion du vecteur natif, mais également de l'effet synergique de cette association nouvellement formée sur la dispersion. Les multiples aspects de ce modèle m'ont poussé à adopter une approche intégrative, à l'interface entre les sphères de la biologie évolutive et la biologie de l'invasion, auxquelles j'ai emprunté et adapté les outils respectifs.

Dans un premier temps, j'ai cherché à caractériser la structure génétique du vecteur natif et à identifier les facteurs qui déterminent cette structure à l'échelle de l'Europe. L'objectif de ce premier travail était de replacer l'espèce dans son contexte macro-évolutif afin de rechercher les facteurs pouvant affecter sa dispersion à large échelle (adaptation locale, processus de spéciation pour une essence hôte). Dans un second temps, j'ai estimé la dispersion de ce vecteur en milieu hétérogène à l'échelle de la péninsule ibérique. Pour cela, j'ai analysé les flux de gènes de cette espèce et leur corrélation avec plusieurs paramètres environnementaux afin d'identifier les barrières et les corridors à sa dispersion. Enfin, j'ai modélisé la dynamique spatio-temporelle des populations de vecteurs infectés par le nématode à travers la barrière pyrénéenne, en incluant non seulement la dispersion du vecteur mais aussi l'effet synergique de cette association phorétique.

Les associations phorétiques nouvellement formées sont des cas d'invasion relativement rares et peu étudiés. L'objectif de cette thèse a donc également été d'apporter des éléments sur le modèle nématode - *Monochamus* pouvant être appliqués à d'autres situations.

Liste des articles publiés ou en préparation relatifs à cette thèse

Article n°1: Haran J, Rousselet J, Roques A, Roux G, (*in prep*). Phylogeographic patterns of *Monochamus galloprovincialis*, the vector of the pinewood nematode in Europe. Which factors contribute to current distribution of intraspecific lineages?

Article n°2: Haran J, Rossi JP, Pajares JA, Bonifacio L, Naves P, Roques A, Roux G, (*in prep*). Combining multi-scale and multi-site replication in spatial genetics: implications for highly dispersive insect species.

Article n°3: Haran J, Roques A, Bernard A, Robinet C, Roux G, 2015. Altitudinal barrier to the spread of an invasive species: could the pyrenean chain slow the natural spread of the pinewood nematode? *PLoS ONE* 10(7): e0134126.

Article n°4 (Annexe n°1): Haran J, Koutroumpa F, Magnoux E, Roques A, Roux G, 2015. Ghost mtDNA haplotypes generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and phylogeographic pattern. *Journal of Zoological Systematics and Evolutionary Research* 53, 109-115.

Article n°5 (Annexe n°2): Haran J, Roux G, 2014. Development of 12 microsatellites loci for the longhorn beetle *Monochamus galloprovincialis* (Coleoptera Cerambycidae), vector of the pinewood nematode in Europe. *Conservation Genetics Resources* 6, 975-977.

Chapitre II

Diversité et structure génétique du vecteur dans son aire native

Chapitre II : Diversité et structure génétique du vecteur dans son aire native

Ce premier chapitre repose sur une exploration de la diversité et de la structure génétique de *M. galloprovincialis* au sein de son aire de distribution. Afin d'explorer au mieux cette structure, nous avons tout d'abord évalué la pertinence des marqueurs génétiques disponibles pour ce type d'étude.

II.1 – Pertinence des marqueurs génétiques pour l'étude de la structure génétique de *M. galloprovincialis*

Comme indiqué dans la section I.1.4B de l'introduction, les outils génétiques sont largement utilisés pour étudier les patrons génétiques des espèces et inférer les flux de gènes. Bien qu'il existe un large éventail d'outils disponibles, ces études sont en général réalisées à partir de deux types de marqueurs génétiques neutres (*i.e.* non soumis à sélection) : les marqueurs mitochondriaux (séquences de gène ou de fragments de gène issus du génome mitochondrial) et les marqueurs microsatellites (courtes séquences d'ADN nucléaire constituées de répétitions de motifs composés de 2 à 10 nucléotides). Chacun de ces marqueurs a une vitesse d'évolution et un mode de transmission propre, apportant une information complémentaire sur la résolution des patrons génétiques, ainsi que sur leur origine. Nous avons donc cherché à étudier la structure génétique et des flux de gènes chez *M. galloprovincialis* à partir de ces deux types marqueurs. Cette première section présente les biais relatifs à l'utilisation des marqueurs mitochondriaux chez cette espèce, ainsi que les marqueurs microsatellites spécifiques développés au cours de cette thèse.

Dans un premier temps, nous nous sommes basés sur les travaux réalisés par Koutroumpa et al. (2009, 2013), afin de réévaluer la pertinence du gène mitochondrial codant pour la cytochrome oxydase 1 (COI) pour l'étude de la structure génétique de *M. galloprovincialis*. La séquence de ce gène est largement utilisée pour la reconstruction de la structure génétique des espèces à large échelle,

en particulier pour les phylogéographies (Galtier et al. 2009). Cependant, chez *M. galloprovincialis*, ce gène présente des paralogues intégrées au sein du génome nucléaire (Koutroumpa et al. 2009). Ces pseudogènes (ou NUMTs, pour « nuclear mitochondrial pseudogènes ») peuvent poser problème dans la reconstruction des lignées évolutives chez cette espèce en raison de la confusion qu'ils entraînent avec la copie « authentique » du gène utilisé (Bensasson et al. 2001). Nous avons donc cherché à évaluer la fréquence des NUMTs chez *M. galloprovincialis*, ainsi que leur impact sur les hypothèses sous-jacentes aux reconstructions phylogéographiques. Afin de placer cette étude dans un contexte plus large, nous avons également évalué notre capacité à détecter ces artefacts dans notre jeu de données. Nos résultats ont montré que les NUMTs ne constituaient pas un phénomène anecdotique et qu'ils pouvaient représenter une fraction très importante d'un jeu de données. Nous avons également montré que les NUMTs pouvaient être très fréquents au niveau intra-individuel, et présenter plusieurs niveaux de divergence avec la copie mitochondriale du gène. Cette étude nous a permis de montrer que les méthodes couramment employées pour identifier ces artefacts sont majoritairement inefficaces pour détecter une fraction importante de NUMTs, en raison de la forte prévalence de NUMTs cryptiques. Les NUMTs cryptiques correspondent à des paralogues dont la faible divergence (quelques mutations seulement) ne permet pas leur détection avec les méthodes employées couramment, telles que la recherche de codons stop, de mutations non synonymes ou encore de décalages de cadre de lecture (Bensasson et al. 2001, Song et al. 2008, Baldo et al. 2011). Enfin, cette étude nous a permis de montrer que les NUMTs peuvent avoir un fort impact sur la reconstruction de la phylogéographie d'une espèce. La présence de ces haplotypes paralogues surnuméraires conduit à un brouillage total du signal phylogéographique, et à une incohérence dans la distribution spatiale des haplotypes. Cette étude, au-delà des aspects fondamentaux relatifs à la prévalence et à la complexité des NUMTs, montrent que l'utilisation des marqueurs mitochondriaux n'est pas pertinente pour l'évaluation de la structure génétique chez *M. galloprovincialis*. Ces marqueurs n'ont donc pas été utilisés pour caractériser les lignées évolutives de cette espèce. L'article méthodologique relatif à cette étude est joint en annexe n°1 de ce manuscrit.

Annexe n°1: Haran J, Koutroumpa F, Magnoux E, Roques A, Roux G 2015. Ghost mtDNA haplotypes generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and phylogeographic pattern. *Journal of Zoological Systematics and Evolutionary Research* 53, 109-115.

Les microsatellites sont des marqueurs hautement polymorphes permettant d'étudier la diversité et la structure génétique des populations d'une espèce à une échelle spatiale relativement fine. En raison de ce fort polymorphisme, ce type de marqueurs est également utilisé pour identifier les flux de gènes ainsi que les événements de migration d'individus sur de longues distances. De part ces caractéristiques, les microsatellites sont adaptés à la résolution des questions soulevées dans cette thèse. Nous avons donc développé un ensemble de 12 loci microsatellites spécifiques à *M. galloprovincialis*. Ce set a été sélectionné à partir d'un pool de 48 loci identifiés par pyroséquençage afin de ne retenir – à partir de la variation au sein de 14 individus issus de l'ensemble de l'aire de distribution de cette espèce – que les loci polymorphes et une spécifiques (copie unique dans le génome). Ces loci ont ensuite été testés sur une population de 30 individus (Orléans, France) afin d'évaluer leur polymorphisme au sein d'une population et leur pertinence chez cette espèce. Les résultats montrent que les loci développés ne présentent pas de déséquilibre de liaison significatif. Un excès d'homozygotes a cependant été détecté pour quatre loci, ce qui laisse supposer que ceux-ci pourraient être affectés par un ou plusieurs allèles nuls au sein de la population étudiée. Les allèles nuls correspondent à un polymorphisme dans les séquences flanquantes des loci qui empêche leur amplification par PCR. Ces allèles non lus peuvent augmenter artificiellement le taux d'homozygotes et déséquilibrer certaines analyses de génétique des populations. Ce type d'artefact est très répandu chez les marqueurs microsatellites (Dakin & Avise 2004) et leur fréquence peut varier spatialement. La prévalence de ces allèles au sein des 12 loci sera donc prise en compte pour chacune des études de ce manuscrit afin d'éviter les biais que ceux-ci peuvent induire. Cette étude a également montré que les loci ont un polymorphisme modéré au sein de la population étudiée, variant de 2 à 8 allèles par locus, pour une moyenne de 3.75 par locus. Ce polymorphisme modéré a également été observé chez *M. alternatus* (Shoda-Kagaya 2007), ce qui suggère que ce caractère est partagé au sein du genre.

Malgré ce faible polymorphisme, et un nombre plus restreint de loci, l'étude menée par Shoda-Kagaya et al. (2007) a permis de mettre en évidence une structure génétique chez *M. alternatus* au Japon. Dans les études de ce manuscrit, nous utiliserons donc le jeu de 12 loci pour étudier la structure génétique à large échelle, mais également à échelle locale pour inférer les flux de gènes. La note méthodologique correspondant au développement et à la validation des 12 loci spécifiques à *M. galloprovincialis* est jointe en annexe n°2 de ce manuscrit.

Annexe n°2: Haran J, Roux G 2014. Development of 12 microsatellites loci for the longhorn beetle *Monochamus galloprovincialis* (Coleoptera Cerambycidae), vector of the pinewood nematode in Europe. *Conservation Genetics Resources* 6, 975-977.

II.2 - Phylogéographie de *M. galloprovincialis* : Quels facteurs peuvent expliquer la structure génétique du vecteur à large échelle ?

Dans cette section, nous explorons les facteurs influençant la mise en place de la structure génétique de *Monochamus galloprovincialis*. Pour cela, nous avons reconstruit la phylogéographie des populations Européennes de cette espèce à partir des 12 loci microsatellites développés pour cette espèce. L'objectif de cette première approche était de déterminer les facteurs pouvant affecter la dispersion du vecteur à large échelle, tels que les processus de spéciation ou d'adaptation locale de certaines lignées à des paramètres environnementaux. Cette étude se rapporte à l'article n°1.

Article n°1: Haran J, Rousselet J, Roques A, Roux G (*in prep*). Phylogeographic patterns of *Monochamus galloprovincialis*, the vector of the pinewood nematode in Europe. Which factors contribute to current distribution of intraspecific lineages?

- Contexte et objectifs :

M. galloprovincialis est une espèce largement répandue en Europe, on la retrouve sur une gamme d'altitudes s'étendant de 0 à plus de 1500 mètres, et sur différentes essences hôtes appartenant au genre *Pinus* (Evans et al. 1996, Koutroumpa et al. 2013). Les différences de climats entre régions au sein desquelles cette espèce se maintient et la diversité de ses arbres-hôtes sont susceptibles d'avoir engendré des phénomènes d'adaptation et de spécialisation des populations (Sexton et al. 2014). Ces phénomènes peuvent fortement affecter la dispersion chez *M. galloprovincialis*, en particulier si les adultes de cette espèce volent à la recherche de conditions climatiques particulières ou vers une essence d'arbre-hôte spécifique. Il est donc important de caractériser ces phénomènes, car ils peuvent par extension affecter la dispersion du PWN. En tant qu'espèce native en Europe, *M. galloprovincialis* a probablement subi les cycles de contraction-expansion de son aire de distribution en raison des oscillations glaciaires du quaternaire (Taberlet 1998, Hewitt 2000). Ces cycles ont pu fortement imprimer la distribution des patrons génétiques de l'espèce. En particulier à travers la recolonisation

postglaciaire de populations isolées dans des zones refuges, au sein desquelles elles ont pu subsister durant les maxima glaciaires.

L'objectif de cette étude est d'étudier la distribution des lignées évolutives de *M. galloprovincialis* afin (i) d'identifier les facteurs environnementaux responsables de la structure génétique des populations de cette espèce. La mise en évidence d'une structure génétique à large échelle a également permis de caractériser (ii) l'histoire quaternaire de cette espèce ainsi que (iii) les phénomènes de migration à longue distance d'origine anthropique.

-Matériel et Méthodes :

Cette étude repose sur le génotypage de 1098 adultes à 12 loci microsatellites. Ces adultes ont été capturés à l'aide de pièges et d'attractifs phéromonaux spécifiques en 45 points couvrant une large part de l'amplitude latitudinale, longitudinale et altitudinale de l'aire de distribution européenne de cette espèce, ainsi qu'une large gamme de pins hôtes. Deux points d'échantillonnage ont été placés dans des zones où l'import massif de bois est susceptible d'avoir entraîné des transports à longue distance de ce coléoptère (port et papeterie).

Le niveau de diversité génétique des 45 populations a été évalué en calculant la richesse allélique et le nombre d'allèles privés (Kalinowski 2005, Peakall & Smouse 2012). La différenciation entre populations a été estimée grâce à l'indice de fixation (F_{ST} , Weir, Cockerham 1984). La structure génétique des populations a été inférée par méthode bayésienne (Pritchard et al. 2000) pour caractériser les principales lignées évolutives de *M. galloprovincialis*. Le nombre optimal de regroupements (ou clusters, qui correspondent aux lignées évolutives) a été déterminé par la méthode du Delta-K (Evanno et al. 2005). Enfin, nous avons reconstruit l'arbre phylogénétique des populations afin de reconstruire les relations de parentés entre populations.

Les facteurs affectant la distribution de la structure génétique ont été évalués par régression multi-variée (Analyse par Redondance: RDA). Nous avons utilisé une analyse par redondance partielle (pRDA) pour examiner la proportion de la variation environnementale expliquant la variation des génotypes, tout en « partitionnant » la localisation géographique des populations. Cette analyse a été

réalisée sur un sous-échantillonnage de 24 populations (627 individus) distribuées dans la partie occidentale de l'aire de distribution. Ce sous-échantillonnage a été choisi car la densité de population est plus importante dans cette zone, et les données environnementales disponibles y sont plus précises. Les populations des îles ont été exclues de cette analyse, car leur divergence liée à l'isolement allopatrique présentait un risque de déséquilibre dans l'analyse. Nous avons testé cinq paramètres environnementaux supposés pouvoir affecter la distribution des lignées évolutives : l'élévation, les précipitations, les températures maximales et minimales ainsi que l'essence hôte. Pour cette analyse, l'essence de pin a été codée sur un gradient altitudinal de Thermo-Méditerranéen à oro-Méditerranéen en suivant la classification de Barbérot (1998): (*P. halepensis* 1, *P. pinaster* 2, *P. nigra* 3, *P. sylvestris* 4). Pour évaluer la proportion de variance génétique uniquement expliquée par chacun des facteurs environnementaux, nous avons « partitionné » les composants de la variance. Pour chaque paramètre séparé et pour tous les paramètres ensemble, trois modèles pour RDA ont été construits : un modèle complet incluant toutes les variables environnementales explicatives, un modèle partiel dans lequel la géographie explique les géotypes en partitionnant les données environnementales et enfin le modèle inverse où les données environnementales expliquent les géotypes en partitionnant la géographie. Nous avons ensuite calculé le pourcentage de contribution du paramètre environnemental seul, de la géographie seule et de leur effet commun. La significativité de la relation entre les paramètres environnementaux et les géotypes a été obtenue par un test de permutation.

La présence de migrants de première génération entre localités distantes a été estimée à l'aide de tests d'assignation (Rannala & Mountain 1997). Nous avons testé la probabilité de chaque individu d'appartenir à l'une des 24 populations, en incluant celle d'origine par méthode bayésienne.

-Résultats :

Nous avons observé une richesse allélique moyenne des populations variant de 1.77 (Palerme) à 3.25 (Castellbell) par raréfaction. Globalement, une plus faible richesse a été observée dans les îles et une richesse maximale a été observée au sein des populations du pourtour méditerranéen, en particulier au centre et sur la côte est de l'Espagne. Les mêmes patrons ont été observés pour les

allèles privés. La différenciation entre populations (estimations de F_{st}) variait dans une gamme de 0.00 à 0.482, avec des maxima observés entre les îles et les populations continentales.

Les analyses par regroupement bayésien ont montré plusieurs niveaux de structure de ces populations. En effet, cette méthode d'analyse offre un continuum de divisions en clusters (K). Parmi cet ensemble de divisions, il peut parfois être difficile de définir un nombre optimal de clusters sur lequel baser les interprétations. Nous avons choisi de retenir les nombres de 2 et 8 groupes car ils correspondent aux niveaux de division qui expliquent le mieux la structuration génétique des populations du jeu de données analysé (Delta K les plus élevés), mais aussi car ils correspondent à la topologie de l'arbre des populations basé sur les distances. Le regroupement des individus en deux lignées sépare les populations au niveau des Pyrénées entre la péninsule ibérique (incluant les populations des îles de Madère, de Majorque et la Corse) et le reste des populations européennes. Le regroupement en 8 lignées permet de distinguer 5 lignées principales européennes continentales (Portugal, Galice / Espagne / France ouest et nord / Italie et sud-est de la France / Europe de l'est et du nord). A ce niveau de regroupement, la Turquie, la Corse et la Tunisie forment des groupes indépendants. L'arbre phylogénétique des populations permet de retrouver une structure similaire à celle obtenue par regroupement bayésien. Les relations au sein de cet arbre montrent une forte divergence des populations du Portugal, des îles, de Tunisie et de Turquie vis-à-vis du reste des populations, et une faible divergence entre populations continentales d'Europe hors péninsule ibérique.

L'ensemble des paramètres environnementaux testés a montré une corrélation significative avec la distribution des génotypes par analyse pRDA et expliquent 40.73 % de la part explicable de la variance génotypique. Les précipitations moyennes annuelles constituent le paramètre qui explique le mieux la variance génétique (21.25 %), suivi de l'élévation (15.89 %) et des températures minimales hivernales (13.20 %) et maximales estivales (13.14 %). A l'inverse, l'essence hôte est le paramètre qui semble expliquer le moins la structure génétique chez *M. galloprovincialis* (7.6 %).

Les tests d'assignation n'ont pas permis d'identifier de migrants de première génération au sein des populations « sensibles » incluses dans l'échantillonnage. Nous n'avons pas non plus observé

de signe indiquant l'existence d'une voie de transport d'individus sur de longues distances à l'échelle de l'ensemble de l'aire d'étude.

Phylogeography of *Monochamus galloprovincialis* (Coleoptera, Cerambycidae), the European vector of the pinewood nematode, reveals a climatic origin of genetic patterns.

Julien HARAN^{1,2}, Jérôme Rousselet, Alain ROQUES¹ & Géraldine ROUX^{1,2}

1 INRA, UR633 Zoologie Forestière, F-45075 Orléans, France.

2 Université d'Orléans, France.

Abstract

Monochamus galloprovincialis (Coleoptera, Cerambycidae) is an oligophagous beetle associated with pine trees in Europe. This species is the native vector of an introduced pine forests pathogen: the pinewood nematode (PWN, *Bursaphelenchus xylophilus*) that is expanding its range in Iberian Peninsula. Determining the factors influencing dispersal behavior of *M. galloprovincialis* is a prerequisite to estimate the potential of the PWN to spread across Europe. In this study, we reconstructed a phylogeography of this species to uncover its quaternary history and to determine whether distribution of lineages was associated with host tree specialization or climatic parameters. A number of 1098 individuals from 45 locations throughout the European range of the species were genotyped at 12 microsatellite loci. Phylogeographic patterns were investigated using bayesian clustering methods and phylogenetic reconstruction of populations. The contribution of environmental parameters in shaping genetic structure was tested using redundancy analysis. Our results show that the species is structured into several evolutionary lineages exhibiting a pattern typical of post-glacial recolonisation. We also show that the distribution of these lineages in Western Europe is mainly determined by precipitations and that contrary to expectations, pine species does not seem to play a significant role in shaping genetic structure in this species. It is therefore suggested that the genetic structure observed in *M. galloprovincialis* is mainly derived from glacial oscillations and the distribution of climatic parameters, and that this species is not in a process of specialization for host-tree species.

Keywords: glacial refugia, microsatellite, host specialization, *Pinus sp.*, redundancy analysis.

Introduction:

Quaternary ice ages have played a significant role in shaping distribution of intraspecific lineages of European biota (Taberlet 1998, Hewitt 2000). Series of glaciation interleaved by interglacial periods have led to cycles of contraction-expansion and shifts of species distribution range. For temperate species, glacial maxima resulted in regression of range and persistence of isolate populations in southern refugia. In contrast, warmer periods defined suitable areas allowing northward recolonisation of taxa (Hewitt 2004).

For closely interacting species (host-parasite, prey-specialist predator...), migration pathways are constrained by the co-occurrence of associated taxa above simple climatic parameters (Nicholls et al. 2010). This is particularly true for herbivorous species, where host species distribution and recolonisation patterns are expected to play a significant role in shaping current genetic structure (Borer et al. 2012). In addition, such case may hinder an additional degree of structure due to host species specialization, which may be enhanced by allopatric isolation in refugia (Kohyama et al. 2014). Understanding the relative contribution of climatic and biotic variables in shaping current genetic structure of an herbivorous species can help to identify its biological requirements and its potential specialization to host species. By extension, it may help to highlight the drivers of gene flow, as dispersal is often affected by processes of local adaptation to environmental conditions (Sexton et al. 2014).

In this study, we investigated quaternary history of *Monochamus galloprovincialis* (Coleoptera Cerambycidae). This beetle is the vector of the pinewood nematode (PWN, *Bursaphelenchus xylophilus*), an introduced pest for pine forests that expands its range in Iberian Peninsula since its introduction in Portugal in 1999 (Mota et al. 1999). The PWN causes the pine wilt disease (PWD), which can lead to tree decline within a few months under suitable climatic conditions (Naves et al. 2007, Kuroda 2008). Determining the origin of genetic structure of *M. galloprovincialis* and underlying drivers of gene flow is of primary importance to estimate the potential expansion of the PWN. *M. galloprovincialis* is a native and widespread species in Europe (Koutroumpa et al. 2013). It

performs its life cycle in the wood of stressed or freshly dead pine trees, mainly *P. pinaster*, *P. sylvestris*, *P. nigra*, *P. halepensis* (Hellrigl 1971, Naves et al. 2006, Koutroumpa et al. 2008). Although considered as an oligophagous species, host choice tests in laboratory conditions have highlighted preference for *P. sylvestris* for oviposition (Naves et al. 2006), suggesting a potential host specialization of the population tested. However, it is not known yet whether *M. galloprovincialis* shows such specialization across its range, or if its genetic diversity is determined by other factors.

Koutroumpa et al. (2013) provided the first overview of genetic structure of this species based on the sequences of the mitochondrial gene of the cytochrome oxidase c subunit I (COI). Overall, haplotypes distribution showed an East-West differentiation in Europe, and a possible effect of elevation rather than host tree and geography on the genetic differentiation. However, this study was focused on species delineation between *M. galloprovincialis* and its sister species *M. sutor*, and was subsequently limited in its geographic extent. In particular, sampling was scarce in Iberian Peninsula, an area which is nevertheless of importance in the evolution and divergence of pines associated insects (Rousselet et al. 2010, Tavares et al. 2014). In addition, the dataset was reduced due to the high prevalence of pseudogenes in COI in this species (Koutroumpa et al. 2009), and recent investigations of the complexity of these artifacts tends to confirm the difficulty to base phylogeographic reconstructions on mitochondrial DNA in this species (Haran et al. 2015a). As a result, evolutionary history and lineage distribution of *M. galloprovincialis* in its distribution range remain poorly known. In the context of the PWN invasion, determining the distribution of genetic lineages of this beetle is also a base to track long-distance human-mediated-jumps that can significantly accelerate the spread of the PWN (Robinet et al. 2009). However, the prevalence and the distribution of such migration events have not been explored yet.

We used microsatellite markers to explore genetic variation of *M. galloprovincialis* in its European range. We first (i) reconstructed the distribution of genetic variation and investigated the quaternary history of this species. We then (ii) identified the main drivers in the genetic make-up of this species. Finally, we (iii) discuss the potential use of phylogeographic patterns to track long human-mediated dispersal events in *M. galloprovincialis*.

Material and Methods:

Sampling and genotyping

Adults were trapped using multifunnel traps baited with specific volatile attractant (Galloprotect[®], SEDQ, Spain). Traps were placed in pine stands during summer, from 2011 to 2014 depending on localities. We collected a total of 1098 individuals from 45 localities distributed across most of the European distribution ranges of *M. galloprovincialis* (Table 1). Average number of individuals per localities was 24.4 (range 14-31). Localities sampled covered a wide range of elevations (1-1326 m) and plots of the main host pine species (*P. pinaster*, *P. sylvestris*, *P. nigra*, *P. halepensis* but also *P. peuce* and *P. brutia*), including mixed pin stands. Two sampling points were placed near a port and a wood industry respectively, in order to detect the existence of potential long-distance migrants coming from trade of wood products (Irun, Spain and Tarascon, France; population labels 15, 26). After collecting, samples were stored in ethanol 96 % at 4°C until DNA extraction.

Genomic DNA was extracted from body parts of specimens (legs, thorax, antennae) using a Nucleospin Kit (Macherey–Nagel, Düren, Germany), following manufacturer's instructions. Each individual was genotyped at 12 microsatellite loci (Mon01, Mon08, Mon17, Mon23, Mon27, Mon30, Mon31, Mon36, Mon41, Mon42 & Mon44), following Haran & Roux (2014). Multiplexed PCR were performed in a 10 µL reaction volume using 1 µL of genomic DNA, 0.4 U of DreamTaq DNA Polymerase (Thermo Scientific[®]), 0.75 µL Dream Taq Green Buffer (including 20 mM MgCl₂, Thermo Scientific[®]), 1 µM Betaine, 0.24 µL dNTP (10 µM) and deionized H₂O. PCR amplifications were run on a Veriti[®] 96 well fast Thermal cycler (Applied biosystems[®]) using the following settings: a first denaturation step at 95 °C over 10 min; 40 cycles of denaturation (30 s at 95 °C), hybridization (30 s at 55 °C) and elongation (1 min at 72 °C); and a final elongation step at 72 °C over 10 min. One µL of PCR product was denatured within a mix of 10 µL of formamide and 0.3 µL of 600 Liz marker before being run on an ABI PRISM 3500 sequencer (Life Technologies[®]). Genotypes were read through GENEMAPPER V 4.1 (Applied Biosystems[®]). Primer sequences and multiplex details are displayed in supporting information (see Appendix S1).

Label	collection site	Country	Host species	Long	Lat	Elev	Year	N	A	AR	P	Ho	He	F_{is}
1	Madeira Island	Portugal	<i>P. pinaster</i>	-16,90	32,75	262	2012	14	2,00	1,79	0	0,172	0,214	0,102
2	Santo Andres	Portugal	<i>P. pinaster</i>	-8,77	38,02	43	2013	18	2,45	2	1	0,242	0,291	0,103
3	Castro Daire	Portugal	<i>P. pinaster</i>	-7,92	40,92	602	2013	24	3,73	2,85	0	0,502	0,501	-0,005
4	Ourense	Spain	<i>P. pinaster</i>	-7,54	42,01	659	2013	24	3,64	2,7	0	0,419	0,442	0,051
5	Vale Feitoso	Portugal	<i>P. pinaster</i>	-6,98	40,07	533	2012	24	3,45	2,84	0	0,427	0,461	0,040
6	Villardeciovos	Spain	<i>P. pinaster</i>	-6,31	41,92	867	2011	30	3,73	2,95	7	0,430	0,527	0,194
7	Villalba de Guardo	Spain	<i>P. nigra/sylvestris</i>	-4,82	42,72	1068	2012	25	3,82	2,91	0	0,418	0,497	0,149
8	Santa Maria la Real	Spain	<i>P. pinaster</i>	-4,40	41,07	904	2011	30	5,00	3,19	3	0,456	0,502	0,107
9	Antequera	Spain	<i>P. pinaster</i>	-4,36	37,32	504	2011	27	4,18	2,99	5	0,450	0,521	0,161
10	Murcia	Spain	<i>P. halepensis</i>	-1,13	37,93	132	2011	24	4,36	3,07	1	0,452	0,501	0,113
11	Altura	Spain	<i>P. nigra/pinaster</i>	-0,64	39,82	362	2011	28	4,55	3,08	2	0,492	0,516	0,056
12	Mallorca-Arta	Spain	<i>P. halepensis</i>	3,23	39,70	128	2011	20	2,18	1,9	0	0,252	0,309	0,180
13	Castellbell	Spain	<i>P. halepensis</i>	1,84	41,62	206	2013	30	4,36	3,25	0	0,518	0,579	0,108
14	Huesca	Spain	<i>P. halepensis</i>	-0,72	42,29	601	2013	25	4,36	3,13	1	0,441	0,524	0,181
15	Irun	Spain	<i>P. pinaster</i>	-1,91	43,33	216	2013	18	3,55	2,85	0	0,409	0,447	0,083
16	Vieuzos	France	<i>P. nigra</i>	0,45	43,24	398	2013	28	3,64	2,68	1	0,366	0,417	0,132
17	Marcheprime	France	<i>P. pinaster</i>	-0,85	44,67	52	2013	30	3,27	2,58	0	0,317	0,412	0,222
18	Sare	France	<i>P. nigra</i>	-1,60	43,26	291	2013	30	3,64	2,71	0	0,346	0,457	0,238
19	Vayrières	France	<i>P. sylvestris</i>	1,34	44,61	279	2012	27	3,27	2,61	1	0,336	0,407	0,200
20	Rennes	France	<i>P. sylvestris</i>	-1,71	48,12	26	2013	18	3,27	2,57	0	0,390	0,381	-0,042
21	Senlis	France	<i>P. sylvestris</i>	2,58	49,20	57	2013	24	3,55	2,79	2	0,380	0,469	0,147
22	Orleans	France	<i>P. sylvestris</i>	1,91	47,83	120	2012	30	3,73	2,73	0	0,388	0,436	0,074
23	Dole	France	<i>P. sylvestris</i>	5,49	47,08	205	2013	25	3,55	2,6	0	0,325	0,413	0,284
24	St Beazely	France	<i>P. nigra/sylvestris</i>	3,01	44,18	630	2012	17	3,45	2,72	0	0,327	0,415	0,182
25	Gignac	France	<i>P. halepensis</i>	3,54	43,64	80	2012	30	4,18	2,83	0	0,370	0,421	0,106
26	Tarascon	France	—	4,65	43,77	10	2014	16	3,73	2,82	0	0,344	0,370	0,078
27	Meyrargues	France	<i>P. halepensis</i>	5,52	43,63	248	2004	30	4,18	2,79	0	0,336	0,428	0,203
28	Briançon	France	<i>P. sylvestris</i>	6,62	44,87	1326	2013	22	3,82	2,79	0	0,341	0,395	0,111
29	Conca	France	<i>P. pinea/halepensis</i>	9,33	41,73	399	2012	31	2,55	2,06	0	0,322	0,330	0,023
30	Tunis	Tunisia	—	10,15	36,87	89	2013	23	3,45	2,85	2	0,473	0,482	0,010
31	Palermo	Italy	—	13,36	38,11	75	2012	29	2,09	1,77	1	0,328	0,285	-0,136
32	Tregmago	Italy	—	11,16	45,52	341	2012	24	4,09	2,89	0	0,380	0,421	0,101
33	Monfalcone	Italy	—	13,53	45,80	69	2012	30	4,27	2,82	2	0,361	0,404	0,107
34	Baden	Austria	—	16,23	47,99	352	2013	21	4,73	2,96	3	0,343	0,434	0,193
35	Zadar	Croatia	—	15,19	44,18	21	2011	15	2,73	2,23	1	0,287	0,315	0,051
36	Sarajevo district	Bosnia	—	18,34	43,88	657	2011	18	4,00	2,9	3	0,382	0,435	0,106
37	Belgrade	Serbia	—	20,40	44,71	128	2013	30	4,45	2,94	0	0,386	0,436	0,107
38	Bitola	Macedonia	<i>P. peuce</i>	21,32	41,01	673	2013	24	3,36	2,45	0	0,282	0,332	0,099
39	Yakoruda	Bulgaria	—	23,65	42,00	968	2013	24	3,55	2,62	0	0,341	0,371	0,084
40	Isparta	Turkey	<i>P. brutia</i>	30,55	37,73	467	2014	24	2,73	2,06	1	0,217	0,253	0,102
41		Poland	—	—	—	—	2011	20	3,27	2,39	1	0,325	0,356	0,103
42	Gotska Sandön	Sweden	<i>P. sylvestris</i>	19,19	58,39	18	2011	29	2,45	2,02	1	0,241	0,288	0,132
43	Grasö	Sweden	<i>P. sylvestris</i>	18,46	60,37	1	2011	17	3,18	2,49	0	0,356	0,371	-0,003
44	Lower sawony	Germany	—	—	—	—	2014	27	3,73	2,61	3	0,389	0,394	0,025
45	Hagueunau	France	<i>P. sylvestris</i>	7,79	48,81	157	2013	24	4,09	2,94	4	0,306	0,414	0,249

Table 1: Sampling details of the 45 populations analyzed in this study. Long: Longitude, Lat: Latitude, Elev: Elevation, Year: Year of collecting, N: number of individual, A: average allelic richness, AR: average allelic richness corrected using rarefaction (11 genes), P: number of private alleles, Ho and He: observed and expected heterozygoties, F_{is} : estimates of F_{is} values.

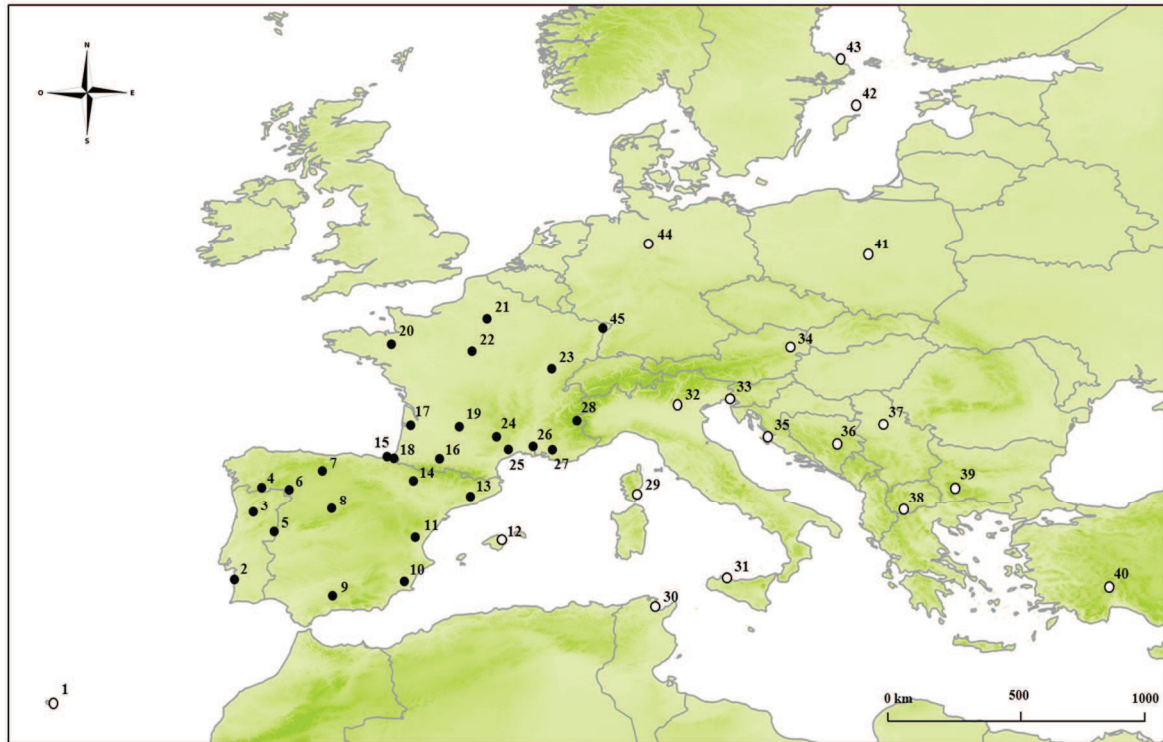


Figure 1: Map of distribution of sampling sites of *M. galloprovincialis*. Black dots correspond to populations used for RDA analysis. Green background refers to elevation.

Microsatellite assessment and genetic structure

We estimated expected and observed heterozygosity (H_e and H_o), departures from Hardy-Weinberg equilibrium (F_{is}) and the number of private alleles using GENAIEX 6.5 (Peakall & Smouse 2012). The mean allelic richness per populations was computed with and without rarefaction using HPRARE (Kalinowski 2005). Deviation from HWE at each locus and linkage disequilibrium between loci were tested using GENEPOP 4.2 (Raymond & Rousset 1995), with a False Discovery Rate Correction (QVALUE, Storey 2002). The frequency of null alleles at each locus and for each population was estimated using FREENA (Chapuis & Estoup 2007).

Population differentiation was estimated using the unbiased Pairwise F_{st} estimates implemented in the program FREENA (Chapuis & Estoup 2007). We used the ENA method to correct the potential positive bias induced by the occurrence of null alleles on F_{st} estimation. Significance of

pairwise differentiation was tested with FSTAT v.2.9.3.2 (Goudet 1995) using 100.000 permutations and a level for significant differentiation set to $p < 0.0001$.

The clustering of populations was made through STRUCTURE 2.3.4 (Pritchard et al. 2000) using the settings “correlated allele frequencies” and “admixture model”, with and without “sampling populations as prior”. The optimal number of cluster (K) was determined using the Delta-K method (Evanno et al. 2005). The genotypes were analyzed for 10 repeats (300.000 burn-in, 600.000 MCMC), for K values ranging from 1 to 15. We used STRUCTURE HARVESTER (Earl & Von Holdt 2012) to visualize the evolution of Delta-K (mean likelihood of K/ variance of likelihood for the same K) through increasing values of K. We used ArcGis 9.3 (ESRI, Redlands, California, USA) to map the membership of populations to the clusters, and R 3.0.2 (R development Core Team 2012) to display the barplots of membership of individuals.

We reconstructed a phylogenetic tree of populations to study genealogical relationships within the species. Phylogeny was computed using POPTREE2 (Takezaki et al. 2009). Analysis were based on Nei’s standard genetic distance (D_{st}) and the UPGMA reconstruction method for a number of 10 000 bootstrap replicates. Trees were visualized and edited in FIGTREE v1.4.2 (molecularevolution.org).

Environmental drivers of genetic structure

To determine which factors shape the distribution of genetic diversity, we conducted a redundancy analysis (RDA, Legendre & Legendre 2012). RDA is a multivariate regression that examines how much of variation in environmental data explains the variation in genotypes. We used a partial RDA (pRDA) to analyze correspondences between variables independently of the geographic location of populations. Calculations were performed on a subset of 24 populations (672 individuals) from Portugal, Spain and France (Islands not included). Such set of populations was chosen to include areas with the highest sampling density, and where accurate environmental data for populations were available. The pRDA analysis was performed at individual level. Genotypes were organized so that each row represents an individual and each is allele coded as a 0 when absent, 1 when single at locus

(heterozygotes) and 2 for homozygotes. Thus individuals are represented as a linear vector of size n , where n is the total number of alleles encountered in all individuals genotyped in the subset.

We considered five environmental parameters as potential drivers of lineages distribution: elevation (E), maximal and minimal temperatures (T_{max} & T_{min}), precipitations (Pr), and pine species (Psp). Elevation was tested because dispersal of *M. galloprovincialis* is known to be affected by topography (Torres-Vila et al. 2015). Therefore, it is suggested that populations from downhill and uphill may be differentiated from each other. In addition, this species exhibit a chromatic variant (*Monochamus galloprovincialis pistosus*) distributed in high latitudes and elevated areas that could represent a distinct evolutionary lineage (Koutroumpa et al. 2013). Maximal and minimal temperatures were tested because the larval development of this beetle occurs mainly during winter, and it has been shown that instars may stop their development and eventually die after exposition to extended periods of cold temperatures (Naves & Sousa 2009). Such parameter may determine the distribution of lineages potentially adapted to cold or warm temperatures. Precipitations were also tested because there is an indication that the distribution of some species of *Monochamus* is driven by this parameter (Kutywayo et al. 2013). Finally, we tested the effect of pine species, because *M. galloprovincialis* is an oligophagous species showing feeding and oviposition preferences for host-tree species (Naves et al. 2006). Therefore, it can be expected that this species is in a process of specialization for host taxa. Temperature and precipitation data were first optimized to determine the combination of data that best explain the genetic variation. We performed a pRDA between genotypes and all combinations of monthly data (each month separate, seasonal mean, annual mean) to determine the best explanatory combination for further analysis. Pine species were coded on an altitude gradient from Thermo-Mediterranean to Oro-Mediterranean following Barb  rot (1998): *P. halepensis*: 1, *P. pinaster*: 2, *P. nigra*: 3 and *P. sylvestris*: 4. This coding system was chosen because it allowed to account for potential specialization for pine groups above the simple species level (*P. halepensis* and *P. pinaster* as thermophile species versus *P. nigra* and *P. sylvestris* as orophile). In addition, it permitted to take in account mixed pine plots. Temperatures and precipitation data (1950-2000) were downloaded from

(Hijmans et al. 2005; <http://www.worldclim.org>). Host pine species associated with populations was estimated based on the dominant tree species in the perimeter of the trapping site (Table1).

To evaluate the ratio of genetic variance uniquely explained by each environmental parameter, we partitioned variance components. For each separate parameter and for all parameters together, we constructed 3 models for RDA: a full model including all environmental and geographic explanatory variables, a partial model in which geography explains genotypes partialling out environmental data and the reverse partial model in which environmental data explains genotypes partialling out geography. We then calculated the percentage of contribution of environmental parameters and geography and joint effect of both to the total explainable variance. Significance of relationship between climate and genotypes was made using permutation test. Calculations were made in R (R development Core Team 2012) using the vegan package.

Detection of human-mediated dispersal

We used assignment tests implemented in GENECLASS 2.0 (Piry et al. 2004), to detect potential long-distance human-mediated transport of individuals of *M. galloprovincialis* across Europe. We tested the likelihood of all individuals to belong to any of the 45 populations tested, including the one of origin ($Likelihood (L) = L_{\text{population of origin}} / L_{\text{maximum in all populations}}$) using bayesian method (Rannala & Mountain 1997). Calculations were performed based on the Monte Carlo resampling method (Paetkau et al. 2004), using 10,000 simulated genotypes. Threshold for migrants detection was set to $p < 0.001$.

Results:

Genetic diversity and distribution of clusters

Number of alleles among loci ranked from 2 (Mon01) to 20 (Mon30) over all populations analyzed. Average allelic richness of population ranged from 1.77 (Palermo, pop. label 31) to 3.25 (Castellbell, pop. label 13) under rarefaction. Overall, allelic richness was the lowest in islands

(Conca, Madeira, Mallorca, Palermo, Gotska Sandom, pop. labels 29, 1, 12, 31, 42) and highest in Spanish populations (central Spain and along the Mediterranean coast). Private allelic richness was also higher in those populations. We found a significant departure from HWE equilibrium in favor of heterozygote deficit for Mon01-27-30-35-42-44 over all populations ($p < 0.001$). Estimated null allele frequency was exceeding 30 % for Mon01 and Mon27, so these loci were not considered for further analysis. We observed a moderate null allele frequency for Mon30 (17.6 %), Mon42 (13.6%) and Mon44 (14.5 %). Other loci were affected by less than 10 %. Pairwise F_{st} among the 45 populations ranked from 0.00 (Castro Daire/Ourense, pop. labels 3, 4) to 0.482 (Conca/Isparta, pop. labels 29, 40). Unbiased F_{st} estimates and their hierarchical distribution are displayed in supporting information (see Appendix S2 and Appendix S3 respectively).

Based on STRUCTURE results, we retained the numbers of 2 and 8 clusters because they gave the best fit to the dataset analyzed (ΔK K2= 498.07, K8=58.93, see Appendix S4 in supporting information). The clustering for K=2 divided samples between the populations from the Iberian Peninsula and several islands (Madeira, Mallorca and Corsica, pop. labels 1, 12, 29) and the rest of populations (Figure 2A). The distribution of populations of the two clusters formed a contact zone along the Pyrenean chain. For K=8, populations grouped into five major continental lineages spread into (I) the Western Iberian Peninsula (Portugal), (II) the Eastern and central Iberian Peninsula, (III) the Western and Northern part of France, (IV) the South-East of France and North Italy, and (V) an Eastern European lineage (including the Balkans, Central Europe and the South of Scandinavia). Populations from Turkey, Tunisia and Western Mediterranean Islands (Mallorca and Corsica, pop. labels 12, 29) formed three additional distinct clusters (Figure 2B and 2C). Phylogenetic relationships within European populations exhibited a consistent arrangement (Figure 3). Spanish and Portuguese populations (including Madeira, pop. labels 1) formed distinct and genetically distant clusters to all other continental populations. Populations from Mediterranean islands (Palermo and Corsica, pop. labels 31, 29), Tunisia and Turkey (Isparta, pop. label 40) formed also distant branches to all other populations. Branches support were high in some deep nodes (Western Iberian Peninsula and Corsica versus all others populations) and generally weak for others splits.

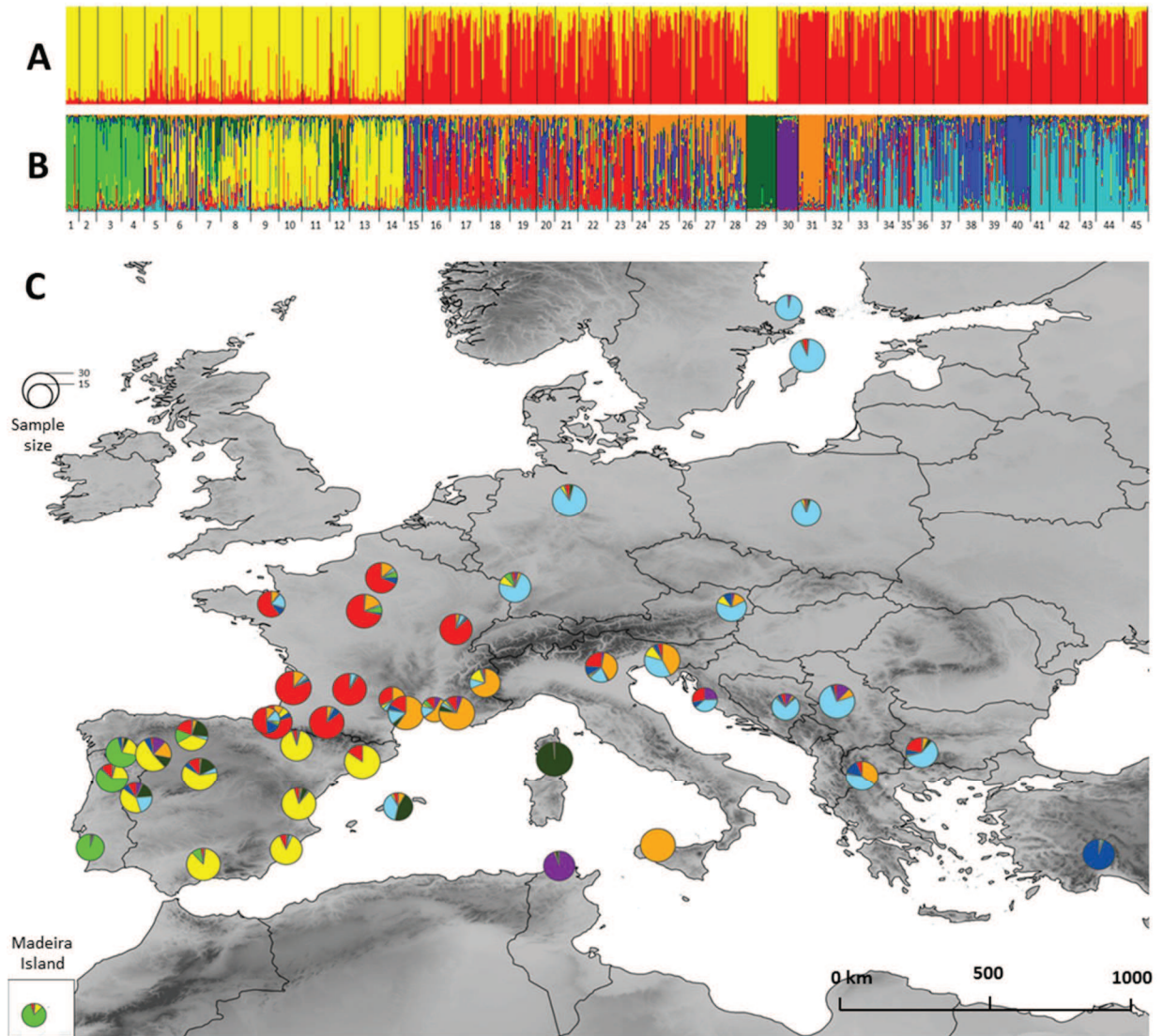


Figure 2: Genetic clusters inferred from STRUCTURE. **A** and **B**: barplots of membership of individuals for $K = 2$ and $K = 8$ (model “no sampling population as prior”). **C**: Membership of populations (model “sampling population as prior”) for $K = 8$.

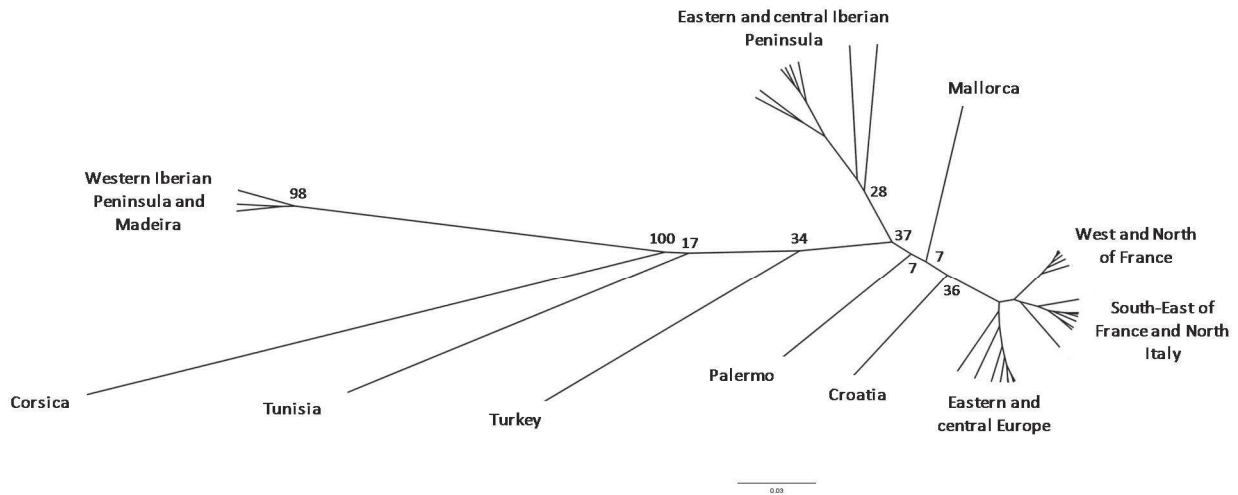


Figure 3: Unrooted tree of populations based on Nei's standard genetic distance (D_{st}) and the UPGMA reconstruction method for a number of 10 000 bootstrap replicates.

Redundancy analysis

The subset of 672 individuals used for redundancy analysis gave a total number of 86 alleles. Monthly temperature and precipitation data were first analyzed with partial RDA to select the combination optimizing correlation with genotype variation. The mean minimal temperatures of winter months (December, January and February) and the mean maximal temperature of January, February and March were the combination best explaining genetic variation. For precipitations, all monthly data had similar correlation with genotypes; consequently, we used the mean annual precipitation for analysis. All parameters tested were moderately but significantly correlated to the genotypes (Table 2). The ratio of explainable variance uniquely explained by each parameter varied. Precipitations gave the best ratio of explainable variance (21.25 %), followed by elevation (15.89 %), mean minimal temperatures of winter (13.20 %) and mean maximal temperature (13.14 %). Pine species was the least important parameter in shaping genetic structure (7.6 %). The combination of all environmental parameters (full model) explained 40.73 % of genetic variation.

Models	Full model		Precipitations		Elevation	
	δ^2	%	δ^2	%	δ^2	%
total variance explainable	1.520*	—	1.149*	—	1.076*	—
pure climate	0.615	40.728	0.244	21.253	0.171	15.892
pure geography	0.187	12.384	0.800	69.625	0.677	62.983
joint climate/geography effect	0.718	47.549	0.104	9.120	0.227	21.124
Models	Min. winter temperatures		Max. summer temperatures		Pine species	
	δ^2	%	δ^2	%	δ^2	%
total variance explainable	1.043*	—	1.042*	—	0.979*	—
pure climate	0.137	13.202	0.136	13.138	0.074	7.603
pure geography	0.864	82.857	0.914	87.744	0.703	71.812
joint climate/geography effect	0.041	3.940	-0.009	-0.882	0.201	20.583

Table 2: Results of RDA analysis. Contribution of each environmental factor to the explainable part of the genetic variance. δ^2 : variance, %: proportion of total explainable variance, *: significance, p -val < 0.001.

Detection of migrants

Using assignment test under Genclass, we detected a number of 30 first generation migrants (see Appendix S5 in supporting information). About 50% of migration events were detected within the 8 lineages identified by clustering methods. The remaining half of migrants was detected between lineages, but without specific trend. Regarding the two populations selected for their exposure to long-distance human transports (Irun, Spain and Tarascon, France, pop. labels 15, 26), we have not observed first generation migrants in significant number, and the occurrence of these few migrants was not higher than in others neighbor populations.

Discussion:

The genetic structure of European biota, as a consequence of quaternary climatic oscillations, has been widely investigated. However, few studies have addressed the question of the environmental origin of lineages distribution, above the effect of historical processes. In this study we provided an example of phytophagous organism for which genetic structure may be determined either by host taxa or climatic conditions. We found that this species exhibit patterns consistent with post glacial recolonisation of Europe after last glacial maxima. We also showed that precipitations are more prevalent than host taxa in shaping genetic structure.

Geographical distribution of genetic lineages

Two distinct lineages in Iberian Peninsula. The genetic structure of continental European populations of *M. galloprovincialis* consisted in five main lineages, two being spread in Iberian Peninsula and three in the rest of Europe. Genetic lineages formed several contact zones along the main mountain systems (Pyrenees and Alps) and along the border between France and Germany. Such arrangement is consistent with general patterns of genetic structure of European biota (Taberlet et al. 1998), and with the genetic structure locally investigated in *M. galloprovincialis* in previous study (Koutroumpa et al. 2013, Haran et al. 2015b). We observed in Iberian Peninsula a clear East-West distribution of clusters, associated with a deep genetic divergence. East-West polarity is known to shape genetic structure and patterns of diversity of organisms in Iberian Peninsula (Gomez & Lunt 2006, Garcia-Barros et al. 2002), in particular through the existence of multiple glacial refugia and important mountain systems in the center of this area. This pattern has also been documented for pine associated species such as *Thaumetopoea pityocampa* (Rousselet et al. 2010) and *Iberorhyzobius rondensis* (Tavares et al. 2014). The case of the moth *T. pityocampa* is particularly interesting to relate to our results, because this species is also oligophagous on pine species and may therefore share similar patterns with *M. galloprovincialis*. The East-West differentiation in *T. pityocampa* formed a contact zone between lineages near the Eastern coast of Iberian Peninsula, and the existence of an ancient barrier to gene flow due to unsuitable environment was suggested to explain such pattern

(Rousselet et al. 2010). In *M. galloprovincialis*, however, the contact zone occurred in the Western side of the Peninsula, and does not correspond to any obvious geographic boundary (mountain range, current lack of host trees). This structure suggests that other factors are active or acted on gene flow in this area for this species compared with *T. pityocampa*. Interestingly, the distribution of the Western Iberian lineage corresponds to the range of a specific lineage of *Pinus pinaster* (Bucci et al. 2007). Over all populations, the mean allelic richness was highest in Central and Eastern Iberian Peninsula, which corresponded also to the areas where the maximum number of private alleles was observed. This part of Iberian Peninsula, and in particular the Eastern coast, is an important glacial refugium for several pines species that persisted in this area during the last glacial maximum (Carrion 2002, Gomez et al. 2005, Fady 2012). It is therefore expected that these areas represented important glacial refugia for *M. galloprovincialis* in Iberian Peninsula. However, contrary to other recolonisation routes highlighted in pine associated species, this lineage did not contribute to the recolonisation of northern latitudes in Europe (Burban et al. 1999, Rousselet et al. 2010), indicating that the Pyrenean chain (including the Western hillside) represented a more important barrier for dispersal for *M. galloprovincialis*.

Three main lineages in the rest of Europe. The three others continental lineages were less differentiated from each other than with lineages found in Iberian Peninsula, suggesting a more recent origin of their divergence. The East-West distribution of lineages confirms the finding of Koutroumpa et al. (2013), based on mitochondrial haplotypes distribution. However, we observed a genetic cluster overlapping the South-East of France and Italy that was not reported in this previous study. Over the three continental European lineages, we generally found higher allelic richness in southern location than in northern ones. Although variations were moderate, such pattern fits the southern richness and northern purity scheme, known for species experiencing successive bottlenecks during last northward postglacial expansion (Hewitt 1999). Such pattern is in agreement with the recent evolutionary of host pines species of *M. galloprovincialis* that were mainly concentrated in glacial refugia in Mediterranean peninsulas and expanded their ranges northward during Holocene (Fady 2012). Based on our results, we suggest that the distribution of Western French cluster corresponds to the expansion from the

refugium of *P. pinaster* (Bucci et al. 2007) described in South-West of France, which is a refugium documented for several other species (Sommer and Nadachowski 2006). It is suggested that the most northern location of France were colonized following pine afforestation in this area in the 19th century (Pines naturally absent in this area, Euforgen 2009). The distribution of the Central and Eastern lineage and its allelic richness suggest that a refugium existed for *M. galloprovincialis* in Balkans, as widely documented for European species (Taberlet et al. 1999, Schmitt et al. 2007). The Italian peninsula was also described as a major glacial refugium, and it is suggested that the lineage distributed in South-East of France and Italy is the result of post glacial expansion from this area. This pattern shows that the southern part of the Alpine arc did not constituted a barrier to contact between lineages as documented for other species (Hewitt 1999, Schmitt 2007). A similar structure was observed in *T. pityocampa* (Rousselet et al. 2010) and associated specialist parasitoid (Auger-Rozenberg et al. 2015) indicating that a continuous corridor of favorable conditions occurred along the Mediterranean coast for pine associated species at least during the last range expansion. However, only the western French and the Eastern European clusters contributed to the recolonisation of northern latitudes, indicating that the Alpine arc and the mid elevations of Massif Central probably constituted barriers to northward range expansion during last warmer period.

Islands and other continents. Populations from Mediterranean islands (Mallorca, Corsica, Sicily) exhibited a strong genetic divergence with continental populations and a poor allelic richness. Such observation is consistent with the hypothesis of an ancient colonization of these areas and subsequent genetic drift due to geographic isolation in a small area. By contrast, the population of Madeira Island was closely related to the Western Iberian lineage, and in particular the population from Ourense ($F_{st} = 0.067$). *M. galloprovincialis* was detected for the first time in 1992 in Madeira, and a Spanish origin of populations was proposed (Erber & Franquinho Aguiar 1996) without precise location. Our results show that the beetle was probably introduced from the west Iberian cluster, comprising Galicia (Spain) and Portugal. The population from Gotska Sandom (Baltic Sea) was also weakly differentiated from the neighbor continental population (Graso), indicating a rather recent colonization of this island, probably occurring with range expansion during Holocene.

North Africa was an important glacial refugium for European species (Husemann et al. 2014), and contributed significantly to the recolonisation of northern latitudes through Gibraltar and Strait of Sicily. Our sampling was insufficient in this area to explore the existence of such refugium for *M. galloprovincialis*. However, the genetic divergence of the population from Tunisia and its high allelic richness suggests that the species probably subsisted there over long periods and during last glacial maxima. The Turkish population was also divergent suggesting, that populations of the species were probably isolated in Near East over long periods.

Altogether, our results show that *M. galloprovincialis* has a genetic structure typical of post glacial recolonisation routes of temperate European species, which is probably explained by retraction-expansion of range of pines trees during glacial oscillations (Fady 2012). The distribution of lineages was consistent with other pine associated species. However, the arrangement of contact zones between lineages and the recolonisation routes of *M. galloprovincialis* were different, suggesting that lineage distribution is constrained by other factors in this species. In Europe, *M. galloprovincialis* exhibits two chromatic variants. *M. galloprovincialis galloprovincialis*, the typical form, is found mainly in lowlands valleys whereas *M. galloprovincialis pistor*, with darker integuments, is inhabiting Eastern Europe and elevated areas of Western Europe (above 800 m, Vives 2000). It is generally admitted that these two variants could represent evolutionary lineages adapted to specific environmental conditions. Although we observed a genetically structured eastern lineage, we did not observe this lineage across elevation in western Europe. Our results confirm findings of previous studies (Koutroumpa et al. 2013, Wallin et al. 2013) and show that such variants probably does not corresponds to a genetically divergent subspecies.

Environmental drivers of genetic structure

When partialling out geographical distances, we found a dominant effect of precipitations to explain the distribution of genetic variation in *M. galloprovincialis*. We also found that such effect is not restricted to a specific season as all monthly precipitation explained similarly variance of

genotypes. Several elements tend to show that precipitations play an important role in shaping distribution of *Monochamus* species. For example, Kutuwayo et al. (2013) showed using distribution models that precipitations is the main factor explaining the distribution of *Monochamus leuconotus* in Africa. However, determining the exact underlying mechanism of this effect is challenging, in particular because precipitations affect directly other parameters such as the growth and the distribution of host trees. Kong et al. (2006) have shown that humidity affects longevity in this genus. It can be proposed that evolutionary lineages are constrained by humidity of the dead wood used as substrate by larval stages, and that this humidity is determined by precipitations. In a lesser extent, we found an effect of elevated areas (and associated temperatures) in shaping genetic structure of *M. galloprovincialis*. Such effect was already observed on the distribution of mitochondrial haplotypes (Koutroumpa et al. 2013). Elevated areas are known to hinder dispersal of *M. galloprovincialis* (Haran et al. 2015b) and *M. alternatus* (Shoda-Kagaya 2007). We show here that such conditions may – in a certain extent – also drive population divergence. By extension, this result suggests a reduced dispersal between populations of elevated and low areas. Pine species showed a weak but significant effect on the genetic make-up of *M. galloprovincialis*. This correlation was not supported in a previous study based on mitochondrial sequences (Koutroumpa et al. 2013). It is though that the weak effect observed may result from covariance with precipitations, as this factor determines distribution of pines tree. For example, the Western Iberian lineage corresponds to important mean precipitations and to the distribution range of *P. pinaster*. Our results suggest that host tree is not an important factor driving genetic structure of *M. galloprovincialis*. This is in agreement with previous studies at local scale (Haran et al. 2015b) and shows that in spite of preferences for some species such as *P. sylvestris* in laboratory conditions (Naves et al. 2006) the species does not seem to be constrained by host species in the wild.

Our results show that distribution of intraspecific lineages is mainly determined by climatic parameters and tends to show that host specialization has not occurred yet in this species. These results support the idea that climatic conditions explain the location of the contact zones identified in *M. galloprovincialis*. The maintaining of such contact zones especially in absence of obvious

geographical barrier suggests that dispersal is limited between areas exhibiting different climatic condition. Such dispersal limitation may be attributed to local adaptation, which is a common pattern in biology (Sexton et al. 2014). Overall, correlations between genotypes and environmental variables were significant but weak. Such weak signal is expected because genetic variation in *M. galloprovincialis* is – as shown above – deeply influenced by differentiation due to quaternary history and so, intrinsically driven by geography (location of glacial refugia).

Human mediated dispersal

In this study, we showed that *M. galloprovincialis* forms several genetic lineages with a quite clear geographic distribution. In particular, we observed an important distance between populations from Iberian Peninsula and the rest of European continental populations. This divergence is a favorable situation to detect long distance migrants between differentiated lineages and we searched for signature of such events. Among the two populations sampled in areas subject to introduction of PWN (Irun, Spain and Tarascon, France, pop. labels 15, 26), we did not find any signature of recent migration events from distant locations. Individuals of these populations clustered with individuals from neighbor populations located in areas non-subject to PWN introduction and individual assignment did not support a specific trend of first generation migrants in those populations. More generally, we did not find a specific pathway of long distance jumps between populations over the whole dataset. This result indicates that such events are absent or not frequent enough to be detected based on our dataset. Long distance dispersal of *M. galloprovincialis* resulting from anthropic activities is a critical parameter in the context of the spread of the PWN, because long-distance jumps of PWN-infected beetles may significantly accelerate the spread of the PWN (Robinet et al. 2009). *Monochamus* species are particularly subject to long distance transportation of larvae and adults through trade of wood logs and wood products from pine trees (Makihara 1998, Mille et al. 2015). Identifying the areas where long distance transports of *M. galloprovincialis* occurs may help to identify the potential pathways of PWN invasion and the development of monitoring programs. The

genetic patterns highlighted in this study constitute a base for future assignment of individuals suspected to come from a distant location through human transport.

Conclusion:

This study aimed at exploring the drivers of the genetic make-up of a xylophagous species associated with several pine species in Europe. Using a sampling of populations at wide scale, we show that climatic parameters are more important in shaping the genetic structure than host plant taxa in *M. galloprovincialis*, at least among its Western distribution range. Such conclusion show that contrary to expectations, host plant taxa does not necessary influence genetic structure of associated insects (Kohyama et al. 2014), even if some preference may exist under laboratory conditions. The genetic structure highlighted in this study also showed that distribution of genetic lineages in this species has likely been influenced by quaternary glacial oscillations. Observed patterns were generally consistent with the ones found in other pine associated species, suggesting a common history of those species with changes in pines distribution over quaternary times. Over all we explain the distribution of genetic lineages of *M. galloprovincialis* as a combination of long term isolation of population in glacial refugia and expansion of these populations in areas with specific climate conditions. In addition, we suggest that contact zones between lineages remained over time due to local adaptation to environment and associated biased dispersal.

In the context of the current range expansion of the PWN, our results show that a climatic boundary limits *M. galloprovincialis* dispersal between the Western Iberian lineage (Portugal and Galicia, where the PWN maintains) and the rest of the Iberian peninsula. This point is interesting because this limited dispersal may slow the natural vector-mediated spread of the PWN in such areas. However, the exact role of such boundaries on the spread of the PWN is difficult to predict because the PWN may spread using a relay of vectors from both sides without dispersal of vectors across the

boundary. The limiting factors of *M. galloprovincialis* dispersal in those areas must be investigated in further details to clarify the potential of such boundaries to prevent PWN dispersal.

Acknowledgments:

We would like to thank all the REPHRAME collaborators and colleagues for making this study possible through the sampling of *M. galloprovincialis*. Our thanks go in particular to J. Pajares, E. de Soussa, P. Naves, L. Bonifacio, M. Glavendekic, H. Jactel, G. David, D. Doichev, M. Ben Jaama, T. Schroeder, G. Hoch, M. Dautbasic, N. Lackovic, M. Schroeder, I. Papazova, R. Tomov, T. Manole, V. Kramarets, I. Matsiakh, G. Csoka, A. Hopf, A. Battisti, M. Avci, A.M. Chappe (ANSES, France). We also wish to thanks the County Administrative Board of Gotland for permission to trap beetles in the National Park of Gotska Sandön. We warmly thank M.A. Auger-Rozenberg for usefull comments on early version of this manuscript.

This work was funded by the European project REPHRAME KBBE.2010.1.4-09 (FP7 Project, Analysis of the potential of the pine wood nematode (*Bursaphelenchus xylophilus*) to spread, survive and cause pine wilt in European coniferous forests in support of EU plant health policy) and the COST grant (COST-STSM-FP1002-14177). First author was funded by the French Ministry of Research and Education

References:

- Auger-Rozenberg MA, Torres-Leguizamon M, Courtin C, Rossi JP, Kerdelhué C, 2015. Incongruent evolutionary histories of two parasitoids in the Mediterranean Basin: influence of host specialization and ecological characteristics. *Journal of Biogeography*. doi:10.1111/jbi.12495.
- Barbéro M, Loisel R, Quézel P, Richardson DM, Romane F, 1998. Pines of the Mediterranean Basin. In D.M Richardson (ed.), *Ecology and biogeography of Pinus*. p. 153-170. Cambridge University Press, Cambridge, Royaume-Uni.
- Borer M, Alvarez N, Buerki S, Margraf N, Rahier M, 2010. The phylogeography of an alpine leaf beetle: divergence within *Oreina elongata* spans several ice ages. *Molecular Phylogenetics and Evolution* 57, 703–709.
- Bucci G, Gonzalez-Martinez SC, Le Provost G, Plomion C, Ribeiro MM, 2007. Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Molecular Ecology* 16, 2137–2153.
- Burban C, Petit RJ, Carcreff E, Jactel H, 1999. Rangewide variation of the maritime pine bark scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. *Molecular Ecology* 8, 1593–1602.
- Carrion JS, 2002. Patterns and processes of Late Quaternary environmental change in a montane region of southwestern Europe. *Quaternary Science Reviews* 21, 2047–2066.
- Chapuis MP, Estoup A, 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24, 621–631.
- Earl DA, VonHoldt BM, 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4, 359–361.
- Erber D, Franquinho Aguiar AM, 1996. New and remarkable species of the coleopterous fauna of Madeira. *Boletim do Museu Municipal do Funchal (Historia Natural)* 48, 41–62.
- EUFORGEN 2009, www.euforgen.org, n.d.
- 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.
- Evans HF, 2014. Rapport final du projet Européen REPHRAME, pp. 57. <http://www.rephrame.eu/pwn.php>.
- Fady B, 2012. Biogeography of neutral genes and recent evolutionary history of pines in the Mediterranean Basin. *Annals of Forest Science* 69, 421–428.
- Garcia-Barros E, Gurrea P, Lucianez MJ, Cano JM, Munguira ML, Moreno LC, et al., 2002. Parsimony analysis of endemism and its application to animal and plant geographical distributions in the Ibero-Balearic region (western Mediterranean). *Journal of Biogeography* 29, 109–124.
- Gomez A, Lunt DH, 2006. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia – evolutionary perspectives on the origins and conservation of European biodiversity* (ed. by S. Weiss and N. Ferrand), pp. 155–188. Kluwer, Dordrecht.
- Gomez A, Vendramin GG, Gonzalez-Martinez SC, Alia R, 2005. Genetic diversity and differentiation of two Mediterranean pines (*Pinus halepensis* Mill. and *Pinus pinaster* Ait.) along a latitudinal cline using chloroplast microsatellite markers. *Diversity and Distributions* 11, 257–263.
- Goudet J, J., 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* 86, 485–486.

- Haran J, Koutroumpa F, Magnoux E, Roques A, Roux G, 2015a. Ghost mtDNA haplotypes generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and phylogeographic pattern. *Journal of Zoological Systematics and Evolutionary Research*. doi: 10.1111/jzs.12095
- Haran J, Roques A, Bernard A, Robinet C, Roux G, 2015b. Altitudinal barrier to the spread of an invasive species: Could the Pyrenean chain slow the natural spread of the pinewood nematode? *PLoS ONE* 10(7):e0134126.
- Haran J, Roux-Morabito G, 2014. Development of 12 microsatellites loci for the longhorn beetle *Monochamus galloprovincialis* (Coleoptera Cerambycidae), vector of the pinewood nematode in Europe. *Conservation Genetics Resources* 6, 975–977.
- Hellrigl KG, 1971. La bionomie des espèces de *Monochamus* (Coleoptera, Cerambycidae) et leur importance pour la sylviculture et l'économie du bois. *Redia* 52, 367–511.
- Hewitt GM, 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359, 183–195.
- Hewitt GM, 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405, 907–913.
- Hewitt GM, 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 68, 87–112.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965–1978.
- Husemann M, Schmitt T, Zachos FE, Ulrich W, Habel JC, 2014. Palaeoarctic biogeography revisited: evidence for the existence of a North African refugium for Western Palaeoarctic biota. *Journal of Biogeography* 41, 81–94.
- Kalinowski S, 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5, 187–189.
- Kohyama T, Matsumoto K, Katakura H, 2014. Deep phylogeographical structure and parallel host range evolution in the leaf beetle *Agelasa nigriceps*. *Molecular Ecology* 23, 421–434.
- Kong WN, Wang H, Li J, Zaho F, Ma RY, 2006. The effects of temperature and humidity on the longevity of *Monochamus alternatus* Hope (Coleoptera: Cerambycidae). *Journal of Shanxi Agricultural University* 294–295 [in Chinese].
- Koutroumpa FA, Lieutier F, Roux-Morabito G, 2009. Incorporation of mitochondrial fragments in the nuclear genome (Numts) of the longhorned beetle *Monochamus galloprovincialis* (Coleoptera, Cerambycidae). *Journal of Zoological Systematics and Evolutionary Research* 47, 141–148.
- Koutroumpa FA, Rougon D, Bertheau C, Lieutier F, Roux-Morabito G, 2013. Evolutionary relationships within European *Monochamus* (Coleoptera: Cerambycidae) highlight the role of altitude in species delineation. *Biological Journal of the Linnean Society of London* 109, 354–376.
- Koutroumpa FA, Vincent B, Roux-Morabito G, Martin C, Lieutier F, 2008. Fecundity and larval development of *Monochamus galloprovincialis* (Coleoptera Cerambycidae) in experimental breeding. *Annals of Forest Science* 65, 707.
- Kutywayo D, Chemura A, Kusena W, Chidoko P, Mahoya C, 2013. The impact of climate change on the potential distribution of agricultural pests: the case of the coffee white stem borer (*Monochamus leuconotus* P.) in Zimbabwe. *PLoS ONE* 8(8): e73432.
- Legendre P, Legendre L, 2012. Numerical Ecology. Elsevier, Amsterdam, The Netherlands.
- Mamiya Y, 1983. Pathology of the Pine Wilt Disease caused by *Bursaphelenchus xylophilus*. *Annual Review of Phytopathology* 21, 201–20.
- Mille C, Cazères S, Sudre J, 2015. Interception de *Monochamus galloprovincialis* (Olivier, 1795) (Cerambycidae, Lamiinae, Monochamini) en Nouvelle-Calédonie. *Bulletin Mensuel de la Société Linnéenne de Lyon* 84, 55–58.

- Mota M., Braasch H., Bravo M.A., Penas A.C., Burgermeister W., et al., 1999. First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. *Nematology* 1, 727–734.
- Naves P, de Sousa E, 2009. Threshold temperatures and degree-day estimates for development of post-dormancy larvae of *Monochamus galloprovincialis* (Coleoptera: Cerambycidae). *Journal of Pest Science* 82, 1–6.
- Naves PM, Camacho S, de Sousa E, Quartau JA, 2007. Transmission of the pine wood nematode *Bursaphelenchus xylophilus* through oviposition activity of *Monochamus galloprovincialis* (Coleoptera : Cerambycidae). *Entomologica Fennica* 18, 193–198.
- Naves P, Sousa E, Quartau J, 2006. Feeding and oviposition preferences of *Monochamus galloprovincialis* for some conifers under laboratory conditions. *Entomologia Experimentalis et Applicata* 18, 99–104.
- Nicholls JA, Preuss S, Hayward A, Melika G, Csoka et al., 2010. Concordant phylogeography and cryptic speciation in two Western Palaearctic oak gall parasitoid species complexes. *Molecular Ecology* 19, 592–609.
- Paetkau D, Slade R, Burden M, Estoup A, 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* 13, 55–65.
- Peakall R, Smouse PE, 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28, 2537–2539.
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudoin L et al., 2004. GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity* 95, 536–539.
- Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Rannala B, Mountain JL, 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the USA* 94, 9197–9201.
- Raymond M, Rousset F, 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86, 248–249.
- R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Robinet C, Roques A, Pan H, Fang G, Ye J et al., 2009. Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS One* 4(2):e4646.
- Rousselet J, Zhao RX, Argal D, Simonato M, Battisti A et al., 2010. The role of topography in structuring the demographic history of the pine processionary moth, *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae). *Journal of Biogeography* 37, 1478–1490.
- Schmitt T, 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology* 4, doi:10.1186/1742-9994-4-1.
- Sexton JP, Hangartner SB, Hoffmann AA, 2014. Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution* 68, 1–15.
- Shoda-Kagaya E, 2007. Genetic differentiation of the pine wilt disease vector *Monochamus alternatus* (Coleoptera: Cerambycidae) over a mountain range – revealed from microsatellite DNA markers. *Bulletin of Entomological Research* 97, 167–174.
- Sommer RS, Nadachowski A, 2006. Glacial refugia of mammals in Europe: evidence from fossil records. *Mammal Review* 36, 251–265.
- Storey JD, 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B* 64, 479–498.
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF, 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7, 453–464.

- Takezaki A, Nei M, Tamura K, 2010. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with windows interface. *Molecular Biology and Evolution* 27, 747–752.
- Tavares C, Roques A, Courtial B, Branco M, Jactel H, Lopez-Vaamonde C, 2014. Phylogeography of the ladybird *Iberorhizobius rondensis*, a potential biological control agent of the invasive alien pine bast scale *Matsucoccus feytaudi*. *BioControl*. doi: 10.1007/s10526-014-9622-7.
- Torres-Vila L, Zugasti C, De-Juan JM, Olivia MJ, Montero C, Mendiola FJ et al., 2015. Mark-recapture of *Monochamus galloprovincialis* with semiochemicalbaited traps: population density, attraction distance, flight behaviour and mass trapping efficiency. *Forestry* 88, 224–236.
- Vives E, 2000. Coleoptera Cerambycidae. In: Museo Nacional de Ciencias Naturales, eds. Fauna Iberica. Madrid: Museo Nacional de Ciencias Naturales.
- Wallin H, Schroeder M, Kvamme T, 2013. A review of the European species of *Monochamus* Dejean, 1821 (Coleoptera, Cerambycidae) – with a description of the genitalia characters. *Norwegian Journal of Entomology* 60, 11–38.

Supplementary material:

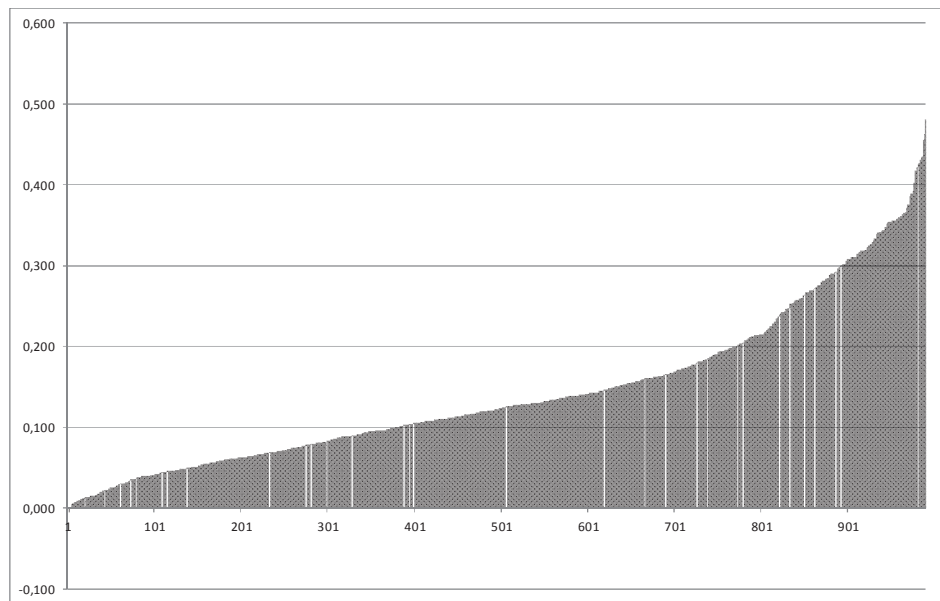
Appendix S1: Detail of primer sequence and multiplexes used in this study.

Locus	Primers sequence 5' -> 3'	Multiplex	Marquage	Amount (pmol)	Motif	Size range
Mon 01	F: TTCACGCACATCATTCTTTG R: TCAAGCAGGAAACGAAAAGC	1	PET	2.4	(aac)5	122-146
Mon 08	F: TGGTGTCTGTAGAACGCTTCA R: GCTTATTAGCTCTCATCAGTATGCTC	3	NED	1	(tatc)5	192-196
Mon 17	F: TAGTTTTACTGGGGCCAATG R: GAACTCATGAACGGATATAAATGAA	3	VIC	1.6	(gt)6	149-153
Mon 23	F: ATTTATTCCAAATTGCCAATACTAC R: GTGTAAGGTGGAAGTGTCAAAGC	1	NED	2	(ca)7	142-144
Mon 27	F: ACAATCTCTTTTCGATACCGTTGA R: TTTGCTACAAAGATGTTCTTAAAAGT	3	NED	2	(tg)7	118-124
Mon 30	F: TTTCTAGTTTGCTGTATCCCG R: AAAGCGGGTGTGAAGTACCA	1	VIC	1.6	(ag)8	236-248
Mon 31	F: GTAAAGAAACCCAACCAACCG R: TTATCCTCACCGGACCGTTA	2	NED	0.64	(ag)8	146-150
Mon 35	F: TTTCATCTGACTTAATTTCTTTCTC R: AGGGACGTGCAGATTAGGAA	2	6FAM	7.2	(tc)8	110-126
Mon 36	F: ACGGTGCAGAACTAAAGTTAGCC R: GACTCGGACGGAGCTTCT	2	PET	0.8	(tg)8	196-200
Mon 41	F: ACGGTAGCGCAA TCTTGAGT R: ACGTGAGCAGTCCTGTTGC	3	VIC	0.8	(ct)9	116-120
Mon 42	F: CCTTATTTAGGTCAAGAATTCGC R: GCGTGTCAATTATCCAAGGAC	1	6FAM	1.6	(gt)9	138-146
Mon 44	F: AACCTGGACCTAGCTCGGAA R: TAGGAGAAGGTGGAGCAGGA	2	VIC	0.64	(act)10	98-113

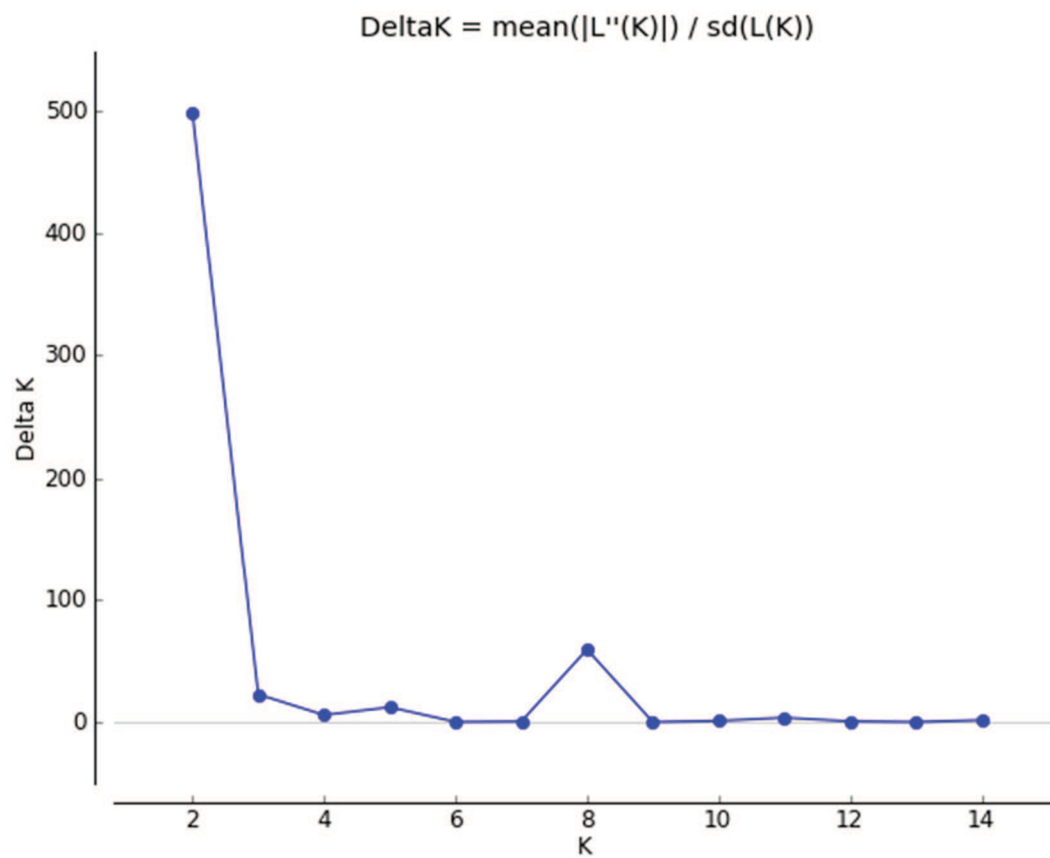
Appendix S2: Pairwise unbiased F_{st} estimates and significant of differentiation. (***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, NS: non significant)

		Madeira island	Santo Andres	Castro daire	Ourense	Vale feitoso	Zamora Villar	Villalba de Guardo	SegoviaSanta Maria	Malaga	Murcia Valle	CuencaAltura	MallorcaArta	Castellbell	Huesca	Irun	Vieuzos	Marcheprime	Sare	Vayrières	Bretagne (Rennes)	Senlis	Orleans	Jura (Dole)	St Beuzely	Gignac	Tarascon	Meyrargues	Briançon	Conca	Tunis	palermo	tregmago	Monfalcone	Baden	Zadar	Sarajevo district	serbia	Bitola	Yakoruda	Isparta	poland	Gotska Sandön	Grasö	Lower sawony	Hagueunau			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45			
Madeira island	1	**																																															
Santo Andres	2	0,127																																															
Castro daire	3	0,079	0,087	NS																																													
Ourense	4	0,063	0,064	0,000																																													
Vale feitoso	5	0,271	0,340	0,127	0,168																																												
Zamora Villar	6	0,284	0,332	0,156	0,200	0,074																																											
Villalba de Guardo	7	0,259	0,283	0,125	0,147	0,050	0,096																																										
SegoviaSanta Maria	8	0,299	0,356	0,171	0,203	0,039	0,061	0,047																																									
Malaga	9	0,276	0,319	0,122	0,159	0,045	0,096	0,075	0,061																																								
Murcia Valle	10	0,308	0,364	0,163	0,204	0,041	0,068	0,064	0,022	0,027		NS		NS	NS																																		
CuencaAltura	11	0,311	0,365	0,165	0,211	0,050	0,056	0,086	0,033	0,042	0,007			NS	NS																																		
MallorcaArta	12	0,322	0,392	0,193	0,231	0,111	0,128	0,113	0,109	0,167	0,112	0,111																																					
Castellbell	13	0,261	0,315	0,131	0,179	0,046	0,055	0,071	0,049	0,031	0,012	0,020	0,127		NS																																		
Huesca	14	0,288	0,350	0,149	0,194	0,036	0,069	0,068	0,029	0,021	0,001	0,017	0,133	0,008																																			
Irun	15	0,228	0,327	0,136	0,180	0,076	0,086	0,130	0,122	0,145	0,120	0,115	0,142	0,076	0,113		NS	NS	NS	NS	*	NS	NS	NS	*	**	*																						
Vieuzos	16	0,285	0,359	0,178	0,218	0,081	0,129	0,121	0,137	0,159	0,145	0,142	0,170	0,101	0,136	0,029		NS	NS	NS	**	NS	NS	NS	**																								
Marcheprime	17	0,269	0,354	0,162	0,201	0,073	0,133	0,117	0,129	0,153	0,130	0,139	0,151	0,104	0,117	0,039	0,015		NS	NS	NS	NS	NS	NS	**																								
Sare	18	0,246	0,320	0,143	0,186	0,068	0,095	0,091	0,103	0,129	0,105	0,111	0,112	0,068	0,105	0,013	0,009	0,014		NS	NS	NS	NS	**	*																								
Vayrières	19	0,290	0,356	0,174	0,212	0,086	0,139	0,107	0,135	0,151	0,132	0,147	0,157	0,097	0,136	0,039	0,007	0,015	0,002		**	NS	NS	NS	*																								
Bretagne (Rennes)	20	0,302	0,402	0,182	0,220	0,116	0,169	0,149	0,153	0,186	0,160	0,171	0,184	0,142	0,143	0,089	0,076	0,026	0,063	0,080		*	NS	**																									
Senlis	21	0,237	0,311	0,127	0,167	0,067	0,111	0,106	0,121	0,111	0,111	0,125	0,151	0,075	0,101	0,029	0,018	0,013	0,009	0,012	0,059		NS	NS	NS	NS																							
Orleans	22	0,230	0,301	0,134	0,166	0,085	0,113	0,102	0,122	0,143	0,130	0,140	0,139	0,101	0,128	0,039	0,023	0,012	0,008	0,011	0,047	0,015		NS	NS	**	*	*																					
Jura (Dole)	23	0,283	0,341	0,161	0,196	0,092	0,140	0,095	0,123	0,143	0,132	0,138	0,174	0,098	0,125	0,063	0,030	0,025	0,024	0,012	0,077	0,017	0,026		NS																								
St Beauzely	24	0,308	0,361	0,147	0,195	0,088	0,101	0,114	0,120	0,124	0,118	0,121	0,162	0,079	0,102	0,061	0,042	0,032	0,030	0,040	0,081	0,016	0,036	0,042		NS	NS	NS	NS																				
Gignac	25	0,290	0,367	0,163	0,216	0,094	0,105	0,146	0,141	0,158	0,128	0,127	0,130	0,110	0,122	0,062	0,074	0,039	0,052	0,083	0,057	0,046	0,056	0,098	0,036		NS	NS	NS																				
Tarascon	26	0,301	0,372	0,161	0,216	0,117	0,117	0,157	0,150	0,178	0,149	0,139	0,116	0,120	0,140	0,058	0,081	0,060	0,053	0,075	0,105	0,071	0,052	0,103	0,046	0,026		NS	**																				
Meyrargues	27	0,265	0,345	0,154	0,198	0,095	0,099	0,143	0,139	0,155	0,139	0,134	0,138	0,114	0,126	0,060	0,077	0,049	0,058	0,087	0,069	0,058	0,041	0,108	0,042	0,011	0,022		NS																				
Briançon	28	0,274	0,376	0,164	0,207	0,105	0,113	0,163	0,145	0,156	0,134	0,128	0,161	0,115	0,121	0,062	0,088	0,046	0,071	0,102	0,056	0,054	0,056	0,106	0,049	0,015	0,054	0,016																					
Conca	29	0,440	0,463	0,310	0,339	0,244	0,231	0,214	0,215	0,281	0,270	0,235	0,263	0,256	0,257	0,318	0,308	0,308	0,289	0,334	0,348	0,309	0,314	0,291	0,310	0,326	0,358	0,318	0,344																				
Tunis	30	0,348	0,355	0,200	0,241	0,162	0,154	0,173	0,172	0,160	0,171	0,178	0,246	0,156	0,165	0,198	0,223	0,209	0,183	0,215	0,223	0,181	0,174	0,226	0,161	0,170	0,178	0,146	0,193	0,360																			
palermo	31	0,433	0,455	0,256	0,301	0,201	0,160	0,195	0,212	0,240	0,213	0,219	0,257	0,198	0,210	0,190	0,174	0,146	0,154	0,168	0,216	0,154	0,143	0,188	0,105	0,096	0,112	0,103	0,139	0,435	0,235																		
tregmago	32	0,272	0,361	0,142	0,195	0,081	0,119	0,125	0,126	0,139	0,127	0,134	0,123	0,095	0,113	0,045	0,059	0,040	0,031	0,051	0,070	0,040	0,038	0,081	0,034	0,028	0,016	0,039	0,050	0,355	0,175	0,141		*															
Monfalcone	33	0,276	0,372	0,158	0,208	0,067	0,109	0,140	0,115	0,126	0,107	0,107	0,130	0,089	0,101	0,031	0,071	0,045	0,040	0,065	0,079	0,036	0,052	0,085	0,043	0,022	0,031	0,036	0,031	0,358	0,177	0,132	0,015		*														
Baden	34	0,267	0,356	0,153	0,196	0,078	0,140	0,138	0,132	0,153	0,133	0,150	0,163	0,119	0,119	0,063	0,088	0,048	0,062	0,088	0,037	0,056	0,065	0,108	0,080	0,042	0,081	0,054	0,064	0,356	0,175	0,182	0,046	0,038															
Zadar	35	0,359	0,418	0,200	0,247	0,117	0,199	0,161	0,179	0,157	0,163	0,198	0,242	0,126	0,163	0,119	0,111	0,103	0,084	0,066	0,150	0,073	0,091	0,110	0,096	0,134	0,137	0,128	0,168	0,430	0,184	0,247	0,094	0,098	0,107														
Sarajevo district	36	0,267	0,362	0,157	0,194	0,060	0,140	0,116	0,109	0,137	0,117	0,136	0,146	0,111	0,111	0,067	0,089	0,051	0,066	0,080	0,057	0,069	0,067	0,099	0,088	0,070	0,092	0,080	0,085	0,321	0,166	0,205	0,051	0,055	0,014	0,095		NS											
serbia	37	0,247	0,340	0,152	0,184	0,067	0,138	0,106	0,097	0,124	0,108	0,130	0,128	0,096	0,106	0,063	0,072	0,051	0,047	0,055	0,052	0,047	0,049	0,077	0,0																								

Appendix S3: Distribution of unbiased pairwise F_{st} values.



Appendix S4: Evolution of DeltaK statistic among an increasing number of clusters (K).



Appendix S5: Assignment tests based on GeneClass analysis. Detection of significant first generation migrants are displayed in red.

		Madeira	Castro daire	Vale feitoso	Villardecievros	Villardecievros	Antequera	Castellbell	Conca	Sarre	Marcheprime	St Beazely	Gignac	Tarascon	Briançon	Hageunau	Gotska	Poland	Croatia	Serbia	Macedonia	Palermo
	Code	1	3	5	6	6	9	13	29	18	17	24	25	26	28	45	42	41	35	37	38	31
SantoAndres	2	3,8	5,1	10,1	12,7	12,3	11,8	13,7	13,2	10,5	11,3	9,7	9,2	9,3	10,9	8,3	12,0	9,9	13,6	9,1	11,2	14,1
Castro Daire	3	4,7	7,3	12,9	14,4	13,6	9,9	11,2	17,4	10,5	11,9	12,2	11,1	10,8	11,2	11,5	17,6	17,6	15,0	13,7	12,9	19,5
Ourense	4	13,7	9,0	15,9	21,0	17,8	14,3	15,3	25,1	15,5	15,6	14,6	12,8	11,5	14,0	10,9	15,4	14,7	16,0	14,1	17,5	18,2
Villardecievros	6	21,6	15,2	14,6	9,6	9,6	11,9	8,1	15,7	10,3	12,8	9,1	13,1	15,5	11,3	17,4	23,8	16,9	16,4	15,3	13,4	16,3
Villardecievros	6	10,7	8,8	9,0	8,0	8,3	14,1	9,0	15,3	8,5	9,8	10,6	9,1	9,4	6,7	5,9	7,9	8,0	10,6	9,0	7,9	16,9
Villardecievros	6	14,6	10,9	9,5	8,9	10,0	13,6	12,4	7,1	12,4	14,7	13,9	15,0	15,6	17,8	13,8	14,8	12,1	18,3	14,1	15,5	21,2
Altura	11	18,1	15,8	17,0	10,3	11,4	15,2	12,4	20,3	14,0	17,8	18,0	16,6	17,3	15,8	18,3	20,5	16,0	19,6	17,4	18,5	26,0
Antequera	9	8,9	7,4	8,3	10,6	8,9	10,6	9,3	19,2	6,8	6,6	8,1	6,9	6,7	5,8	7,6	13,2	13,4	9,4	9,1	5,5	11,6
Castellbell	13	16,4	12,2	11,0	10,2	9,8	10,7	9,8	15,6	5,4	7,0	8,8	7,3	7,8	11,7	10,0	11,8	8,4	9,8	8,3	10,3	15,2
Villalba de Guardo	7	15,9	11,5	10,4	13,0	11,2	10,8	11,7	20,6	9,0	10,9	12,1	10,4	12,8	12,5	10,9	19,2	17,2	12,6	10,8	8,0	17,0
Villalba de Guardo	7	24,1	16,2	13,1	11,8	9,7	13,8	11,0	16,9	16,2	20,9	16,7	17,9	20,5	17,7	21,4	28,5	18,6	21,0	17,9	17,5	26,5
Vieuzos	16	22,3	15,3	12,2	9,9	9,9	10,4	8,9	24,5	7,6	6,9	9,0	9,5	10,1	11,5	13,9	18,0	9,8	10,8	12,0	12,4	9,9
Rennes	20	9,1	7,8	7,4	9,0	8,2	11,0	9,2	11,8	6,9	7,7	5,8	7,2	6,1	8,3	5,5	2,2	4,1	5,6	3,6	6,1	10,4
Rennes	20	15,9	10,9	9,3	9,1	10,7	11,8	11,9	11,9	8,5	8,5	6,6	7,0	8,1	6,0	11,5	15,0	12,2	9,9	9,1	7,8	9,6
Vayrières	19	18,3	15,3	12,7	17,2	11,3	11,9	10,3	19,2	6,2	9,4	8,7	7,8	11,2	12,4	10,1	12,2	10,5	11,4	10,8	13,3	17,6
Orléans	22	22,3	16,3	13,6	12,9	12,1	13,3	13,4	20,3	7,1	10,9	9,9	8,4	13,9	11,3	12,2	19,0	16,0	14,1	13,1	13,3	14,2
Senlis	21	12,7	10,2	9,0	13,2	11,7	9,9	11,0	20,0	10,2	7,5	5,9	7,8	7,1	7,3	10,4	10,1	12,1	6,9	7,4	9,2	8,6
Dole	23	17,2	8,7	5,9	8,4	7,5	7,6	9,1	21,9	6,3	5,9	8,5	6,2	7,4	7,7	7,6	8,0	6,3	4,3	5,8	6,9	9,3
Dole	23	14,9	10,1	7,7	11,0	12,3	12,7	12,6	15,8	12,0	11,4	10,9	10,0	10,3	14,0	10,6	13,8	10,6	12,7	10,4	12,3	17,4
St Beazely	24	15,8	11,5	8,8	10,0	9,6	6,6	7,1	18,8	7,6	9,5	11,9	9,4	13,8	10,5	10,3	17,0	14,4	12,1	15,5	13,5	16,2
Gignac	25	22,9	13,1	12,4	17,7	12,4	12,8	13,2	25,4	12,5	10,8	9,2	11,2	8,0	11,3	11,5	11,5	11,5	8,3	10,8	12,7	13,0
Gignac	25	16,8	12,2	11,3	5,9	9,2	10,5	11,7	15,6	7,0	6,7	5,7	6,4	6,0	6,3	11,2	15,3	12,7	10,2	10,9	5,8	2,5
Tarascon	26	14,0	10,9	8,9	10,9	11,1	12,7	12,3	15,8	9,9	9,5	8,7	7,1	8,7	8,7	6,0	5,6	4,0	7,4	5,3	5,1	12,2
Meyrargues	27	23,7	15,3	13,1	16,8	11,6	12,0	13,6	23,4	10,1	11,2	11,6	6,2	10,6	9,2	12,9	16,3	15,6	14,0	11,8	13,3	14,6
Meyrargues	27	11,4	6,8	6,2	7,5	6,8	10,5	8,7	5,5	6,8	8,9	9,1	9,5	10,6	11,5	8,4	8,4	7,8	12,9	8,0	11,2	14,7
Hageunau	45	20,0	13,3	8,3	13,4	8,8	10,6	8,6	24,3	10,4	12,4	11,9	12,5	14,2	13,9	10,3	9,4	6,6	8,1	6,3	9,0	22,1
Hageunau	45	18,7	12,2	9,1	11,6	7,5	12,9	7,8	15,5	10,9	13,2	14,0	11,5	13,8	14,4	12,5	14,1	11,5	15,8	8,9	9,8	22,4
Bulgaria	39	17,1	9,2	10,1	10,0	8,0	8,6	10,0	20,9	7,0	8,6	6,2	6,6	7,9	7,2	12,4	12,9	13,4	8,9	9,3	8,5	10,1
Macedonia	38	15,0	12,8	9,0	15,7	10,3	11,0	9,9	18,3	10,8	11,0	9,9	8,1	7,4	12,0	8,1	8,8	9,4	9,7	5,8	7,8	15,5
Macedonia	38	21,8	14,6	11,8	13,2	13,8	17,8	17,0	22,2	13,3	14,4	13,3	12,9	13,1	13,0	10,2	15,9	12,5	11,1	7,7	10,1	18,3

Chapitre III

**Effet des paramètres environnementaux sur
la dispersion de *M. galloprovincialis* :
l'importance des reliefs**

Chapitre III : Effet des paramètres environnementaux sur la dispersion de *M. galloprovincialis* : l'importance des reliefs

Après avoir défini la structure génétique du vecteur et cherché à identifier les facteurs qui la modulent, j'aborde dans ce chapitre la question de la dispersion du vecteur en milieu hétérogène. Pour cela j'ai utilisé une approche de génétique spatialisée (ou génétique du paysage), afin de tester l'existence de caractéristiques environnementales constituant des barrières ou des corridors à sa dispersion. Les contraintes relatives à l'espèce étudiée et à la zone d'étude choisie (péninsule ibérique) m'ont encouragé à adopter une attitude exploratoire quant au dispositif expérimental de cette étude. Ce chapitre contient donc autant de considérations fondamentales relatives à l'application des méthodes de génétique spatialisée que de résultats portant sur la dispersion de l'espèce étudiée. Ce chapitre s'appuie sur l'étude de l'article n°2.

Article n°2: Haran J, Rossi JP, Pajares JA, Bonifacio L, Naves P, Roques A, Roux G (*in prep*). Combining multi-scale and multi-site replication in spatial genetics : implications for highly dispersive insect species.

-Contexte et objectifs:

Les approches de génétique spatialisée sont largement utilisées pour identifier les structures et les paramètres environnementaux formant des barrières ou des corridors à la dispersion des espèces (Manel & Holderegger 2013). Les barrières à la dispersion de *M. galloprovincialis* sont importantes à caractériser car elles peuvent représenter des barrières à l'expansion naturelle du PWN. Les études de génétique spatialisée sont communément réalisées au sein d'une aire bien déterminée, c'est à dire à une échelle et en une localité unique. Un tel dispositif expérimental est adapté aux espèces dispersant relativement peu et fortement affectées par les paramètres environnementaux, ce qui est souvent le cas pour les espèces faisant l'objet de plan de conservation, et pour lesquelles les approches de génétique

spatialisée ont été développées (Cushman et al. 2006, Castillo et al. 2014). Le cas présenté dans cette thèse est différent car *M. galloprovincialis* est une espèce voilière à fort pouvoir dispersif (David et al. 2014), ce qui ne permet pas a priori de définir une échelle d'étude optimale pour étudier les flux de gènes chez cette espèce. *M. galloprovincialis* est également une espèce native en Europe et présente plusieurs lignées génétiques différenciées issues des patrons de recolonisation postglaciaire. Une telle différenciation génétique peut affecter une analyse de génétique spatialisée si elle implique un biais dans la dispersion (Pflüger & Balkenhol 2014) ou si elle masque une structure génétique plus fine issue de l'effet du paysage ou des paramètres environnementaux sur les flux de gènes. Une approche de génétique spatialisée pour cette espèce doit donc prendre en compte cette différenciation pour éviter les erreurs de type II. Enfin, *M. galloprovincialis* est une espèce se développant aux dépens des pins, et la distribution de ces arbres a beaucoup varié en Europe sous l'effet des activités humaines. Une telle variation est susceptible d'avoir affecté la mise en place de la structure génétique de cette espèce de manière non homogène dans l'espace et dans le temps. L'effet des paramètres environnementaux sur la dispersion doit donc être testé en plusieurs localités afin de prendre en compte cette hétérogénéité et de pouvoir généraliser l'inférence faite à partir des patrons génétiques. Plusieurs études ont montré l'importance de l'échelle (Cushman & Landguth 2010, Angelone et al. 2011, Galpern et al. 2012, Keller et al. 2013) et dans une moindre mesure des réplifications (Short Bull et al. 2011) pour faire une inférence correcte de l'effet des paramètres environnementaux sur la dispersion. En revanche, aucune étude n'a, à notre connaissance, combiné ces deux aspects sur des données empiriques.

L'objectif de cette étude est d'évaluer (i) l'effet des paramètres environnementaux sur la dispersion de *M. galloprovincialis*, une espèce voilière à fort pouvoir dispersif montrant une structure génétique affectée par les oscillations glaciaires. Plus généralement nous chercherons au travers de cet exemple à explorer (ii) l'effet du dispositif expérimental sur l'inférence faite en génétique spatialisée par une approche multi-sites et multi-échelles.

-Matériel et Méthodes :

Cette étude est basée sur l'analyse des génotypes de 992 individus génotypés à 10 loci microsatellites. Les individus ont été collectés en 132 sites (agrégats) distribués sur l'ensemble de la péninsule ibérique. Nous avons tout d'abord évalué la structure génétique des agrégats par ACP et par regroupement bayésien afin d'identifier les principales lignées évolutives présentes au sein de l'aire d'étude. Nous avons ensuite cherché à évaluer la variation spatiale dans la structure génétique de *M. galloprovincialis* afin d'identifier les échelles auxquelles une variation génétique existe. Pour cela nous avons reconstruit un variogramme à partir de l'ensemble des génotypes des agrégats. Cette approche géostatistique décrit la dissimilarité (ici génétique) entre individus en fonction de la distance séparant ces individus. Elle permet donc de mesurer l'échelle globale des flux de gènes et l'échelle à laquelle un effet des paramètres environnementaux peut être détecté.

Pour tester l'effet des paramètres environnementaux sur la dispersion de *M. galloprovincialis* en plusieurs sites et à plusieurs échelles, nous avons généré un ensemble d'aires d'études emboîtées formant des cercles de diamètres variant de 220 km à 1000 km (pas de 20 km) et centrées sur chaque point d'échantillonnage. Les analyses de génétique spatialisée ont ensuite été réalisées à partir des agrégats contenus dans chaque aire formée, et nous avons suivi l'évolution de la détection de l'effet des paramètres environnementaux en fonction de l'échelle et de la localisation de ces aires. Les corrélations entre distances génétiques et distances géographiques (incluant la résistance des paramètres environnementaux) ont été réalisées par tests de Mantel partiels, qui permet de s'affranchir des distances géographiques euclidiennes en les partitionnant. Malgré les biais associés à ce test (Guillot et Rousset 2013), cette approche reste largement utilisée en génétique spatialisée et s'est montrée capable d'identifier les facteurs affectant les flux de gènes tout en rejetant les hypothèses alternatives non valides (Cushman et al. 2006, Cushman & Landguth 2010, Galpern et al. 2012, Castillo et al. 2014). Les distances génétiques ont été calculées sur la base de distances inter-individuelles en utilisant le pourcentage de dissimilarité de Bray-Curtis. La résistance des paramètres environnementaux, quant à elle, a été estimée en suivant le modèle d'isolement par résistance (IBR, McRae 2006). Nous avons testé l'effet de quatre paramètres sur la dispersion de *M.*

galloprovincialis (ou hypothèses d'isolement par résistance): l'élévation (E), les températures minimales hivernales (T) et la densité en pins en tant que barrière (Pr) ou corridor (Pc) à la dispersion. Afin de tester l'importance relative de ces paramètres et leur cumul, nous avons réalisé une analyse de commonalité (Prunier et al. 2014). Les distances de résistance issues de chaque paramètre et l'ensemble de leurs combinaisons ont été régressées avec les distances génétiques (modèles de résistance des paramètres environnementaux). Nous avons ensuite suivi l'évolution des coefficients de commonalité, qui correspondent à part de la variance explicable par chacun des modèles construits.

Les analyses de corrélation entre les génotypes et les paramètres environnementaux ont été réalisées sur l'ensemble du jeu de données et au sein du principal groupe (ou « cluster ») identifié par analyse par regroupement, afin d'évaluer l'effet des lignées évolutives sur l'inférence faite au sein des aires chevauchant plusieurs lignées. Nous avons également mesuré l'amplitude de variation des paramètres testés au sein de chaque aire formée afin d'évaluer l'effet du contraste sur la détection des paramètres environnementaux (Jaquière et al. 2011, Cushman et al. 2013).

-Résultats:

Les agrégats de *M. galloprovincialis* se structurent en deux lignées principales en péninsule ibérique, l'une distribuée sur une frange atlantique (Portugal et Galice) et l'autre dans le reste de la péninsule. Le variogramme montre plusieurs paliers, suggérant que cette espèce se structure sur plusieurs niveaux hiérarchiques et emboîtés de variation génétique. Nous avons observé une augmentation progressive de la dissimilarité (semi-variance) avec la distance jusqu'à un premier palier pour une distance d'environ 190 km. Ce résultat indique qu'il existe une importante auto corrélation spatiale des génotypes sous ces distances. Le palier observé entre 200 et 400 km tend à montrer qu'il existe un premier niveau de variation génétique à ces échelles. L'augmentation de la semi-variance au-delà de ce premier niveau met en évidence l'existence d'une variation génétique présente à large échelle (environ 1000 km) superposée à ce premier niveau – plus local – de variation génétique.

Nous avons observé une corrélation significative entre l'ensemble des quatre hypothèses de résistance des paramètres environnementaux et la distribution des génotypes de *M. galloprovincialis*

indiquant un effet barrière aux flux de gènes des paramètres testés. Cependant la détection de cet effet a montré de fortes variations selon l'échelle et la zone d'étude considérée. Globalement, nous avons observé une augmentation de la détection de l'effet des paramètres environnementaux sur la dispersion en augmentant l'échelle d'étude, cet effet étant plus rarement détecté aux échelles les plus faibles. A échelle fixe, nous avons observé une distribution spatiale non homogène de cet effet. Les paramètres *T*, *E* et *Pr* ont principalement été détectés dans la moitié supérieure de l'aire d'étude tandis que le paramètre *Pc* a été détecté principalement – à fréquence moindre – dans le sud de la péninsule. Cette distribution hétérogène est particulièrement visible pour les échelles d'étude petites et moyennes. En comparant la variation des paramètres environnementaux au sein des aires montrant un effet significatif et non significatif, nous avons observé que cette hétérogénéité ne correspondait pas aux variations de contraste des paramètres dans ces aires (pour les paramètres *E*, *Pr* et *Pc*). Ce constat s'applique uniquement aux échelles les plus faibles, car les aires montrant un effet significatif à large échelle sont généralement associées à des contrastes plus forts des paramètres testés.

Nos résultats montrent également un effet des lignées évolutives sur la détection de l'effet des paramètres environnementaux sur la dispersion. Les aires d'étude chevauchant les deux lignées mises en évidence (lignées ibériques est et ouest) ont montré un effet non significatif tandis qu'un effet a été détecté dans les mêmes conditions lorsque les analyses ont été réalisées au sein d'une seule lignée.

L'analyse de commonalité montre que l'ensemble des modèles de résistance construits expliquent jusqu'à 24 % de la variance totale de la structure génétique de *M galloprovincialis*. Les modèles incluant les paramètres *T* et *Pr* seuls (effets uniques) présentent la contribution la plus élevée à la variance explicable (plus de 20 %) du modèle de régression total. Au sein de l'ensemble des effets uniques et cumulés, le cumul des paramètres *E* et *T* est le modèle montrant la meilleure contribution à cette variance explicable totale (plus de 50 %).

Combining multi-scale and multi-site replication in spatial genetics: implications for highly dispersive insect species.

Julien HARAN^{1,2}, Jean-Pierre ROSSI³, Juan PAJARES⁴, Luis Bonifacio⁵, Pedro Naves⁵, Alain ROQUES¹ & Géraldine ROUX^{1,2}

1 INRA, UR633 Zoologie Forestière, F-45075 Orléans, France.

2 Université d'Orléans, France.

3 INRA, UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Montpellier, France.

4 Univ Valladolid INIA Palencia, Sustainable Forest Management Res Inst, Palencia, Spain.

5 Instituto Nacional de Investigacao Agraria e Veterinaria – INIAV Oeiras, Portugal

Abstract

Inferring the effect of environmental features on gene flow of species is fundamental to understand population structure, dynamics and evolution. However, such inference may be challenging due to the complexity of environmental factors, which acted on various spatial and temporal scales and together established current genetic patterns. For instance, recolonisation after glacial oscillations acted at larger time scales than alterations in gene flow caused by human land use at the landscape level. In addition, occasional human transport can confuse genetic signals of gene flow from which effect of environmental features on dispersal of species is inferred. In this study, we explored the resistance of environmental features (temperatures, elevation and host tree density) to dispersal of a flying long-horned beetle (*Monochamus galloprovincialis*) in a region that underwent deep changes in landscape structure, namely, the Iberian Peninsula. We combined a multiple scale and multiple site landscape genetic approach to assess variation in the detectability of environmental effects on gene flow. Analyses were based on a sample of 992 individuals spread over 132 locations and genotyped at ten microsatellite loci. We generated more than 3×10^4 study areas ranging in scale from 220 to 1000 km and tested for general effects of environmental features on gene flow of *M. galloprovincialis* in all of them. Generally, elevated areas, low winter temperatures and high pine

densities were detected as factors hindering gene flow. However, we also show that there were no general and recurrent effects of environmental features detected among scales and locations, independent of variation in environmental features. We also show that conflicting signals between alternative resistance models may be detected among replication. Our results suggest that distinct evolutionary lineages may bias detection of the effect of environmental features on gene flow. This empirical study highlights the importance to consider both scale and geographic replications when searching for general effects of environmental factors on gene flow in spatial genetics studies.

Keywords: gene flow, landscape genetics, insect dispersal, *Monochamus galloprovincialis*, Iberian Peninsula.

Introduction

Landscape genetics examines the relation between landscape and environmental features and genetic structure (Manel et al. 2003, Manel & Holderegger 2013). It allows to infer which environmental features facilitate or hinder gene flow (Zeller et al. 2012), which is a key factor to understand the persistence and evolution of species and has significant consequences for conservation planning (Castillo et al. 2014, Van Strien et al. 2014). As an emerging and fast moving field, the landscape genetic toolbox is far from being established, and an important effort of method optimization is still required to make correct and optimal inferences about the effects of landscape and environmental features on gene flow (Anderson et al. 2010, Cushman et al. 2013, Manel & Holderegger 2013). Landscape genetic analyses are usually conducted in a single scale and in a single location (Zeller et al. 2012) and therefore provide results that are strictly speaking only applicable to the particular area studied. Indeed, genetic structure is determined by multiple micro and macro-evolutionary processes acting at diverse spatial and temporal scales rarely homogeneously distributed across a study species' distribution range. For example, in addition to contemporary or historical environmental effects on dispersal (Zellmer & Knowles 2009), the genetic structure of organisms is often influenced by historic differentiation due to quaternary climate oscillations (Hewitt 2000), or by biased dispersal due to local adaptation to specific environmental conditions (Sexton et al. 2014,

Pflüger & Balkenhol 2014). The diverse factors acting at different temporal and spatial scales may generate genetic patterns that may be inconsistent across locations or regions, which results in conflicting signals of environmental factors acting on gene flow. This may drastically impede the ability to infer the general drivers of gene flow. To overcome this problem, several authors have pointed to the importance to match study design to the process investigated (Anderson et al. 2010, Cushman & Landguth 2010, Galpern et al. 2012, Keller et al. 2013), or stressed considering landscape-level replications in landscape genetic analysis (Short Bull et al. 2011).

The scale of study is fundamental in landscape genetics, because species respond to environmental features at a continuous range of scales. This point has been particularly highlighted for organisms exhibiting wide home-ranges like large mammals (Galpern et al. 2012, Zeller et al. 2014). The scale of the studied area should exceed the scale of recurrent gene flow (*i.e.* expected dispersal distances of study species). The importance of spatial scale has been highlighted in overview articles (Anderson et al. 2010, Manel & Holderegger 2013), in empirical studies (Angelone et al. 2011, Galpern et al. 2012, Dudaniec et al. 2013, Keller et al. 2013) and in simulation exercises (Cushman & Landguth 2010). Despite an increasing number of studies explicitly accounting for scale effects, landscape genetics studies still rarely consider scale effects (Zeller et al. 2012) and how it affects inference on the detection of general effects of environmental features on dispersal and gene flow. As a result, extrapolation of landscape genetic results obtained from a single study area is problematic.

Replication in landscape genetics consists of landscape-level replication of sampling areas. Few studies have included replication in landscape genetics studies (Driezen et al. 2007, Kindall & Van Manen 2007, Zalewski et al. 2009), and the number of replication is often low due to the sampling effort required. Short Bull et al. (2011) evaluated the influence of landscape on dispersal of the American black bear (*Ursus americanus*) among 12 study areas. They found that the landscape features influencing gene flow were consistent among replications. They could thus conclude about general landscape effects on dispersal of black bears. However, these authors also showed that the features mostly influencing gene flow in particular study areas varied greatly, because of variations in landscape structure.

The scale of processes affecting genetic variation and the location where these processes act are usually unknown when designing sampling a strategy and when no pilot studies are available. In addition, landscape and environmental features and their variation are rarely homogenously distributed across species' range. This again stresses the importance to consider scale and replication in landscape genetics studies. In this study, we used an approach based on nested sampling units distributed across a large area and therefore including a high number of locations. As a study system, we investigated how environmental features influences gene flow of *Monochamus galloprovincialis* (Coleoptera, Cerambycidae), the European vector of the pinewood nematode (*Bursaphelenchus xylophilus*, PWN), in Iberian Peninsula. *M. galloprovincialis* is a xylophageous beetle that performs its life cycle in the wood of declining pine trees (*Pinus pinaster*, *P. sylvestris*, *P. nigra*, *P. halepensis*, Naves et al. 2006, Hellrigl 1971). The study species is rather long-lived and is thought to have good dispersal abilities (David et al. 2013, Mas et al. 2013). However, the role of environmental features as barriers or corridors to dispersal of this species is still not understood (but see Haran et al. 2015), and the scale of effective gene flow is unknown.

The Iberian Peninsula is a complex area regarding the genetic structure of biota. Repeated glacial oscillations have involved dramatic changes in ecosystems and in landscapes, which have shaped the genetic structure of most organisms through the existence of multiple glacial refugia (Gomez & Lunt 2006). The genetic legacy of these historical processes is the existence of multiple evolutionary lineages. These lineages are expected to cause bias if found in the same study area, because their genetic divergence can misbalance the detection of a weaker genetic structure derived from recent gene flow. Conversely, the landscapes of the Iberian Peninsula (forest cover in particular) experienced a series of recent changes due to human activities (Ruiz-Benito et al. 2012, Abel - Schaad et al. 2014, Lopez-merino et al. 2014, Serra et al. 2014). Such changes may result in the fragmentation or connectivity of the populations of organisms. Recent environmental changes are critical because the time lag between the causal effects and the genetic reaction may be too short to be detected.

The primary goal of the present study was to investigate how scale and location of affects landscape genetics results. In particular, we explored the detection of environmental features fostering

or hindering gene flow in *M. galloprovincialis* at scales ranging from 220 to 1000 km in 132 locations spread across the Iberian Peninsula. For this aim, we used individual based landscape genetic analysis including partial Mantel tests and commonality analysis.

Material and Methods:

Sampling and genotyping

The study area covered the entire Iberian Peninsula (582 000 km²) with altitudes ranging from sea level up to 2444 m. *M. galloprovincialis* specimens were trapped between 2011 and 2013 at 137 sites spread over the Iberian Peninsula. We used multifunnel traps and a volatile attractant (Galloprotect, SEDQ, Spain) placed in pine stands during summer to catch flying adults. A total of 1050 individuals were collected, and stored in 96.66 % ethanol at 4°C. Details of sampling localities, host trees and year of collecting are given in table S1 (supporting information). Individuals collected at the same locality were considered as one deme. The distribution of sites covered most of the pines forests found in Iberian Peninsula (Fig. S1; supporting information).

DNA was isolated from two legs per individuals using a Nucleospin Kit (Macherey-Nagel, Düren, Germany). Specimens were genotyped at 12 microsatellite loci (Mon01, Mon08, Mon17, Mon23, Mon27, Mon30, Mon31, Mon36, Mon41, Mon42 & Mon44) following Haran and Roux-Morabito (2014). Details of primer sequences and the protocol for genotyping are given in table S2 (supporting information). Negative and ambiguous amplification of particular loci were repeated once and considered as null when still unsatisfactory. Individuals exceeding two missing loci were removed for analysis. Deviation from Hardy Weinberg equilibrium (HWE) and linkage equilibrium at each locus was estimated on a subsample of three demes of larger size ($n > 19$, figure S1). We used GENEPOP 4.2 (Raymon & Rousset 1995) to test for these disequilibria, and significance was assessed using a False Discovery Rate (FDR) correction (Q-value; Storey et al. 2002). The frequency of null

alleles at each locus and for each population was tested using FREENA (Chapuis and Estoup 2007). Loci exceeding a rate of 7 % of null alleles across populations were discarded from further analysis.

Genetic structure

Clustering of genotypes. We used the bayesian approach implemented in STRUCTURE 2.3.4 (Pritchard et al. 2000) to identify the main genetic lineages among Iberian demes. STRUCTURE assigns individuals to a predefined number of clusters based on allelic composition and linkage disequilibrium. We used the Delta K method (Evanno et al. 2005) to determine the number of clusters (K) that best fit the data. Genotypes were analyzed using default parameters (admixture model, correlated alleles frequency). We made ten repeats of a 200.000 burn-in period followed by 500.000 replicates of Markov Chain Monte Carlo (MCMC), for K values ranging from 1 to 10. Results were uploaded in STRUCTURE HARVESTER (Earl et al. 2012) to determine the optimal K. We also analyzed the structure of genotypes of individuals through a principal component analysis (PCA) using the R package adegenet (Jombart 2008) in R 3.0.2 (R development Core Team 2012). To account for potential confounding effects of differentiated evolutionary lineages (STRUCTURE clusters) on the inference of gene flow, landscape genetic analyses were performed twice, once within the main cluster identified by STRUCTURE and PCA, and once with the whole dataset including all clusters.

Geostatistical treatment of the genetic structure. The scores of the sampling locations upon axis 1 of the PCA are linear descriptors of the allele frequencies and, as such, can be used as a univariate statistical measure of genetic composition (Rossi et al 2014). We analyzed the score of sample points upon axis 1 by means of a variogram (Goovaerts, 1997) to explore the scale at which variation in genetic structure occurs. Let $z(u_\alpha)$, with $\alpha=1, 2, \dots, n$, be a set of n values of sample scores upon a PCA axis where u_α is the vector of spatial coordinates of the α th observation. In geostatistics, spatial dependence is generally described in terms of dissimilarity between observations expressed as a function of the separating distance. The average dissimilarity between data separated by a vector h is

measured by the experimental semi-variance $\hat{\gamma}(h)$, which is computed as half the average squared difference between the data associated with every data pair:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{x=1}^{N(h)} [z(u_x) - z(u_x + h)]^2 \quad (1)$$

where $N(h)$ is the number of data pairs for a given lag vector h , $z(u_x)$ and $z(u_x + h)$ the score values of all sample locations separated by a vector h . The more alike the observations at points separated by h are, the smaller $\hat{\gamma}(h)$ and vice versa. The plot of $\hat{\gamma}(h)$ against h is called a semi-variogram of variogram for short and represents the average rate of change of z with distance. Its shape describes the pattern of spatial variation in terms of general form, scales and magnitude.

Variograms are good tools to depict spatial structures and to analyze nested patterns (Burrough 1983). When different sources of variation shape the investigated pattern at different spatial scales, the resulting variogram exhibits different plateaus in association with different scales (Robertson & Gross 1994, Rossi 2003). A theoretical model is often fitted to the observed variogram and the model parameters are used for interpolation purposes (Goovaerts 1997). In our study we fitted an exponential model, because it is supposed to well summarize spatial structure (Goovaerts 1997). Three main parameters are estimated during model fitting; (i) the range, (ii) the sill, which denotes the plateau of the variogram, and (iii) the nugget, which is the intercept of the variogram and quantifies the amount of variance occurring at scales below the minimum inter-sample distance plus measurement errors. Variograms were computed using the R package GEOR (Ribeiro & Diggle 2001).

Landscape genetics analysis

Genetic distances. We computed genetic distances between demes using an individual-based metric (Shirk et al. 2010, Prunier et al. 2013). We first constructed a matrix where each individual is a row and alleles are columns and where genotypes were coded for each allele as 0 when absent, 1 when single at a locus (heterozygotes) or 2 for homozygotes (Shirk et al. 2010). Thus individuals are

represented as a linear vector of size n , where n is the total number of alleles encountered in all individuals genotyped. We then generated a semi matrix of distance between all pairs of individuals. We computed the Bray-Curtis percentage of dissimilarity (Legendre & Legendre 1998) to estimate differentiation between all pair of individuals. Calculations were performed in R 3.0.2 (R Development Core Team 2013) using the VEGAN package.

Spatial distances. We selected the environmental features considered as the most likely to influence dispersal of *M. galloprovincialis* given the existing knowledge on species requirements. Apart from Euclidian distances (null model), we considered three environmental features to be potential drivers of dispersal (pine density, winter temperatures and elevation).

We modeled environmental resistance as a function of pine density as this parameter determines the volume of resource available for the *M. galloprovincialis* and is thought to affect its foraging dispersal. As the dispersal behavior of this beetle in reaction to pine density is not known, we modeled this parameter according to two alternative scenarii. (1) High pine densities are positively correlated with beetle dispersal. In this scenario, a dense pine cover represents a corridor for dispersal due to the high amount of resources available. Conversely, a low pine density would represent a barrier. (2) High pine densities are negatively correlated with beetle dispersal. For this second scenario, it was assumed that a dense pine cover provides enough resources for local populations, which therefore would not need to disperse. This scenario thus assumes more dispersal in low pine cover areas. To model resistance based on pine density, we considered the sum of densities of all pine species encountered in a grid cell. Indeed, in the Iberian Peninsula, *M. galloprovincialis* is performing its life cycle in stressed or fresh dead wood of the most widespread pine species: *Pinus pinaster*, *P. nigra*, *P. sylvestris*, *P. halepensis* and *P. radiata* (Hellrigl 1971, Naves et al. 2006) and shows no specialization for any of these host species (Haran et al. 2015).

Resistance was modeled as a function of winter temperatures, as winter temperatures is likely to determine survival or developement rate of larval instars in *M. galloprovincialis*. This species performs its larval phase during winter, and instars may stop their development and eventually die after exposition to extended periods of cold temperatures (Naves & Sousa 2009). As no precise

threshold is known, we consider that resistance increase when winter temperatures decrease. We used the mean minimum temperatures of winter months (December, January and February) to represent variations in temperatures. Elevation is often a proxy for temperature. We hypothesized that resistance to dispersal increases when elevation increases. We kept temperatures and elevation as distinct environmental features for the analysis, because temperature and altitude may not covary similarly at large scales (North to South of Spain; for collinearity see below). A summary of resistance scenarios of environmental features is given in Table 1.

Table 1: Summary of environmental features tested in isolation by resistance (IBR) models.

Environmental features	Code	Associated IBR hypotheses
Elevation	<i>E</i>	High elevations = resistance to dispersal
Pine density	<i>Pc</i>	High pine density = corridors to dispersal
	<i>Pr</i>	High pine density = resistance to dispersal
Mean min. Winter temperatures	<i>T</i>	Low winter temperatures = resistance to dispersal

Resistance distances were computed using the package GDISTANCE (van Etten 2012) in R 3.0.2 (R Development Core Team 2013). Raster layers of environmental features were imported at a resolution of 10 x 10 km. Such resolution was chosen because the mean flight distance of *M. galloprovincialis* is about 16 km in flight mills (David et al. 2014). Temperature data (1950-2000) were downloaded from Hijmans et al. (2005; <http://www.worldclim.org>) and pine density from Tröltzsch et al. (2009; <http://www.efi.int/>). For control purpose, resistance distances were also measured on layers with a resolution of 1 x 1 km. As temperature, elevation and pine density are continuous parameters, we did not assign particular resistances to particular values, but directly used the values. Pairwise resistance distances were estimated based on random walk probabilities (Chandra et al. 1997, McRae 2006) and computed using the command `commuteDistance` (package GDISTANCE). Resistance distances were chosen instead of least cost distances (LCD) because they are thought to be more reliable biologically and produce fewer artifacts on long distances (McRae 2006). We constructed a semi matrix of resistance distance between each pair of individuals. Values were normalized for further analysis. Collinearity was estimated using the variance inflation factor

(VIF) based on the formula $VIF = 1/(1-R^2)$, where R^2 is the r-squared value of regression between variables. VIF values > 10 are usually considered as evidence for collinearity between environmental features (O'Brien 2007). We did not detect collinearity between environmental features over the whole area of study ($VIF < 1$ for all pairwise comparisons).

Correlation analysis. We tested correlation between the response (genetic distances matrix, G) and resistance distances (resistance matrices; IBR) and geographic distances (Euclidian distance; IBD) using partial Mantel tests (Cushman & Landguth 2010). Partial Mantel tests measure association between two distances matrices while partialling out a third distance matrix. This approach is widely used in the field of landscape genetics (Cushman et al. 2006, Cushman & Landguth 2010, Galpern et al. 2012, Castillo et al. 2014) and have shown to efficiently infer the drivers of gene flow (Cushman & Landguth 2010b). Partial Mantel tests have received criticism regarding their statistical performance (Guillot & Rousset 2013), however, recent works have highlighted the validity of this test on empirical data (Diniz-Filho et al. 2013), and it's complementary to other approaches (Kierepka et al. 2015). In this study, we used a causal modeling framework (Cushman et al. 2006). We first used simple Mantel tests to correlate IBD with G. We then tested the effect of IBR in partial Mantel tests. Support for IBR was considered when: (1) IBR should be significantly correlated to G after partialling out IBD ($p < 0.05$) and IBD should be non-significant with IBR partialled out ($p \geq 0.05$). Mantel and partial Mantel tests were performed using the *vegan* package in R 3.0.2 (R Development Core Team 2013) with 10^3 permutations. To estimate the relative importance of the environmental features in shaping genetic structure, we conducted a commonality analysis (Prunier et al. 2014). The response G was regressed onto each resistance matrices separate and each combination using the *YHAT* package in R 3.0.2 (R Development Core Team 2013). Commonality analysis was also used as a complementary approach with Mantel tests, to detect supported isolation by resistance hypothesis.

Multiple scales and multiple locations analysis

We considered various spatial scales and various locations in the above landscape genetic analysis by generating nested sampling areas spread over the full extent of the Iberian Peninsula. Sampling areas were constructed as circles of diameters ranging from 220 to 1000 km (steps of 20 km) and centered at each sampling location. Mantel tests were performed between all individuals found within each area defined. Areas of diameter below 220 km were not included in the analysis, as it was often too small to gather neighboring demes for Mantel tests. We then tracked the evolution of the number of areas with supported IBR hypothesis and the mean significant Mantel r with increasing the scale. The geographic distribution of areas with a supported IBR hypothesis was obtained by summing the number of times that each individual was included in a sampling area with IBR hypothesis support among all scales. Obtained numbers (frequencies) were corrected accounting for intrinsic variation due to overlapping of sampling areas. Frequencies at each point were interpolated using the Inverse Distance Weighted method (IDW) in ARCGIS 9.3 (ESRI, Redlands, CA, USA) to visualize variation in spatial distribution of areas which supported each IBR hypothesis. Landscape genetics analyses have been shown to perform better in a contrasted landscape (Jaquiéry et al. 2011, Cushman et al. 2013). We extracted resistance values of raster cells within each sampling area and computed the standard deviation (SD) of these values to determine whether support of IBR hypotheses was due to variation in the environmental features tested. We then calculated mean standard deviation of areas with supported and non-supported IBR hypotheses among scales of study. Commonality analyses (see above) were performed within each sampling area generated. We tracked the development of commonality coefficients (percentage of variance explained by unique and cumulated IBR hypothesis) among scales and locations. The sampling area maximizing commonality coefficients was chosen for representation of the relative importance of environmental features in shaping genetic structure.

Results

Sampling and genotyping:

Despite intensive trapping, *M. galloprovincialis* was not recorded in five localities in the Central lowlands of Castilla y Leon, central Galicia and Asturias districts. We finally obtained a sampling of 1050 individuals at 132 sites. Seventy-seven sites had a size below 10 and 55 above or equal to this value, average sampling size per location was 7.68 individuals. Among the 3 populations of larger size tested ($n > 19$), two loci exhibited substantial null allele frequencies ($> 7\%$) and were not considered for further analysis (Mon 01 and Mon 27). Significant heterozygote deficit was detected at four loci (Mon 30, 35, 42, 44). Corresponding null alleles frequencies were low ($< 7\%$), so these loci were retained. No significant linkage disequilibrium was detected among pairs of loci. After removal of incomplete genotypes ($n = 58$) and biased loci, we obtained a total number of 992 individuals genotyped at ten loci. The average number of alleles per locus was 10.2 (range: 6-24).

Genetic structure

Individuals clustered into two lineages under STRUCTURE analysis (Delta $K_2 = 1061.47$; delta $K_3 = 101.68$). Clusters showed a clear geographic structure, exhibiting a split between Portugal and western Galicia (West Iberian cluster) versus the rest of Iberian Peninsula (Fig. 1A). Higher values of K , although less supported by Delta K method, gave additional geographic clusters. Demes from the North-East half of Spain exhibited a higher level of structure than the remaining demes (Fig. 1B). PCA gave similar results on the first axis, splitting demes into two distinct clusters (Fig. 1C). Additional splits were not observed with PCA. Estimates of population differentiation (F_{st}) between the three populations of large size ($n > 19$) was moderate (Castro Daire /Catsellbell: 0.12; Castro Daire/Vale Feitoso: 0.15; Catsellbell/ Vale Feitoso: 0.07).

Variogram. Data points were grouped into 26 distance classes ranging from 0 to 1252 km, with a spatial lag of 50 km. The variogram reveals that the first axis of the PCA corresponds to a

highly spatially structured pattern (Fig. 2). Moreover, the structure at hands corresponded to nested patterns possibly originating from superimposed sources of genetic variation. The semi-variance first progressively increased with increasing lag distance up to a distance of about 190 km and then reached a plateau. For distance of about 400 km, the semi-variance increased again and leveled off for distances larger than 1000 km. The shape of this variogram therefore revealed the presence of a long-range spatial variation superimposed over a more local, *i.e.* short-scale genetic structure occurring at scales of 200 to 400 km. For scales below 200 km, the variogram show that genotypes were strongly spatially autocorrelated (*i.e.* non-independent).

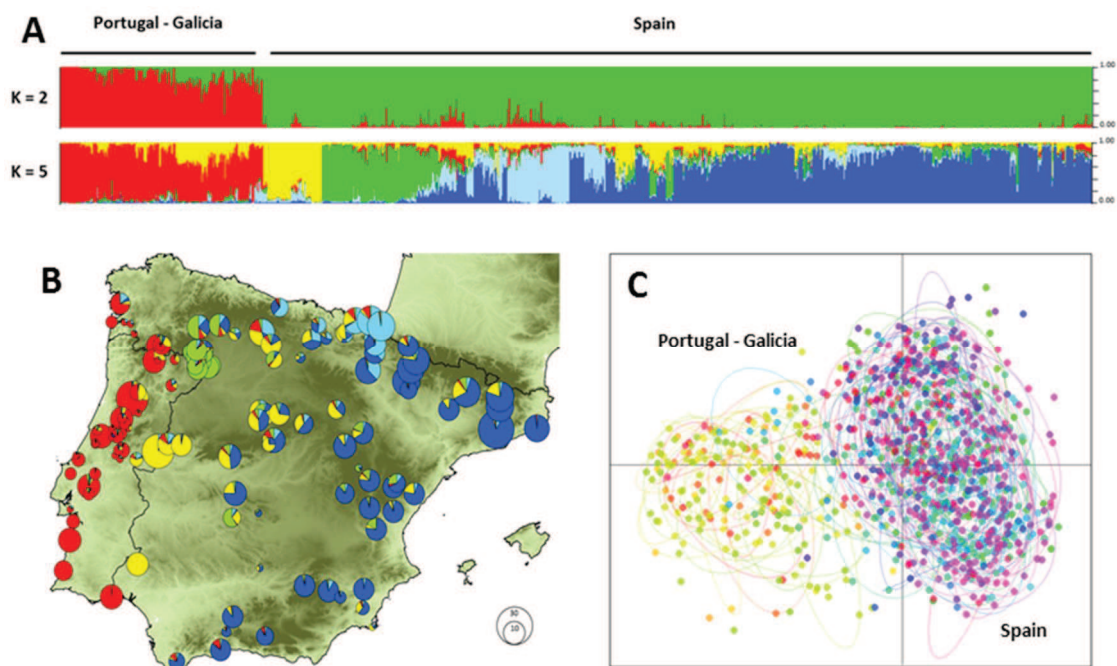


Figure 1: Genetic clustering of 992 individuals of *Monochamus galloprovincialis* sampled at 132 locations. **A:** Assignment of individuals to clusters based on a STRUCTURE analysis for K=2 and K=5. **B:** Assignment of demes to clusters for k=5, displayed in geographic context (Iberian Peninsula, size of pies refer to the size of demes). **C:** First PCA axis of individuals.

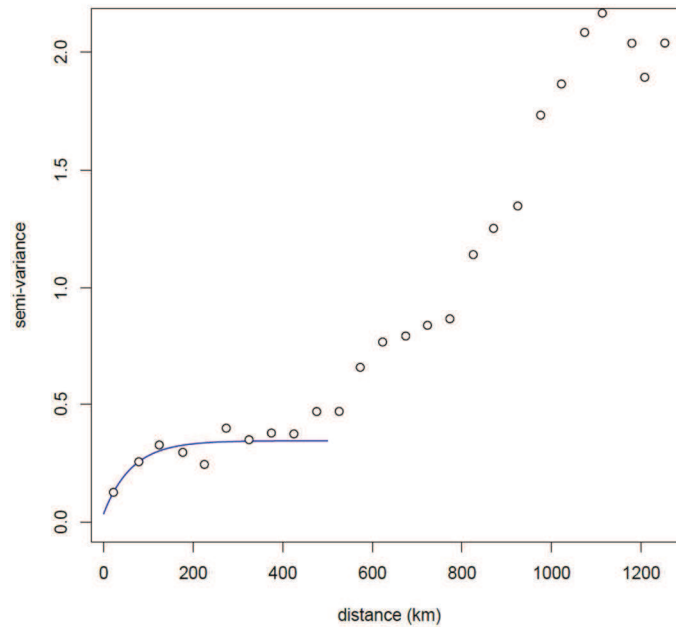


Figure 2: Empirical semi-variogram of genotypes of *Monochamus galloprovincialis*. The variogram was fitted with an exponential model to highlight the first plateau. Data points are shown with a spatial lag distance of 50 km.

Landscape genetics analysis

Analyses were conducted both on the whole dataset (992 individuals, 132 localities) and within the Spanish cluster (790 individuals, 87 localities), for a total of 116 and 102 alleles analyzed respectively. Grain sizes of 1 x 1 km and 10 x 10 km resulted in similar results. Null distances were not encountered at grain 10 x 10 km, as none of the sampling sites fell with neighbor sites in the same grain. Below we thus focus on results obtained for grain 10 x 10 km.

Over the whole area of study, we generated a total of 30 576 sampling areas. The mean number of individuals within sampling areas varied from 89.18 (SD: 42.42) at the smallest scale (220 km) to 644.58 (SD: 158.07) at the largest scale (1000 km; Fig. S2, supporting information). Significant effects of environmental features were detected for all IBR hypotheses tested with Mantel test, but the frequency of areas exhibiting an IBR effect varied among scales and locations. The number of areas showing a significant effects of environmental features generally increased with increasing scale (Fig. 3A), but each of the four IBR hypotheses showed a different pattern. Significant effects of

environmental features for *E*, *Pr* and *T* hypotheses were detected in about 15-25% of the areas at smallest scale (220 - 300 km). The frequency of *E* and *Pr* gradually increased to reach 90% and 60 % for areas at 1000 km. The frequency of areas with a supported *T* hypothesis increased among scale to reach a peak around 600 km (≈ 80 % of areas) and subsequently decreased again. Significant *Pc* hypotheses were encountered at a lower frequency. The number of positive areas ranged from 0 to 1.11 %, for an average number of 4.07 areas for each scale considered. No specific trend was observed when scale increased for the *Pc* hypothesis. Significant isolation by distance (IBD) was observed for ≈ 60 % of areas at smallest scale. A first plateau of about 85 % of areas was reached for scales ranging between 400 and 700 km, and a second plateau of almost 100 % of areas was reached for scales above 700 km. Best values of mean Mantel r were observed at small scales and generally decreased when scale increased (Fig. 3B). IBR hypothesis *T* showed the best Mantel r among all IBR hypotheses for scales above 360 Km.

Interpolation of supported IBR hypotheses and IBD was based on areas of scales ranging from 220 to 600 Km, because at these scales most of variation in the detection of effects of environmental features was found (Fig. 3A). For most IBR hypotheses (*E*, *Pr* and *T*) and IBD, effects were mainly detected in the northern part of the area of study, corresponding to Cantabrian chain and the western half of the Pyrenees (Fig. 4). By contrast, these IBR hypotheses were the least frequently detected in a region comprising the eastern side of the Iberic and Betic mountain systems. For the IBR hypothesis *Pc*, significant effects were detected mainly in Andalucía, along the Betic system. Conversely, low or no effects for this hypothesis were detected in the Northern half of the Iberian Peninsula.

The effect of coexistence of different STRUCTURE clusters on the inference of effects of environmental features was mainly observed at the smallest scales (220 - 300 km). The inclusion of the west Iberian cluster in the analysis has led to a loss of supported effect for areas where the two clusters overlapped at several scales and for several IBR hypotheses. This effect was the most striking for areas located near the north-eastern part of Portugal (Fig. 5). Interpolation did not highlight this effect clearly, due to the “smoothing” effect of wider areas for which differentiation between clusters impacted less on the detection of IBR hypotheses. As a result, the distribution of supported hypotheses

was generally similar between that performed on the whole dataset and on the Spanish cluster only (Fig. 4).

For hypotheses *E*, *Pr* and *Pc*, the variation of environmental features was on average lower in areas exhibiting significant effects for scales up to scales of 400 - 600 km (Fig. 6). Above this scale, the mean standard deviation (SD) of significant areas was either equal, either higher to the mean SD of non-supported areas. For the *T* hypothesis, mean SD of significant areas was above the mean for non-supported areas for most of the scales.

Regression models gave a maximum explained variance of 24 % over all sampling areas through commonality analysis (Table 2). Best values were obtained in various locations for medium size scales (520-620 km). Relative importance of unique and common effects of IBR hypotheses were constant between the three areas exhibiting maximum explained variance. The features *T* and *Pr* uniquely contributed to more than 20 % of the total variance explained (20.77 to 32.65 % and 21.82 to 35.24 % respectively). The best contribution to the total variance explained was observed for the common effects of *E* and *T* (54.31 to 56.43 %).

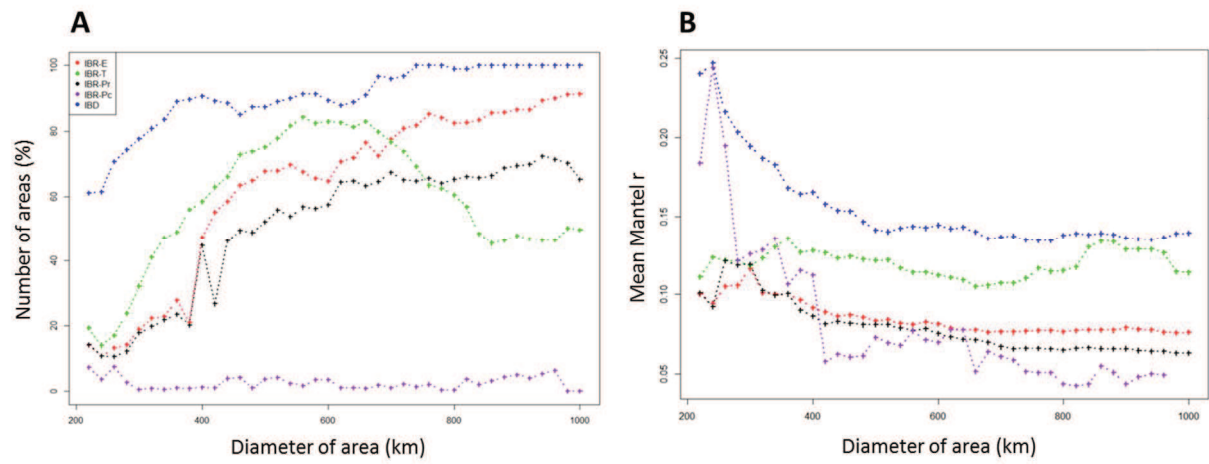


Figure 3: Development of the number of areas with supported IBR hypotheses for Mantel tests (A) and of mean partial Mantel r (B) of areas with support of IBR hypotheses with increasing scale (whole dataset). IBR-*E*: Elevation, IBR-*T*: Mean minimum Winter temperatures, IBR-*Pr* and IBR-*Pc*: pine densities as a resistant feature and as a corridor respectively, IBD: Isolation by distance.

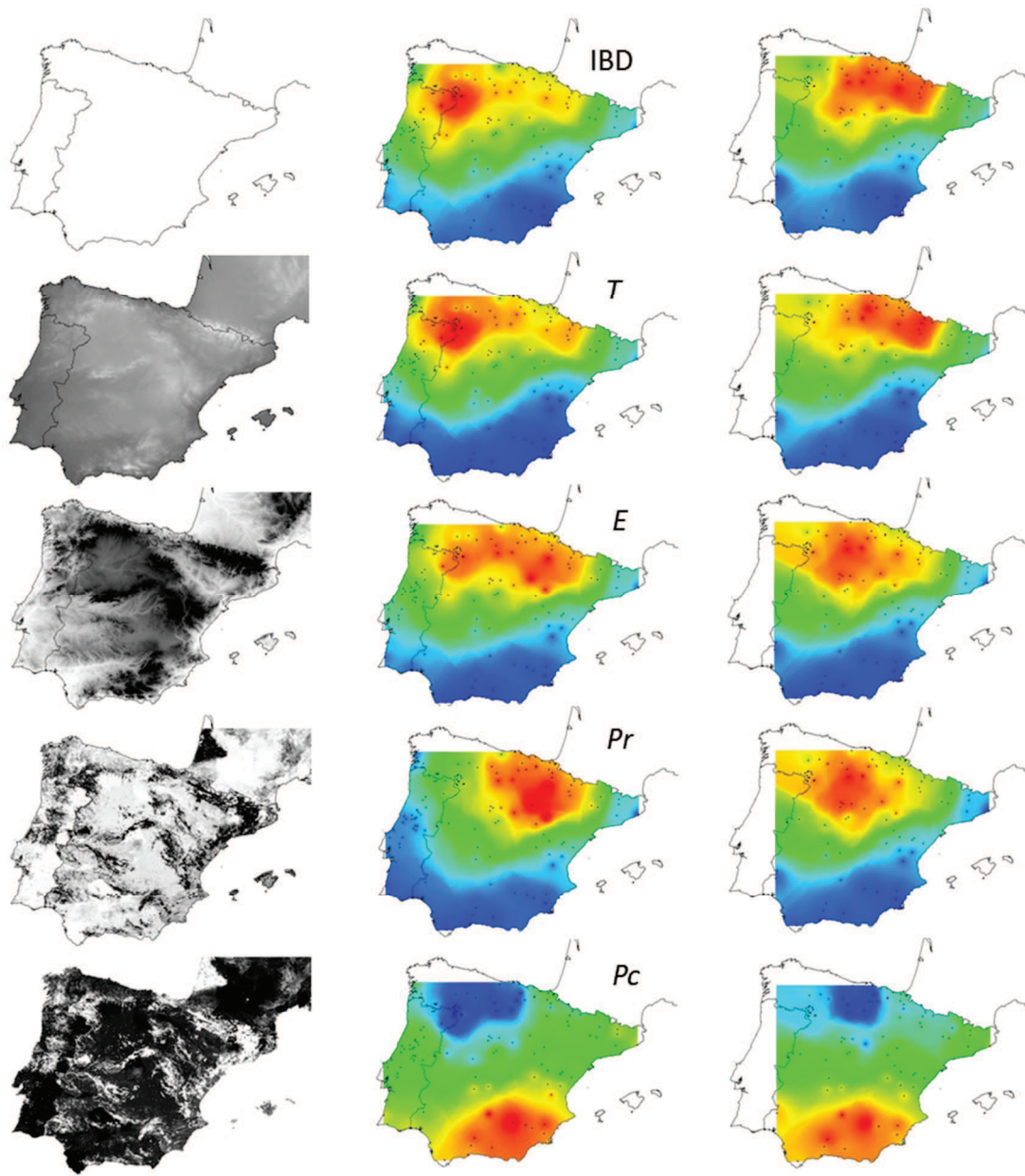


Figure 4: Distribution of supported IBR hypotheses through Mantel tests for all environmental features tested (Euclidian distances, *IBD*; mean winter temperatures, *T*; elevation, *E*; high pine densities as barriers, *Pr*; high pine densities as corridors, *Pc*). Grey maps refer to the distribution of environmental features associated with resistance models. Colored maps refer to interpolations of supported IBR hypotheses on the whole dataset (central column) and within the western Iberian cluster only (right column). From blue to red: low to high frequency of supported resistance models.

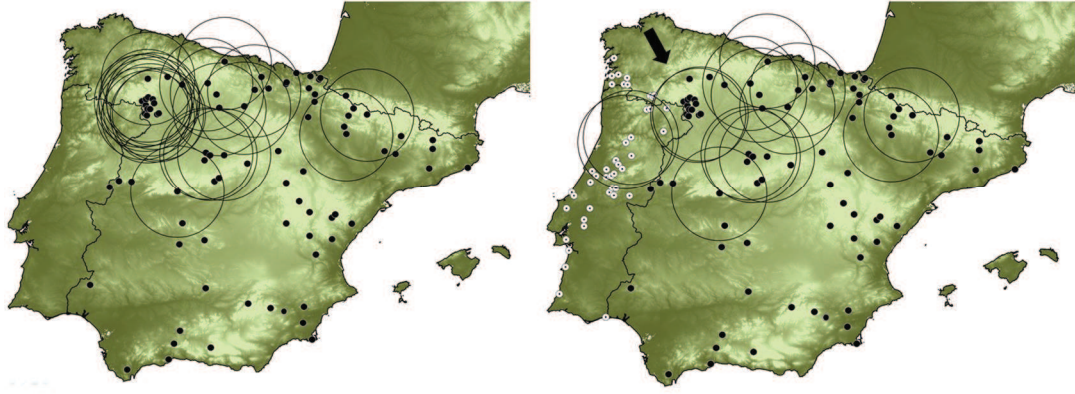


Figure 5: Example of the distribution of areas with supported IBR hypotheses through Mantel tests (background: elevation as resistance to dispersal (E); diameter of areas of 220 km) with and without the western Iberian cluster (white dotted spots).

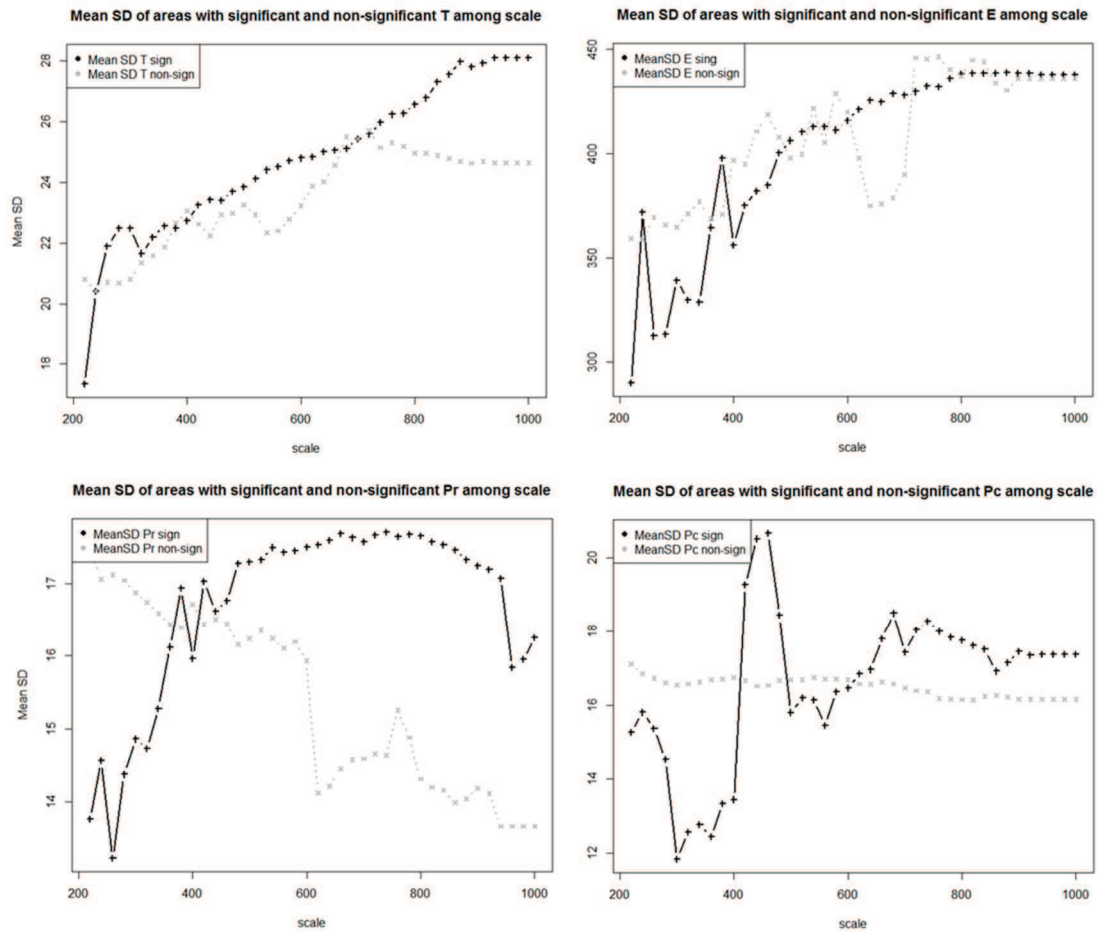


Figure 6: Development of spatial heterogeneity (mean standard deviation, SD) of environmental features in areas with supported and non-supported resistance hypotheses through Mantel test with increasing scale. Mean SD: mean standard deviation, T , E , Pr and Pc refer to IBR hypotheses tested, sign: significant, non-sign: non-significant.

	Code pop	85		130		131	
	Scale	620		540		520	
	N	225		254		244	
	IBR hypotheses	Coef.	% Total	Coef.	% Total	Coef.	% Total
<i>1st-order</i>	<i>E</i>	0,008	3,408	0,001	0,351	0,002	0,807
	<i>T</i>	0,050	20,775	0,070	32,651	0,059	28,108
	<i>Pc</i>	0,004	1,806	0,008	3,678	0,008	3,811
	<i>Pr</i>	0,085	35,235	0,047	21,817	0,046	22,214
<i>2nd-order</i>	<i>E,T</i>	0,136	56,426	0,117	54,314	0,115	54,953
	<i>E,Pc</i>	-0,003	-1,255	-0,001	-0,243	-0,001	-0,430
	<i>T,Pc</i>	0,001	0,426	0,007	3,236	0,010	4,893
	<i>E,Pr</i>	-0,002	-0,897	0,031	14,375	0,020	9,721
	<i>T,Pr</i>	-0,013	-5,398	0,018	8,406	0,014	6,540
	<i>Pc,Pr</i>	0,024	10,049	0,008	3,579	0,011	5,333
<i>3rd-order</i>	<i>E,T,Pc</i>	0,014	5,724	-0,005	-2,311	-0,001	-0,474
	<i>E,T,Pr</i>	-0,023	-9,712	-0,069	-32,007	-0,047	-22,356
	<i>E,Pc,Pr</i>	0,003	1,313	0,026	12,118	0,030	14,509
	<i>T,Pc,Pr</i>	-0,009	-3,795	0,006	2,961	0,010	4,950
	<i>E,T,Pc,Pr</i>	-0,034	-14,106	-0,049	-22,926	-0,068	-32,581
	Sum	0,240	100	0,216	100	0,209	100

Table 2: Commonality coefficients of both unique and common effects for the three sampling areas with the highest variance explained. *Code pop*: code of population of the center of sampling area. *Scale*: diameter of sampling area (km). *N*: number of individuals in sampling area. *Coef.*: percentage of variance explained by environmental features (IBR hypotheses). *% Total*: percentage of contribution of environmental features to the total variance explained.

Discussion:

Species are expected to generally show a consistent response to environmental features across their distribution range. However, our ability to make inference about the effect of environmental features may vary due to multiple evolutionary processes acting on genetic structure at different spatial and temporal scales. In this study, we explored potential barriers and corridors to dispersal and gene flow of a flying insect in a large area with a complex glacial and post-glacial history and dramatic landscape changes, which occurred recently. Based on replication of scales and locations of study areas, we showed that historically differentiated lineages may impede the detection of IBR hypothesis when not considered properly in the sampling design. We also found evidence for consistent effects of environmental features on gene flow at both local and large scales, but we observed a heterogeneous distribution of these effects among locations, especially at the lowest spatial scales.

Implications of genetic clusters for inference

The Iberian Peninsula was one of the most important Pleistocene glacial refugia for European biota. Numerous species exhibit deeply divergent lineages within this area resulting from the complexity of glacial oscillations and the associated landscape changes (Gomez & Lunt 2006). The genetic split observed between western and eastern Iberian populations of *M. galloprovincialis* was consistent with general patterns of endemism in the Iberian Peninsula (García-Barros et al. 2002) and with genetic patterns of other pine-associated insects (Rousselet et al. 2010, Tavares et al. 2014). The western Iberian lineage showed a clear divergence from the rest of the Iberian populations. Its distribution followed the range of the western cluster of *P. pinaster* in the Iberian Peninsula, which was presumably a refugium for this tree (Bucci et al. 2007) and the insect. In this study, we found that historically differentiated lineages may cause bias in landscape genetics analysis. Our results showed that the effect of environmental features was less often detected when sampled areas were occupied by two distinct genetic clusters, especially at the smallest scales (220-300 km). Two processes may explain such an effect. (i) It is first assumed that the occurrence of a deep genetic divergence between genetic clusters within a unique analysis may mask the genetic structure derived from the effect of environmental features on dispersal, resulting in type II errors (false negative). (ii) This bias may also be due to non-homogenous dispersal between genetic clusters due to specific environmental conditions. This case is expected if individuals preferentially disperse within local environmental conditions due to local adaptation (Pfuger & Balkenhol 2014), which is a common situation in nature (Sexton et al. 2014). Differences in response to environmental features cause erroneous results in correlation analysis, because such analysis usually assumes that dispersal is similar over the whole area of study.

Our results call for specific attention to consider genetic structure and evolutionary context when designing sampling areas in landscape genetics analysis. This is especially true for species with good dispersal abilities (large and/or flying organisms). For such organisms, inference of the effect of environmental features on dispersal and gene flow is usually assessed at large spatial scales to match dispersal distances (Zeller et al. 2012), and large area are more likely to include several evolutionary

lineages. The genetic structure of *M. galloprovincialis* in the Iberian Peninsula was suitable to explore the effects of such historical genetic structure. The genetic divergence between the western and eastern lineage was strong and the limits of the clusters were quite sharp, making their effect on inference obvious. It is however more difficult to evaluate the effect of weaker and more diffuse historical genetic structure on the detection of supported IBR hypotheses. Such a structure was observed in our study in the northern part of Spain, and it was potentially also derived from quaternary glacial oscillations. In which degree historical genetic structure are problematic in landscape genetics analysis is a stimulating question that should be addressed in future work. We suggest that a larger sampling area increases the signal of the environmental features on gene flow by gathering more individuals from both lineages and “balance” their conflicting effect. Thus, widening the spatial scale of area of a study could represent a solution to overcome bias due to evolutionary history.

Effect of scale and location on inference

We observed a notable influence of scale on the detection of supported IBR hypotheses with Mantel tests for most environmental features tested (*E*, *T* and *Pr*). Support was scarcely detected at the lowest spatial scale (220-400 km) and generally more often detected with increasing scale. Indeed, 190-400 km corresponded to the distances at which the variogram given in Figure 2 showed a first plateau of genetic dissimilarity. This correspondence suggested that at this range of scales, dissimilarity between individuals was often not appropriate to show a significant effect of environmental features on gene flow. By contrast, the peak (for *T*) or inflection of curves (for *Pr*, *E*) of number of areas with supported IBR hypotheses observed at scales ranging from 400 to 600 km corresponded to the increase of dissimilarity in the variogram. Thus, scales above 400 km seemed more appropriate to gather a genetic structure in *M. galloprovincialis* that was determined by the environmental features. Interestingly, we observed that the development of the variation of frequency of areas with support was specific to each environmental features tested. Similar results were observed for a large mammal (Zeller et al. 2014) and for insects (Rasic & Keyghobadi 2012) when multiple scales were considered.

The shape of the variogram showed important autocorrelation of genotypes below scales of 190 km. This autocorrelation indicated that genotypes were quite similar and that genetic structure was weak for distances up to about 200 km. Weak genetic differentiation at such scale was shown, based on estimates of population differentiation (F_{st}), for *M. alternatus* and *M. galloprovincialis* in lowland valleys (Kawai et al. 2006, Shoda-Kagaya 2007, Haran et al. 2015). Direct measures of the dispersal ability of *Monochamus* species show that adults may fly over long distances, ranging from 2 to 22 km in the field (Takasu et al. 2000, Linit & Akbulut 2003, Hernandez et al. 2011, Mas et al. 2013, David et al. 2014). These flight performances are thought to cause intensive gene flow and generate the weak genetic structure observed in this study at small spatial scales. This weak genetic structure was sufficient to detect IBD in a large proportion of the areas at small scales (<220 km), but IBR was rarely supported at such scales. Our results illustrate a general problem of landscape genetics analysis performed on species with good dispersal abilities. This is particularly true for flying species, which are naturally less affected by environmental features than non-flying species. For such species, the combination of intensive dispersal and gene flow and a limited number of environmental features affecting dispersal make inference difficult at small spatial scales (Dreier et al. 2014). Considering a continuous range of scales in analysis prevented us to base our conclusions at a scale at which the effect of environmental features could not be detected. Our observations are consistent with the cases of large mammals for which multiple scales, including very large scales, have been used to deal with uncertainties regarding scale of gene flow (Galpern et al. 2012, Zeller et al. 2014).

Based on landscape-level replications of sampling areas across the Iberian Peninsula, we have shown the existence of a heterogeneous distribution of supported resistance models. Most variation in the distribution of support for IBR was observed at small and intermediate scales (220-600 km). Supported effects were mainly detected in the north-central part of the Iberian Peninsula. Conversely, effects were less supported in the rest of Iberian Peninsula (center, south and coasts). Interestingly, the distribution of genetic clusters identified by STRUCTURE showed a consistent pattern. Most diversity was observed in the upper part of the Iberian Peninsula. By contrast, the other parts of the Peninsula exhibited a single genetic cluster. Two hypotheses may explain this spatial heterogeneity in supported

resistance models. A first hypothesis is that differences of variation in environmental features exist across replication. An area exhibiting contrasted environmental features is known to affect more dispersal and thus to increase the chance to detect their effect (Short Bull et al. 2011, Cushman et al. 2013). However, our results showed that at the smallest spatial scales, variation of environmental features in areas with supported IBR hypothesis was not higher than for non-supported areas, for most resistance models. This indicated that the distribution of variation of environmental features was not the main factor determining heterogeneity in support of resistance models. A second hypothesis is the existence of a conflicting signal due to inclusion of different evolutionary lineages in a study, as discussed above. Such an effect probably explained the lack of support along the western Iberian coast. Indeed, the western Iberian cluster formed a narrow band, and areas of study almost systematically overlapped with the eastern cluster there. Within the eastern Iberian cluster however, we observed a lack of detection of supported IBR hypotheses in areas that covered only one cluster (eastern Iberian coast). A large part of the heterogeneity was therefore not due to conflicting signal between clusters or evolutionary lineages.

These results suggested that at scales of about 220-600 km, *M. galloprovincialis* was structured according to environmental features in some areas but not in others, independently of artifacts or variations of these environmental features. This observation is surprising, because a native species is generally expected to show a homogeneous response to environmental features, at least within a genetic lineage. Determining the exact origin of such heterogeneity is challenging. We believe that this variation was a legacy of changes in distribution of host trees in the Iberian Peninsula. Distribution and density of pine trees have been strongly affected by anthropogenic activities during the last centuries (Ruiz-Benito et al. 2012, Lopez-Merino et al. 2014), resulting in local extinction, but also connectivity and fragmentation of pine tree cover across time. For example, Abel-Schaad et al. (2014) have shown that pine trees locally disappeared from the Central Iberian System during the middle ages. By contrast, these areas have been afforested at 80% with pines trees during 1940-1950. It is assumed that such recent modifications have dramatically affected the distribution and abundance of *M. galloprovincialis*, and that time since these modifications happened is too short for appearance

in the genetic structure of the beetle. This is especially expected in a species showing intensive gene flow and for which the effect of environmental features is better detected at large spatial scales. Lag times until new barriers become detectable is an important field of exploration in landscape genetics (Murphy et al. 2008, Landguth et al. 2010). The effect of difference in the dispersal abilities of different organisms on the length of such time lags is an interesting question that should be addressed in future research.

Strength of the effects of environmental features

For all environmental features tested, average partial Mantel coefficients (r) of supported IRB hypotheses were moderate (from 0.045 to 0.371), but consistent with values observed for wide-ranging organisms like large mammals (Short Bull et al. 2011, Galpern et al. 2012) or flying insects (Dupas et al. 2014). The decrease of mean Mantel r with increasing scale was interesting. It suggested that in areas exhibiting support for IRB hypotheses, correlation is stronger at small spatial scales than at larger scales. Such a situation is expected because larger areas in this study (600-1000 km) often harbored two evolutionary lineages derived from glacial oscillations, and we have shown that such spatial overlap among lineages could impede inference in landscape genetics analysis. Conversely, small areas with support for IRB hypotheses showed the highest mean Mantel r values. This result suggested that areas with significant IRB hypotheses exhibit a “pure” effect with less conflicting signal. Therefore, our results show a tradeoff between sampling small areas where effects of environmental features are strong but scarcely detected and sampling large surfaces, where this effect is weaker but often detected.

Elevated areas are barriers to dispersal for *M. galloprovincialis*

One of our hypotheses was that elevated area constitutes barriers to gene flow for *M. galloprovincialis*. The two resistance models (T and E) support this hypothesis (Fig. 3) and corroborate observation made on *M. alternatus* across the Ohu chain mountain in Japan (Shoda-

Kagaya 2007) and on *M. galloprovincialis* across the Pyrenees (Haran et al. 2015). Several factors may explain this result. Temperature affects larval development and survival in *M. galloprovincialis* (Naves & Sousa 2009) and its ability to complete its development in one or two years (Tomminen 1993, Naves et al. 2007b, Koutroumpa et al. 2008,). Therefore, low winter temperatures likely constitute a factor that prevents migration across elevated area by impeding or slowing species development. Topography may also explain the effect of elevation on dispersal. For example, Torres-Vila et al. (2015) have shown using mark-release-recapture experiments that adults tend to fly downhill.

The effect of pine densities

Effect of pine on dispersal was modeled according two mutually exclusive hypotheses: high densities of pines represent barriers (*Pr*) or corridors (*Pc*) to dispersal. Our results show that *M. galloprovincialis* is mainly structured according to the first hypothesis. Therefore, this species seem not to disperse when its resource is abundant. This result contradicts with expectations of a species performing its life cycle on declining pine trees. Indeed, the scattered nature of such resource is expected to enhance dispersal along pine forests. However, *M. galloprovincialis* is known to develop on dead branches stemming from a self-pruning process encountered in pines. Dead branches represent a resource that is more widely distributed in space and time than dead trees. The philopatric behavior of *M. galloprovincialis* in relation to available resources is consistent with observation of flight of this species in the field (Torres-Vila et al. 2015), or with the behavior of the pine processionary moth (*Thaumetopoea pityocampa*), another oligophagous pine-associated insect (Demolin 1969). The prevalence of the first hypothesis suggests, conversely, that low pine densities are not barriers to dispersal. This is in agreement with the suggestions of Torres-Vila et al. (2015) that dispersal of *M. galloprovincialis* tends to be enhanced across open areas. In fact, the Iberian Peninsula contains several wide areas where pine tree forests are absent (center of Castilla y Leon for example), and our results suggested that such areas do not represent barriers to dispersal. Rossi et al. (2015) have

shown using observed and simulated data that areas without pine forests still show a homogenous distribution of scattered trees planted for ornamental use. We suggest that pine trees out of forests are providing a scattered but homogeneously distributed of resource that allows dispersal of *M. galloprovincialis* across non-forested areas.

Our results showed that genotypes of *M. galloprovincialis* are also structured according the second hypothesis (P_c) in some areas (Fig. 3 and 4). This hypothesis was encountered at a smaller frequency in the south of the study area. The geographic consistency of this supported hypothesis tends to show that this signal corresponds to a true effect of this environmental feature and is not originated from type I error that may rise with the use of partial Mantel tests (Guillot & Rousset 2013). This subtle conflicting signal indicates that in some areas pines may represent corridors to dispersal of *M. galloprovincialis*, namely in areas where the species showed weaker degree of genetic structure (South and South east of the peninsula). We suggest that recent population expansion along re-planted pine forests may explain such effect.

Conclusions

In this study, we highlighted that elevated areas and dense pine cover constitute barriers to dispersal of *M. galloprovincialis*. We also showed that this species exhibit intensive gene flow at local scale of less than about 200 km. These results are generally consistent with previous studies on the genetic structure and dispersal abilities of this species. In the context of the invasion of the pinewood nematode, our results suggest that natural dispersal of this nematode may be enhanced in flat and open areas. However, higher elevations and pine forests could slow down the expansion of this pest across the Iberian Peninsula and potentially the whole Europe.

Our results exemplify the importance to simultaneously consider a continuous range of scales and multiple locations when exploring the effect of environmental features on dispersal in highly mobile species. Multiple scales allow to infer the effect of environmental features at the appropriate

extent for each features tested while preventing to focus analysis at an extent where intensive gene flow make inference impossible due to lack of genetic structure. In addition, replications can also help to identify variation in inference due to conflicting signals in genetic structure, and therefore allow for generalizing conclusions regarding the effects of environmental features on dispersal and gene flow. As a result, the combination of both multiple spatial scale and replications across various locations in landscape genetics analysis provides more general picture of the effects of environmental features on gene flow and reduces the versatility of results.

Acknowledgments

This work was supported by the European project REPHRAME KBBE.2010.1.4-09 (FP7 Project, Analysis of the potential of the pine wood nematode (*Bursaphelenchus xylophilus*) to spread, survive and cause pine wilt in European coniferous forests in support of EU plant health policy). First author was funded by the French Ministry of Research and Education. We warmly thank Rolf Holderegger and Bertrand Gauffre for valuable comments on early versions of this manuscript. We also thank Jérôme Rousselet and Christelle Robinet for interesting exchanges and help regarding the methodology.

References

- Abel-Scaad D, Lopez-Saez JA, Pulido F, 2014. Heathlands, fire and grazing. A paleoenvironmental view of Las Hurdes (Cáceres, Spain) history during the last 1200 years. *Forest Systems* 23, 247–258.
- Anderson CD, Epperson BR, Fortin MJ, Holderegger R, James PMA et al., 2010. Considering spatial and temporal scale in landscape-genetic studies of gene flow. *Molecular Ecology* 19, 3565–3575.
- Angelone S, Kienast F, Holderegger R, 2011. Where movement happens: scale-dependent landscape effects on genetic differentiation in the European tree frog. *Ecography* 34, 714–722.
- Bucci G, Gonzalez-Martinez SC, Le Provost G, Plomion C, Ribeiro MM, 2007. Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Molecular Ecology* 16, 2137–2153.
- Burrough PA, 1983. Problems of superimposed effects in statistical study of the spatial variation in soil. *Agricultural Water Management* 6, 123–143.
- Castillo JA, Epps CW, Davis AR, Cushman SA, 2014. Landscape effects on gene flow for a climate-sensitive montane species, the American pika. *Molecular Ecology* 23, 843–856.
- Chandra AK, Raghavan P, Ruzzo WL, Smolensky R, Tiwari P, 1997. The electrical resistance of a graph captures its commute and cover times. *Computational Complexity* 6, 312–340.
- Chapuis MP, Estoup A, 2007. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24, 621–631.
- Cushman SA, Landguth E, 2010. Scale dependent inference in landscape genetics. *Landscape Ecology* 25, 967–979.
- Cushman SA, McKelvey KS, Hayden J, Schwartz MK, 2006. Geneflow in complex landscapes: testing multiple models with causal modeling. *American Naturalist* 168, 486–499.
- Cushman SA, Shirk AJ, Landguth E, 2013. Landscape genetics and limiting factors. *Conservation Genetics* 14, 263–274.
- David G, Giffard B, Piou D, Jactel H, 2014. Dispersal capacity of *Monochamus galloprovincialis*, the European vector of the pine wood nematode, on flight mills. *Journal of Applied Entomology* 138, 566–576.
- Démolin G, 1969. Comportement des adultes de *Thaumetopoea pityocampa* Schiff. Dispersion spatiale, importance écologique. *Annales des Sciences Forestières* 26, 89–102.
- Diniz-Filho JAF, Soares TN, Lima JS, Dobrovoski R, Landeiro VL et al., 2013. Mantel test in population genetics. *Genetics and Molecular Biology* 36, 475–485.
- Dreier S, Redhead JW, Warren IA, Bourke AF, Heard MS et al., 2014. Fine-scale spatial genetic structure of common and declining bumble bees across an agricultural landscape. *Molecular Ecology* 23, 3384–3395.
- Drizen K, Adriaensen F, Rondinini C, Doncaster CP, Matthysen E, 2007. Evaluating least-cost model predictions with empirical dispersal data: a case-study using radiotracking data of hedgehogs (*Erinaceus europaeus*). *Ecological Modelling* 209, 314–322.

- Dudaniec RY, Rhodes JR, Wilmer JW, Lyons M, Lee KE et al., 2013. Using multilevel models to identify drivers of landscape-genetic structure among management areas. *Molecular Ecology* 22, 3752–3765.
- Dupas S, Le Ru B, Branca A, Faure N, Gigot G et al., 2014. Phylogeography in continuous space: coupling species distribution models and circuit theory to assess the effect of contiguous migration at different climatic periods on genetic differentiation in *Busseola fusca* (Lepidoptera: Noctuidae). *Molecular Ecology* 23, 2313–2325.
- Earl DA, VonHoldt BM, 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4, 359–361.
- 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.
- Galpern P, Manseau M, Wilson P, 2012. Grains of connectivity: analysis at multiple spatial scales in landscape genetics. *Molecular Ecology* 21, 3996–4009.
- Garcia-Barros E, Gurrea P, Lucianez MJ, Cano JM, Munguira ML, Moreno LC, et al., 2002. Parsimony analysis of endemism and its application to animal and plant geographical distributions in the Ibero-Balearic region (western Mediterranean). *Journal of Biogeography* 29, 109–124.
- Gómez A, Lunt DH, 2006. Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. *Phylogeography of southern European refugia – evolutionary perspectives on the origins and conservation of European biodiversity* (ed. by S. Weiss and N. Ferrand), pp. 155–188. Kluwer, Dordrecht.
- Goovaerts P, 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, Oxford.
- Guillot G, Rousset F, 2013. Dismantling the Mantel tests. *Methods in Ecology and Evolution* 4, 336–344.
- Haran J., Roques A., Barnard A., Robinet C., Roux G., 2015. Altitudinal barrier to the spread of an invasive species: could the Pyrenean chain slow the natural spread of the pine wood nematode? *PLoS ONE* 10(7):e0134126.
- Haran J, Roux-Morabito G, 2014. Development of 12 microsatellites loci for the longhorn beetle *Monochamus galloprovincialis* (Coleoptera Cerambycidae), vector of the pinewood nematode in Europe. *Conservation Genetics Resources* 6, 975–977.
- Hellrigl KG, 1971. La bionomie des espèces de *Monochamus* (Coleoptera, Cerambycidae) et leur importance pour la sylviculture et l'économie du bois. *Redia* 52, 367–511.
- Hernandez R, Ortiz A, Pérez V, Gil JM, Sanchez G, 2011. *Monochamus galloprovincialis* (Olivier, 1795) (Coleoptera: Cerambycidae), comportamiento y distancias de vuelo. *Boletín de Sanidad Vegetal Plagas* 37, 79–96.
- Hewitt GM, 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405, 907–913.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965–1978.
- Jaquière J, Broquet T, Hirzel AH, Yearsley J, Perrin N, 2011. Inferring landscape effects on dispersal from genetic distances: how far can we go? *Molecular Ecology* 20, 692–705.

- Jombart T, 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405.
- J van Etten, 2012. gdistance: distances and routes on geographical grids. R package version 1.1-4.
- Kawai M, Shoda-Kagaya E, Maehara T, Zhou ZH, Lian CL et al., 2006. Genetic structure of pine sawyer *Monochamus alternatus* (Coleoptera: Cerambycidae) populations in northeast Asia: Consequences of the spread of pine wilt disease. *Environmental Entomology* 35, 569–579.
- Keller D, Holderegger R, van Strien MJ, 2013. Spatial scale affects landscape genetic analysis of a wetland grasshopper. *Molecular Ecology* 22, 2467–2482.
- Kierepka EM, Latchi EK, 2015. Performance of partial statistics in individual-based landscape genetics. *Molecular Ecology* 15, 512–525.
- Kindall JL, Van Manen FT, 2007. Identifying habitat linkages for American black bears in North Carolina, USA. *Journal of Wildlife Management* 71, 487–495.
- Koutroumpa F, Vincent B, Roux-Morabito G, Martin C, Lieutier F, 2008. Fecundity and larval development of *Monochamus galloprovincialis* (Coleoptera Cerambycidae) in experimental breeding. *Annals of Forest Science* 65, 707.
- Legendre P, Legendre L, 1998. Numerical Ecology. Elsevier, Amsterdam.
- Linit MJ, Akbulut S, 2003. Pine wood nematode phoresis: the impact on *Monochamus carolinensis* life functions. *Nematology Monographs and Perspectives* 1, 227–237.
- Lopez-Merino L, Martinez Cortizas A, Reher GS, Lopez-Saez JA, Mighall TM, Bindler R, 2014. Reconstructing the impact of human activities in a NW Iberian Roman mining landscape for the last 2500 years. *Journal of Archaeological Science* 50, 208–218.
- Manel S., Holderegger R, 2013. Ten years of landscape genetics. *Trends in Ecology & Evolution*, 28, 614–621.
- Manel S, Schwartz MK, Luikart G, Taberlet P, 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18, 189–197.
- Mas H, Hernandez R, Villaroya G, Sanchez G, Pérez-Laorga E et al., 2013. Dispersal behavior and long distance flight capacity of *Monochamus galloprovincialis* (Olivier 1795), In: Schröder, T (ed.), Pine Wilt Disease Conference 2013, pp. 22, Braunschweig, ISSN: 1866-590X.
- McRae B, 2006. Isolation by Resistance. *Evolution* 60, 1551–1561.
- Murphy MA, Evans JS, Cushman SA, Storfer A, 2008. Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. *Ecography* 31, 685–697.
- Naves P, de Sousa E, 2009. Threshold temperatures and degree-day estimates for development of post-dormancy larvae of *Monochamus galloprovincialis* (Coleoptera: Cerambycidae). *Journal of Pest Science* 82, 1–6.
- Naves PM, Camacho S, de Sousa E, Quartau JA, 2007. Transmission of the pine wood nematode *Bursaphelenchus xylophilus* through oviposition activity of *Monochamus galloprovincialis* (Coleoptera : Cerambycidae). *Entomologica Fennica* 18, 193–198.

- Naves P, Sousa E, Quartau J, 2006. Feeding and oviposition preferences of *Monochamus galloprovincialis* for some conifers under laboratory conditions. *Entomologica Experimentalis et Applicata* 120, 99–104.
- O'Brien RM, 2007. A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity* 41, 673–690.
- Pflüger FJ, Balkenhol N, 2014. A plea for simultaneously considering matrix quality and local environmental conditions when analyzing landscape impacts on effective dispersal. *Molecular Ecology* 23, 2146–2156.
- Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Prunier JG, Colyn M, Legendre X, Nimon KF, Flamand MC, 2014. Multicollinearity in spatial genetics: separating the wheat from the chaff using commonality analyses. *Molecular Ecology* 24, 263–283.
- Prunier JG, Kaufmann B, Fenet S, Picard D, Pompanon F, Joly P, 2013. Optimizing the trade-off between spatial and genetic sampling efforts in patchy populations: towards a better assessment of functional connectivity using an individual-based sampling scheme. *Molecular Ecology* 22, 5516–5530.
- Rasic G, Keyghobadi N, 2012. From broad scale patterns to fine-scale processes: habitat structure influences genetic differentiation in the pitcher plant midge across multiple spatial scales. *Molecular Ecology* 21, 223–236.
- Raymond M, Rousset F, 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86, 248–249.
- R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ribeiro PJ, Diggle PJ, 2001. geoR: a package for geostatistical analysis. *R-NEWS* 1, 14–18.
- Robertson GP, Gross KL, 1994. Assessing the heterogeneity of belowground resources: quantifying pattern and scale. In: Caldwell M, Pearcy R (eds.), *Exploitation of Environmental Heterogeneity by Plants*. Academic Press, pp. 237–253.
- Rossi JP, 2003. Short-range structures in earthworm spatial distribution. *Pedobiologia* 47, 582–587.
- Rossi JP, Garcia J, Roques A, Rousselet J, 2015. Trees outside forests in agricultural landscapes: spatial distribution and impact on habitat connectivity for forest organisms. *Landscape Ecology*. doi: 10.1007/s10980-015-0239-8.
- Rossi JP, Nardin M, Godefroid M, Ruiz-Diaz M, Sargent AS, et al., 2014. Dissecting the space-time structure of tree-ring data sets using the partial triadic analysis. *PLoS ONE* 9, e108332.
- Rousselet J, Zhao RX, Argal D, Simonato M, Battisti A et al., 2010. The role of topography in structuring the demographic history of the pine processionary moth, *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae). *Journal of Biogeography* 37, 1478–1490.
- Ruiz-Benito P, Gomez-Aparicio L, Zavala MA, 2012. Large-scale assessment of regeneration and diversity in Mediterranean planted pine forests along ecological gradients. *Diversity and Distributions* 18, 1092–1106.

- Serra P, Vera A, Francesc Tulla A, Salvati L, 2014. Beyond urbanerural dichotomy: exploring socioeconomic and land-use processes of change in Spain (1991-2011). *Applied Geography* 55, 71–81.
- Sexton JP, Hangartner SB, Hoffmann AA, 2014. Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution* 68, 1–15.
- Shirk AJ, Wallin DO, Cushman SA, Rice CG, Warheit KI, 2010. Inferring landscape effects on gene flow: a new model selection framework. *Molecular Ecology* 19, 3603–3619.
- Shoda-Kagaya E, 2007. Genetic differentiation of the pine wilt disease vector *Monochamus alternatus* (Coleoptera: Cerambycidae) over a mountain range – revealed from microsatellite DNA markers. *Bulletin of Entomological Research* 97, 167–174.
- Short Bull RA, Cushman SA, Mace R, Chilton T, Kendall KC, et al., 2011. Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology* 20, 1092–1107.
- Storey JD, 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B* 64, 479–498.
- Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, 2007. Putting the “landscape” in landscape genetics. *Heredity* 98, 128–142.
- Takasu F, Yamamoto N, Kawasaki K, Togashi K, Kishi Y et al., 2000. Modeling the expansion of an introduced tree disease. *Biol Invasions* 2, 141–150.
- Tavares C, Roques A, Courtial B, Branco M, Jactel H, Lopez-Vaamonde C, 2014. Phylogeography of the ladybird *Iberorhizobius rondensis*, a potential biological control agent of the invasive alien pine bast scale *Matsucoccus feytaudi*. *BioControl* doi: 10.1007/s10526-014-9622-7.
- Tomminen J, 1993. Development of *Monochamus galloprovincialis* Olivier (Coleoptera, Cerambycidae) in cut trees of young pines (*Pinus sylvestris* L.) and log bolts in southern Finland. *Entomologica Fennica* 4, 137–142.
- Torres-Vila L, Zugasti C, De-Juan JM, Olivia MJ, Montero C, Mendiola FJ et al., 2015. Mark-recapture of *Monochamus galloprovincialis* with semiochemicalbaited traps: population density, attraction distance, flight behaviour and mass trapping efficiency. *Forestry* 88, 224–236.
- Tröltzsch K, Van Brusselen J, Schuck A, 2009. Spatial occurrence of the major tree species group in Europe derived from multiple data sources. *Forest Ecology and Management* 257, 294–302. doi:10.1016/j.foreco.2008.09.012
- Van Strien MJ, Keller D, Holderegger R, Ghazoul J, Kienast F, Bolliger J, 2014. Landscape genetics as a tool for conservation planning: predicting the effects of landscape change on gene flow. *Ecological Applications* 24, 327–339.
- Zalewski A, Pierniey SB, Zalewska H, Lambin X, 2009. Landscape barriers reduce gene flow in an invasive carnivore: geographical and local genetic structure of American mink in Scotland. *Molecular Ecology* 18, 1601–1615.
- Zeller KA, McGarigal K, Beier P, Cushman SA, Winston Vickers T, Boyce WM, 2014. Sensitivity of landscape resistance estimates based on point selection functions to scale and behavioral state: pumas as a case study. *Landscape Ecology* 29, 541–557.
- Zeller KA, McGarigal K, Whiteley AR, 2012. Estimating landscape resistance to movement: a review. *Landscape Ecology* 27, 777–797.

Zellmer AJ, Knowles LL, 2009. Disentangling the effects of historic vs. contemporary landscape structure on population genetic divergence. *Molecular Ecology* 18, 3593–3602.

Supporting information

Table S1: Sampling details of the 132 demes. (Long. and Lat. refer to geographic coordinates of sampling sites; N. is the number of individuals of demes)

Code	Locality	Country	Long.	Lat.	N.
Eastern Iberian lineage					
1	Zamora Tabaras	Spain	-6,04	41,87	10
2	Zamora ferreras	Spain	-6,08	41,84	10
3	Zamora Villar	Spain	-6,31	41,92	10
4	Justel	Spain	-6,29	42,14	8
5	San Vitero	Spain	-6,36	41,77	10
6	Puebla De Sanabria	Spain	-6,58	42,05	8
7	Manzanal Infantes	Spain	-6,23	42,06	5
8	Espadanedo	Spain	-6,43	42,16	2
9	Manzana De Arriba	Spain	-6,45	41,98	9
10	Mahide	Spain	-6,38	41,91	5
11	Figueruela De Arriba	Spain	-6,43	41,80	5
12	Rosinos	Spain	-6,48	42,11	10
13	Ponferrada	Spain	-6,55	42,55	12
14	Quintana de Castillo	Spain	-6,00	42,68	10
15	Leon	Spain	-5,54	42,58	3
16	Villalba de Guardo	Spain	-4,82	42,72	15
17	Lamason	Spain	-4,45	43,25	7
	Villanúno de				
18	Valdavia	Spain	-4,51	42,52	10
19	Cuellar	Spain	-4,36	42,25	6
20	Burgos	Spain	-3,66	42,38	2
21	Ona	Spain	-3,44	42,77	9
22	Villanueva de Mena	Spain	-3,31	43,07	5
23	Espejo	Spain	-3,05	42,85	3
24	Vitoria	Spain	-2,57	43,01	7
25	Aramayona	Spain	-2,56	43,05	5
26	Itziar	Spain	-2,32	43,28	14
27	Irun	Spain	-1,91	43,33	18
28	Sare	France	-1,60	43,26	18
29	Etchalar	Spain	-1,63	43,22	10
30	Lekunberri	Spain	-1,89	43,00	2
31	Marcalain	Spain	-1,69	42,89	11
32	Beriain	Spain	-1,67	42,73	10
33	Falces	Spain	-1,79	42,38	14
34	Col St Martin	Spain	-0,79	42,96	10
35	Canfranc	Spain	-0,52	42,71	17

36	Jaca	Spain	-0,64	42,54	16
37	Huesca 1	Spain	-0,72	42,29	18
38	Huesca 2	Spain	-0,63	42,14	7
39	Torla	Spain	-0,11	42,62	1
40	La Pobla de Segur	Spain	0,94	42,25	18
41	Castillonroy	Spain	0,52	41,88	10
42	Camarassa	Spain	0,84	41,85	6
43	Baga	Spain	1,86	42,24	18
44	Olvan	Spain	1,88	42,05	18
45	Castellbell	Spain	1,84	41,62	30
46	Capmany	Spain	2,96	42,36	11
47	Argelès	France	3,06	42,52	10
48	Tossa de Mar	Spain	2,93	41,72	10
49	Atzeneta	Spain	-0,15	40,21	10
50	Alcala	Spain	-0,64	40,36	7
51	Mora	Spain	-0,72	40,28	10
52	Cuenca Altura	Spain	-0,64	39,82	10
53	Requena	Spain	-1,03	39,44	9
54	Cuenca Talayuelas	Spain	-1,27	39,82	10
55	Albaracin	Spain	-1,36	40,33	9
56	Cuenca Palancares	Spain	-1,96	40,01	8
57	Oribuela	Spain	-1,69	40,53	1
58	Cobeta	Spain	-2,13	40,89	12
59	Muneberga	Spain	-1,70	41,20	10
60	Almazan	Spain	-2,49	41,55	7
61	Catalojas	Spain	-3,30	41,17	8
62	Catalejo	Spain	-3,99	41,26	10
63	Segovia Coca	Spain	-4,51	41,24	10
64	Segovia Navas	Spain	-4,38	41,20	10
65	Segovia Santa Maria	Spain	-4,51	41,07	9
66	Santa Maria	Spain	-4,51	41,08	10
67	Guadarrama	Spain	-4,13	40,66	4
68	Cercedilla	Spain	-4,02	40,75	12
69	El Arenal	Spain	-5,10	40,28	12
70	Espinoso Del Rey	Spain	-4,80	39,62	10
71	Puebla De DonRodrigo	Spain	-4,78	39,16	7
72	Los Yebenes	Spain	-4,12	39,35	1
73	El Centimillo	Spain	-3,87	38,33	1
74	Santiago Murcia	Spain	-2,67	38,16	10
75	Murcia Bebedor	Spain	-2,05	38,14	10
76	Cehegin	Spain	-1,73	38,06	3
77	Murcia Puerto	Spain	-1,16	38,28	9
78	Murcia Valle	Spain	-1,13	37,93	4
79	Huerta-Cartagena	Spain	-0,83	37,60	1
80	Monachil	Spain	-3,48	37,08	7
81	Rute	Spain	-4,36	37,32	9
82	Antequera	Spain	-4,46	37,02	2
83	El Pastor	Spain	-4,53	36,66	8
84	Los Barrios	Spain	-5,56	36,27	6
85	Aroche	Spain	-6,96	37,90	11

86	Ovejuela	Spain	-6,41	40,29	10
87	El Payo	Spain	-6,76	40,25	10

Western Iberian lineage

88	Monte Gordo	Portugal	-7,45	37,18	12
89	Odeceixe	Portugal	-8,77	37,43	8
90	Santo Andres	Portugal	-8,77	38,02	12
91	Comporta	Portugal	-8,78	38,38	4
92	Palmela	Portugal	-8,92	38,59	1
93	Coruche	Portugal	-8,53	38,96	3
94	Santarem	Portugal	-8,57	39,10	10
95	Chamusca	Portugal	-8,48	39,36	3
96	cadaval	Portugal	-9,10	39,24	3
97	Alcobaça	Portugal	-8,98	39,54	4
98	Nazaré	Portugal	-9,06	39,60	1
99	Pombal	Portugal	-8,63	39,91	1
100	Pombal	Portugal	-8,63	39,92	2
101	Montemor	Portugal	-8,65	40,16	2
102	Coimbra	Portugal	-8,52	40,08	12
103	Penacova	Portugal	-8,28	40,27	1
104	Tondela	Portugal	-8,09	40,49	12
105	Tondela 2	Portugal	-8,08	40,51	2
106	Carregal do sal	Portugal	-7,99	40,43	3
107	Oliviera do hospital	Portugal	-7,86	40,36	1
108	Arganil	Portugal	-8,05	40,22	3
109	Gois	Portugal	-8,11	40,16	4
110	Lousa	Portugal	-8,24	40,11	1
111	Alvelos	Portugal	-8,01	39,91	2
112	Oleiros	Portugal	-7,91	39,92	5
113	Serta	Portugal	-8,06	39,79	1
114	Proenca A Nova	Portugal	-7,92	39,75	1
115	Castelo branco	Portugal	-7,50	39,82	4
116	Vale Feitoso	Portugal	-6,98	40,07	19
117	PenlavaDoCastelo	Portugal	-7,69	40,68	1
118	Castro Daire	Portugal	-7,92	40,92	24
119	Lamego	Portugal	-7,81	41,06	6
120	Alfandega Da Fe	Portugal	-6,96	41,34	3
121	Vinhais	Portugal	-7,01	41,83	2
122	Chaves	Portugal	-7,46	41,74	1
123	Redondelo	Portugal	-7,55	41,71	12
124	Vences	Spain	-7,46	42,00	2
125	Matama	Spain	-7,43	42,04	8
126	Ourense	Spain	-7,85	42,26	12
127	Ventin	Portugal	-8,39	42,29	1
128	Ribeira	Spain	-8,23	42,13	1
129	Riva Crecente	Spain	-8,29	42,27	1
130	Amoedos	Spain	-8,55	42,29	1
131	Viga Islas	Spain	-8,90	42,23	3
132	Villagarcia de Aroussa	Spain	-8,78	42,61	9

Table S2: (for table of primer details, please refer to Appendix n°1 of article n°1)

Genotyping protocol

Multiplexed PCR were performed in a 10 µL reaction volume using 25 ng of genomic DNA, 0.4 U of DreamTaq DNA Polymerase (Thermo Scientific®), 0.75 µL Dream Taq Green Buffer (including 20 mM MgCl₂, Thermo Scientific®), 1 µM Betaine, 0.24 µL dNTP (10 µM) and deionized H₂O. PCR amplifications were run on a Veriti® 96 well fast Thermal cycler (Applied Biosystems®) using the following settings: a first denaturation step at 95 °C during 10 min; 40 cycles of denaturation (30 s at 95 °C), hybridization (30 s at 55 °C) and elongation (1 min at 72 °C), and a final elongation step at 72 °C during 10 min. One µL of PCR products were denatured within a mix of 10 µL of formamide and 0.3 µL of 600 Liz marker before being run on an ABI PRISM 3500 sequencer (Life Technologies®). Genotypes were read using the software GENEMAPPER V 4.1 (Applied Biosystems®).

Figure S1: Distribution of sampling sites in the Iberian Peninsula. Black dots refer to populations of size > 19 individuals. Green background refers to elevation (from pale to dark green: low to high elevation).

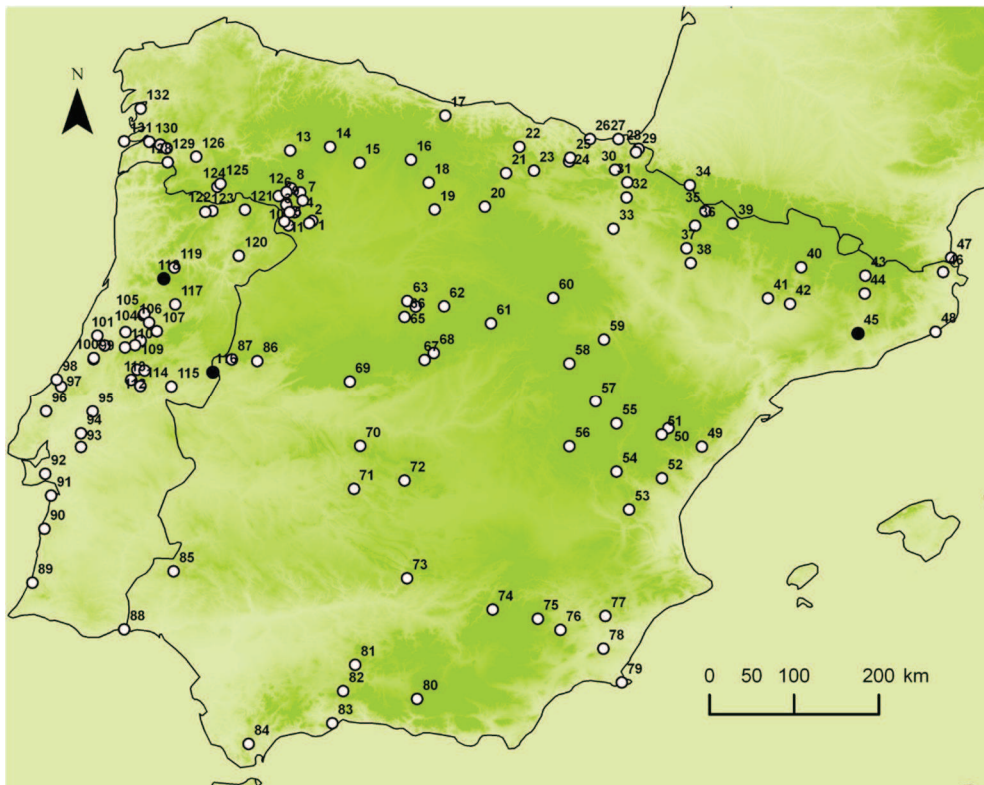
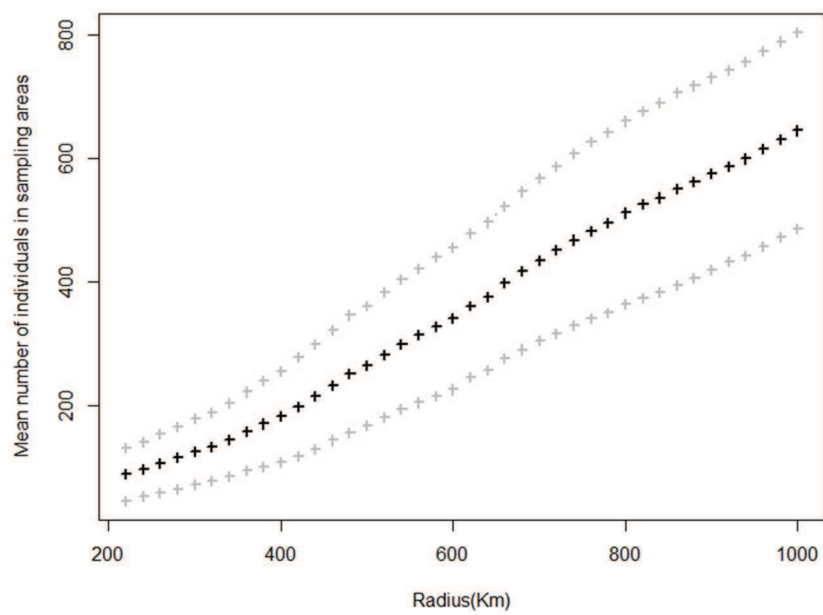


Figure S2: Number of individuals in sampling areas across spatial scale (Mean: black; +/- SD: grey).



Chapitre IV

**Synergie potentielle de la dispersion du nématode et
de *M. galloprovincialis* dans un milieu hétérogène :
le cas de la barrière Pyrénéenne**

Chapitre IV : Synergie potentielle de la dispersion du nématode et de *M. galloprovincialis* dans un milieu hétérogène : le cas de la barrière Pyrénéenne

Dans les chapitres II et III, j'ai étudié les flux de gènes chez l'espèce vectrice *M. galloprovincialis* à plusieurs échelles spatiales, afin d'identifier les barrières potentielles à l'expansion du PWN. Dans ce chapitre, je cherche à explorer le potentiel de dispersion du PWN par une étude intégrant à la fois une étude des flux de gènes du vecteur et une approche de modélisation de la dispersion prenant en compte l'effet synergique de l'association entre les deux partenaires. Cette étude a été menée au niveau d'une barrière naturelle qui constitue une zone stratégique dans l'expansion du PWN : la chaîne pyrénéenne. Cette étude se rapporte à l'article n°3.

Article n°3: Haran J, Roques A, Bernard A, Robinet C, Roux G, 2015. Altitudinal barrier to the spread of an invasive species: could the pyrenean chain slow the natural spread of the pinewood nematode? *PLoS ONE* 10(7): e0134126.

-Contexte et objectifs :

Les chaînes de montagne européennes constituent des barrières naturelles à la dispersion des espèces endémiques et ont fortement affecté la distribution de leur lignées évolutives (Taberlet et al. 1998, Hewitt 1999, Schmitt 2007). Les ruptures que ces chaînes forment dans la connectivité du paysage peuvent donc potentiellement arrêter ou freiner l'expansion d'espèces invasives. Les Pyrénées forment une chaîne de montagne relativement élevée (jusqu'à 3404 m) qui s'étend sur une longueur de plus de 400 km entre la France et l'Espagne. Par cette disposition, elle se trouve au niveau de la seule zone terrestre entre la péninsule ibérique et le reste de l'Europe. Cette chaîne constitue donc une barrière potentielle dans l'invasion du nématode du pin en Europe.

Dans cette étude, j'ai cherché à explorer le potentiel d'invasion du nématode à travers la chaîne Pyrénéenne. Pour cela, j'ai combiné une approche de génétique des populations pour (i) évaluer les flux de gènes du vecteur au niveau de cette barrière naturelle et de modélisation pour (ii) simuler l'invasion du couple PWN-vecteur à travers le relief en intégrant l'aspect synergique de leur association.

-Matériel et Méthodes :

Pour cette étude, nous avons échantillonné 485 adultes de l'espèce vectrice (*M. galloprovincialis*) sur un ensemble de 37 sites de part et d'autre du relief, formant quatre transects nord-sud. Les patrons de flux géniques ont tout d'abord été étudiés par des approches classiques de génétique des populations. Nous avons évalué le niveau de différenciation entre populations par calcul de F_{st} (Weir & Cockerham 1984) et évalué les flux de gènes en mesurant le niveau d'admixture entre populations des transects (Choisy et al. 2004). La structure génétique globale des populations a été évaluée par regroupement bayésien (Pritchard et al. 2000) et nous avons utilisé les tests d'assignation bayésiens (Piry et al. 2004) pour identifier les migrants de première génération entre populations.

Nous avons utilisé un modèle de type réaction-diffusion pour simuler la dispersion de populations de vecteurs infectés par le PWN au sein de l'aire d'étude. Ce type de modèle simule à la fois la dispersion et la croissance d'une population. Cependant, nous avons dû adapter ce modèle à la spécificité du modèle *Monochamus*-PWN et à l'effet de synergie qui lui est associé. En particulier, nous avons intégré la préexistence de populations de vecteurs non infestées dans la zone d'étude et l'évolution de leur densité consécutive à l'arrivée du PWN. De telles estimations ont été obtenues grâce aux densités de vecteurs piégés dans les zones infestées et non infestées par le PWN. Ces densités ont ensuite été reliées aux densités en pins et extrapolées à l'ensemble de l'aire d'étude. Nous avons également intégré la dynamique d'infestation des populations de vecteurs par le PWN en prenant en compte la latence dans la transmission de ce nématode. Plus précisément, nous avons considéré de manière schématique les quatre étapes chronologiques suivantes : (1) le nématode arrive à l'année $t = n$ en une certaine localité et avant son arrivée ($t < n$) il n'existe pas de vecteurs infestés ;

(2) l'arrivée du PWN entraîne un affaiblissement des arbres et par conséquent une augmentation de la densité de vecteurs par mise à disposition de ressource (arbres dépérissant). A cette étape, il n'existe pas de vecteurs infestés au sein des populations locales puisque ceux-ci émergeront l'année d'après. Il n'y a donc pas « recrutement » de ces populations locales ; (3) l'année suivante ($t = n + 1$), les populations locales sont « recrutées » car les individus émergeant des arbres infestés portent le PWN et le dispersent ; (4) les années suivantes ($t > n + 1$), l'ensemble de la population est infestée et disperse le PWN. Il n'existe pas de recrutement supplémentaire. Ce modèle décrit ainsi de manière simplifiée, mais plus réaliste qu'un modèle d'expansion classique, la synergie potentielle entre le nématode et son vecteur.

Nous avons appliqué ce modèle à notre zone d'étude en considérant un pool initial de 200 adultes infestés situé sur le versant espagnol de la zone d'étude. Nous avons ensuite suivi l'expansion de ces populations infestées sur un nombre de génération suffisant pour que l'ensemble de la zone étudiée soit colonisée. Nous avons simulé cette expansion en faisant varier les seuils de survie estimés du vecteur en fonction de l'altitude (en supposant l'absence de transmission du PWN dans les zones où le vecteur ne se maintient pas) pour décrire l'effet du relief sur la dispersion du PWN. Nous avons également testé l'effet de l'expression de la PWD sur cette expansion, en considérant plusieurs scénarii de changement climatique.

-Résultats :

Au sein de la zone étudiée, *M. galloprovincialis* est structuré en deux lignées génétiques distribuées de part et d'autre de la chaîne pyrénéenne. L'estimation des flux de gènes entre populations montre que les zones les plus élevées du relief constituent une barrière à la dispersion de ce coléoptère. En revanche, nous avons observé un flux de gènes important en plaine et aux extrémités est et ouest de ce relief, où l'altitude est plus faible. Les migrants de première génération ont été détectés sur des patrons similaires, principalement entre populations de plaine d'un même versant et entre populations au niveau des versants est et ouest du relief.

Les résultats de nos simulations ont montré une forte cohérence avec les patrons génétiques. Les simulations montrent que les populations infestées par le nématode du pin ne rencontrent pas de barrières à la dispersion en plaine, même dans des zones où les densités de vecteurs sont très faibles. Nos résultats ont également montré que la chaîne pyrénéenne constitue une barrière partielle à l'expansion du PWN de l'Espagne vers la France. Les divers seuils de survie du vecteur et d'expression de la PWD définissent des zones défavorables à l'expansion du nématode au niveau des régions élevées du relief, mais laissent un corridor à cette expansion aux extrémités latérales de la chaîne. De manière générale, nos résultats ont montré que les seuils d'expression de la maladie définissent des barrières potentielles plus larges à la dispersion du PWN que les seuils basés sur la survie du vecteur.

RESEARCH ARTICLE

Altitudinal Barrier to the Spread of an Invasive Species: Could the Pyrenean Chain Slow the Natural Spread of the Pinewood Nematode?

Julien Haran^{1,2*}, Alain Roques¹, Alexis Bernard¹, Christelle Robinet¹, Géraldine Roux^{1,2}

1 INRA, UR633 Zoologie Forestière, F-45075 Orléans, France, **2** Université d'Orléans, Orléans, France

* julien.haran@orleans.inra.fr



OPEN ACCESS

Citation: Haran J, Roques A, Bernard A, Robinet C, Roux G (2015) Altitudinal Barrier to the Spread of an Invasive Species: Could the Pyrenean Chain Slow the Natural Spread of the Pinewood Nematode? PLoS ONE 10(7): e0134126. doi:10.1371/journal.pone.0134126

Editor: João Pinto, Instituto de Higiene e Medicina Tropical, PORTUGAL

Received: January 7, 2015

Accepted: July 6, 2015

Published: July 29, 2015

Copyright: © 2015 Haran et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded by the European project REPHRAME KBBE.2010.1.4-09 (FP7 Project, Analysis of the potential of the pine wood nematode (*Bursaphelenchus xylophilus*) to spread, survive and cause pine wilt in European coniferous forests in support of EU plant health policy). First author was funded by the French Ministry of Research and Education.

Abstract

Mountain ranges may delimit the distribution of native species as well as constitute potential barriers to the spread of invasive species. The invasive pinewood nematode, *Bursaphelenchus xylophilus*, is a severe forest pest inducing pine wilt disease. It is vectored in Europe by a native long-horned beetle, *Monochamus galloprovincialis*. This study explored the potential of the Pyrenean chain to slow or prevent the natural spread of nematode-infested beetles from the Iberian Peninsula, where the nematode is established and is expanding its range, towards France and the rest of Europe. An analysis of the genetic structure and migration patterns of the beetle populations throughout the Pyrenean mountain range was combined with a spread model simulating the potential movements of nematode-infested beetles across it. The central part of the Pyrenees, which corresponds to the highest elevation zone, was shown to prevent gene flow between the French and Spanish populations of *M. galloprovincialis* on each side of the mountains. Conversely, strong admixture was detected between populations located on both sides of low elevation hills, and especially at the east and west extremities of the mountain range. Simulations of the spread of nematode-infested beetles under various thresholds of beetle survival and pine wilt disease expression gave results consistent with the variation in genetic make-up, suggesting that western and eastern hillsides may represent corridors favoring natural spread of the nematode from the Iberian Peninsula to France. Simulations also showed that temperature rise due to climate change may significantly reduce the extent of the barrier formed by highest elevations. Our results support the hypothesis that the Pyrenean chain represents a partial barrier to the natural spread of nematode-infested beetles. These results, which have to be considered together with potential human-assisted long-distance spread of the nematode, highlight priority zones for future pest monitoring and management programs. More generally, such an integrated approach could be used to assess the role of mountain chains in the potential spread of other invasive pests.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Introduction of species beyond their native ranges has increased dramatically over recent decades due to intensification of international trade [1, 2, 3]. When successfully established in new areas, introduced species may expand their ranges and can cause major environmental disturbances together with significant economic losses [4]. The local range expansion of introduced organisms depends on their dispersal ability, which is a critical parameter for development of containment measures [5]. In most cases, this expansion occurs in heterogeneous environments, where the spatial and temporal distributions of biotic and abiotic constraints vary [6]. This heterogeneity determines the influence and scale of landscape effects on dispersal of species [7]. Depending on biological and physical constraints affecting the dispersing species, the spatial distribution of these constraints may constitute negative barriers or positive corridors determining range expansion of the invading species. Identifying environmental factors underlying the efficiency of dispersal of invasive species is essential in developing suitable management measures. Mountainous areas are major components of landscape heterogeneity, exhibiting contrasting climatic conditions that have historically shaped the genetic structures of species by affecting connectivity of landscapes [8, 9, 10]. The Pyrenean chain is a major mountain range in South-Western Europe. Due to its high elevation (up to 3404 m) and its spatial extent (more than 400 km long) this chain is an altitudinal obstacle between the Iberian Peninsula and the rest of Europe that has strongly affected the distributions and genetic structures of native species. Phylogeographic studies have highlighted the role of this barrier in shaping intraspecific lineages [8] and as a major contact zone for post glacial range expansion in European biota [11]. Owing to the effect of this barrier on the dispersal of native organisms, the Pyrenean chain could therefore play an important role in slowing the spread of invasive species from the Iberian Peninsula to the rest of Europe or conversely.

The pinewood nematode (PWN), *Bursaphelenchus xylophilus* (Steiner & Burher, 1934) Nickle, 1970 (Nematoda, Aphelenchoididae) is the causal agent of the pine wilt disease (PWD). Under suitable climatic conditions, this pest is able to kill susceptible pine trees within a few months [12]. Though it causes limited damage in its native range in North America [13], it has resulted in massive mortality to native pine forests [14] in its area of introduction in East Asia (Japan-1905, China-1982, Taiwan-1985, Korea-1988). PWN was detected for the first time in Europe in Portugal in 1999, in the peninsula of Setubal [15]. Despite intensive containment measures, it has quickly expanded its range through most of Portugal and has entered Spain, where it is under eradication [16]. The natural dispersal of the PWN is exclusively done through the activities of longhorned beetles in the genus *Monochamus* (Coleoptera, Cerambycidae) [17]. So far, the widely distributed *Monochamus galloprovincialis* is the only known PWN vector in Europe [18]. *M. galloprovincialis* females oviposit in stressed or freshly dead pine trees, after which the larvae develop and pupate within the wood [19]. The nematodes migrate to the pupal chamber and moult to a specific larval stage that enters the tracheae of callow adult *Monochamus* prior to their emergence. Transmission of PWN takes place either during maturation feeding in the crowns of living pines (primary transmission) and also during oviposition on declining trees (secondary transmission) [17, 20, 21]. In both cases, individuals carrying the nematode (here after called PWN-infested beetles) can fly and disperse the nematode. Consequently, the natural spread (i.e. non-human mediated) of PWN in Europe depends mainly on the dispersal abilities of the vector beetle.

Little is known on the effective dispersal of *M. galloprovincialis* in the field and how environmental parameters affect this dispersal. Based on flight mill experiments, the physiological flight capacity of this beetle was estimated to be an average of 16 km over the lifespan of the beetle [22] whereas mark-release-recapture experiments conducted in Spain revealed that adult

beetles may disperse up to 22 km in the field [23]. These methods are known to overestimate and underestimate dispersal abilities respectively [22, 24]. In addition, mountain chains are complex environments where variations in topography, pine densities, pine species and temperatures are likely to play a significant role in dispersal of this beetle, and may act as a disrupting factor. To date, no data are available on environmental conditions influencing dispersal of *M. galloprovincialis*.

Microsatellite markers are broadly used to infer population genetic structure [25], gene flow between populations [26], individual migration events [27, 28] and to assess the effects of landscape on dispersal [7]. Estimates of the genetic structure of *M. galloprovincialis* populations across the Pyrenees can thus provide valuable information on the location of either barriers or corridors to dispersal of this species, and therefore act as determinants of the range expansion of PWN. However, for native species such as *M. galloprovincialis*, the observed genetic structure also results from past demographic processes whose time scales may be difficult to determine. Consequently, it is challenging to infer precisely the current dispersal abilities of the vector from the genetic structure of the sampled populations [29]. To test dispersal in current environmental conditions, model-based simulations represent a consistent and complementary approach. Spread models are widely used to describe and simulate potential dispersal of invasive species [30, 31, 32, 33]. Although based on the current known biology of the species, models allow the testing of several scenarios under current and future environmental conditions. Such spread models have been used to predict the spread of the PWN and associated PWD in Japan [34, 35, 36] and China [37]. In Europe, a large-scale spread model was applied to identify the most likely factors determining spread of PWN, taking into account short-distance flight as well as human-mediated longer-distance transportation of infested individuals of *M. galloprovincialis*. This model assumes that the spread potential in Europe is similar to that already observed in China [38]. To date, no study has focused on the potential spread of PWN-infested populations at local scale, in an environment exhibiting strong landscape heterogeneity.

Dispersal of PWN results from a combination of natural vector-mediated spread and human long-distance transportations. Given the progression of the PWN in the Iberian Peninsula [39] and the possible economic consequences for European countries [40], it is crucial to consider both quarantine policy and potential barriers to the natural spread of this alien species in order to define suitable management measures. The aim of the present study is to focus on the potential natural dispersal of PWN through the Pyrenean chain, which is a strategic boundary between the infested area (Iberian Peninsula) and the rest of Europe. Environmental and physical conditions associated with this mountain range can potentially affect the dispersal of *M. galloprovincialis* and the spread of PWN. We address the following questions: 1) Does the Pyrenean chain represents a barrier to natural dispersal of *M. galloprovincialis*? 2) Could this chain slow or stop the natural spread of PWN under current and future climate conditions? To answer these questions, we combined a population genetic analysis and a modeling approach, using microsatellite markers to elucidate the genetic structure and migration patterns of *M. galloprovincialis* and a spread model to simulate the invasion of PWN-infested beetles across this mountain range.

Materials and Methods

Field sampling

The sampling sites were selected in order to cover both sides of the entire Pyrenean chain. They included 9 sites on the Northern side (France) of the Pyrenees and 13 on the Southern side (Spain), forming 4 North-South transects (Fig 1). To detect potential effects of environmental factors (temperature, host tree species and elevation) on genetic differentiation and gene flow of *M. galloprovincialis*, these sampling sites were located along a gradient of elevation on each side

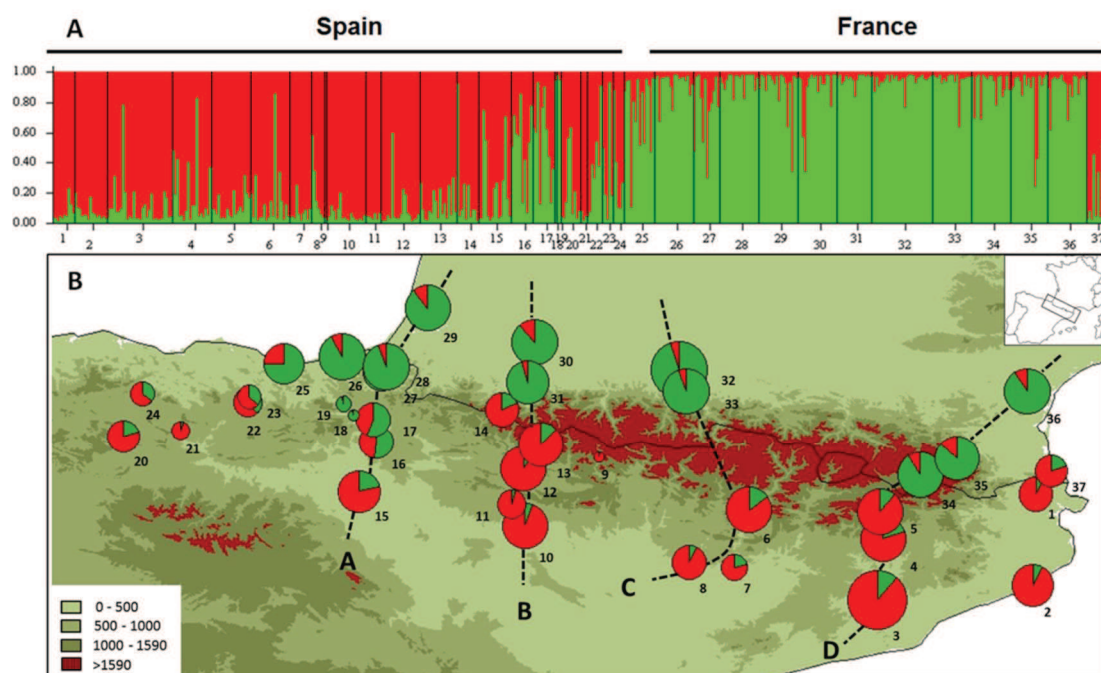


Fig 1. Bayesian clustering of individuals and populations of *M. galloprovincialis*. A) Barplots of individual assignment of the 485 individuals for $K = 2$. B) Membership of populations (sum of membership of individuals). Background refers to elevation.

doi:10.1371/journal.pone.0134126.g001

of the Pyrenean chain. Transects were located close to the main road axes crossing the Pyrenees, as they are mostly situated in valleys. These areas thus represent potential corridors for the dispersal of this beetle through natural or human mediated dispersal (S1 Fig). Additional sites were sampled on the Western side of the chain (populations 25 and 26) and along the Mediterranean Sea (populations 1 and 37) to estimate gene flow along the low elevation coastal areas. In order to increase the coverage of the sampling, several additional populations and isolated individuals were included in this study (2, 9, 18–24). Specimens were trapped at all locations on private land with permission of the land owners. No more specific permissions were required and these field studies did not involve endangered or protected species. Specimens were captured using multifunnel traps and a specific volatile attractant (Galloprotect, SEDQ, Spain). A set of 3 traps was installed on each site from late June to the end of October and specimens were collected every 3 weeks. After identification to species level, specimens were stored in 99.5° ethanol at 4°C. *M. galloprovincialis* was recorded at all sites except in Vielha (Table 1). Despite four months of trapping, this species was difficult to capture in 3 localities above an altitude of 1400 m (Font-Romeu, Col de la Pierre Saint Martin and Canfranc). Overall, we obtained 26 samples (430 specimens) supplemented by specimens from 11 sites collected previously, giving a total number of 37 sampling sites and 485 specimens to be genotyped.

DNA extraction and microsatellite analysis

Genomic DNA was extracted from two legs per individuals using a Nucleospin Kit (Macherey–Nagel, Düren, Germany), following manufacturer's instructions. Each DNA extract was equilibrated with sterile water to a concentration ranging from 10 to 30 ng/μL. Individuals were

Table 1. Details of aggregates of *M. galloprovincialis* considered across the Pyrenees.

Code	Location	Country	Transect	Year	Long.	Lat.	Elevation (m)	<i>Pinus</i>	N	A. rich.	P.	Ho	He	Fis
1	Capmany	Spain	—	2014	2.95	42.36	100	<i>P. halepensis</i>	10	3.03	0.02	0.483	0.586	0.155
2	Tossa de Mar	Spain	—	2012	2.93	41.72	21		15	3.04	0	0.455	0.496	0.153
3	Castellbell	Spain	D	2013	1.84	41.62	190	<i>P. halepensis</i>	30	3.42	0.06	0.493	0.574	0.146
4	Olvan	Spain	D	2013	1.88	42.05	520	<i>P. pinea/nigra</i>	18	3.25	0.04	0.486	0.481	-0.015
5	Baga	Spain	D	2013	1.86	42.24	794	<i>P. sylvestris</i>	18	3.29	0.08	0.474	0.52	0.104
6	La Pobla de S.	Spain	C	2013	0.94	42.25	646	<i>P. halepensis</i>	18	3.55	0.19	0.501	0.568	0.139
7	Camarassa	Spain	—	2013	0.84	41.85	271	<i>P. halepensis</i>	6	—	—	—	—	—
8	Castillonroy	Spain	C	2013	0.52	41.88	402		10	3.31	0.03	0.439	0.494	0.101
9	Torla	Spain	—	2012	-0.11	42.62	1197		1	—	—	—	—	—
10	Huesca 1	Spain	—	2013	-0.63	42.14	419	<i>P. pinea</i>	7	—	—	—	—	—
11	Huesca 2	Spain	B	2013	-0.72	42.29	552	<i>P. halep./sylv.</i>	18	3.45	0.04	0.437	0.539	0.169
12	Jaca	Spain	B	2013	-0.64	42.54	870	<i>P. sylvestris</i>	18	3.39	0.04	0.457	0.532	0.125
13	Canfranc	Spain	B	2013	-0.52	42.71	1502	<i>P. nigra</i>	17	3.72	0.13	0.525	0.581	0.108
14	Col St-Martin	Spain	B	2013	-0.79	42.96	1590	<i>P. uncinata</i>	10	3.45	0.12	0.500	0.524	0.027
15	Falces	Spain	A	2012	-1.79	42.38	360		15	3.48	0.21	0.458	0.517	0.131
16	Beriain	Spain	A	2012	-1.67	42.73	613		10	2.51	0	0.408	0.449	0.106
17	Marcalain	Spain	A	2012	-1.69	42.89	550		10	3.13	0.07	0.430	0.494	0.088
18	Iruzun	Spain	—	2012	-1.83	42.92	553		1	—	—	—	—	—
19	Lekunberri	Spain	—	2012	-1.9	43.00	507		2	—	—	—	—	—
20	Ona	Spain	—	2012	-3.44	42.77	1055		9	—	—	—	—	—
21	Espejo	Spain	—	2012	-3.04	42.81	715		3	—	—	—	—	—
22	Vitoria	Spain	—	2012	-2.57	43.01	760	<i>P. nigra</i>	7	—	—	—	—	—
23	Aramayona	Spain	—	2012	-2.56	43.05	627		5	—	—	—	—	—
24	Villanueva de M.	Spain	—	2012	-3.31	43.07	452		5	—	—	—	—	—
25	Itziar	Spain	—	2013	-2.32	43.28	200	<i>P. nigra</i>	14	2.92	0.01	0.411	0.463	0.094
26	Irun	Spain	—	2013	-1.91	43.33	23	<i>P. pinaster</i>	18	2.95	0	0.399	0.454	0.114
27	Etchalar	Spain	A	2013	-1.63	43.22	273	<i>P. radiata</i>	12	2.8	0	0.298	0.447	0.286
28	Sare	France	A	2013	-1.6	43.26	303	<i>P. nigra</i>	18	2.74	0	0.370	0.449	0.184
29	St-Vincent	France	A	2013	-1.31	43.67	32	<i>P. pinaster</i>	18	2.89	0	0.370	0.435	0.092
30	Orthez	France	B	2013	-0.56	43.43	180	<i>P. nigra</i>	18	2.88	0	0.300	0.424	0.303
31	Oloron	France	B	2013	-0.61	43.15	302	<i>P. strobus</i>	16	2.76	0.03	0.368	0.417	0.091
32	Vieuzos	France	C	2013	0.45	43.24	449	<i>P. nigra laricio</i>	28	2.7	0.03	0.345	0.428	0.198
33	Garros	France	C	2013	0.5	43.09	560	<i>P. nigra laricio</i>	18	2.79	0.03	0.306	0.391	0.145
34	Font-Romeu	France	D	2013	2.14	42.5	1432	<i>P. sylvestris</i>	18	2.83	0.02	0.415	0.474	0.138
35	Prades	France	D	2013	2.4	42.61	410	<i>P. halepensis</i>	17	2.87	0.01	0.400	0.460	0.149
36	Font froide	France	D	2013	2.89	43.09	235	<i>P. pinaster</i>	18	2.75	0.01	0.317	0.421	0.219
37	Argelès	France	—	2013	3.06	42.52	45	<i>P. halepensis</i>	9	—	—	—	—	—
—	Vielha	Spain	—	2013	0.82	42.69	1280	<i>P. sylvestris</i>	0	—	—	—	—	—
Mean South side		Spain								3.329	0.084	0.467	0.522	0.102
(Mean±SD)										0.299	0.065	0.034	0.039	0.051
Mean North side		France								2.823	0.011	0.358	0.438	0.167
(Mean±SD)										0.07	0.01	0.044	0.023	0.072

Code: identification number of aggregates (see Fig 1); Long./Lat., geographic coordinates; *Pinus*, dominant host tree species in sampling site; N, population size; A. rich., mean allelic richness over loci calculated under rarefaction; P., average private allelic richness over loci; He and Ho, expected and observed heterozygosity; F_{is} , multilocus F_{is} estimate; “—”, aggregate not considered as a population for genetic analysis.

doi:10.1371/journal.pone.0134126.t001

genotyped at 12 loci previously described (Mon01, Mon08, Mon17, Mon23, Mon27, Mon30, Mon31, Mon36, Mon41, Mon42 & Mon44) [41]. Multiplexed PCR were performed in a 10 μ L reaction volume using 1 μ L of genomic DNA, 0.4 U of DreamTaq DNA Polymerase (Thermo Scientific), 0.75 μ L Dream Taq Green Buffer (including 20 mM $MgCl_2$, Thermo Scientific), 1 μ M Betaine, 0.24 μ L dNTP (10 μ M) and deionized H_2O . PCR amplifications were run on a Veriti 96 well fast Thermal cycler (Applied biosystems) using the following settings: a first denaturation step at 95°C over 10 min; 40 cycles of denaturation (30 s at 95°C), hybridization (30 s at 55°C) and elongation (1 min at 72°C); and a final elongation step at 72°C over 10 min. One μ L of PCR product was denatured within a mix of 10 μ L of formamide and 0.3 μ L of 600 Liz marker before being run on an ABI PRISM 3500 sequencer (Life Technologies). Genotypes were read using the software GeneMapper V 4.1 (Applied Biosystems). For primer sequences and multiplex details please refer to Haran & Roux [41].

Basic genetic parameters (Heterozygote frequencies: H_e and H_o) and F_{is} were calculated using Genalex 6.5 [42]. The mean allelic richness and private allelic richness of each population was computed by regression using HP-RARE [43], to correct bias involved by unequal sampling size. Deviation from HWE at each locus was tested using Genepop 4.2 [44], with a False Discovery Rate correction [45]. We also used Genepop 4.2 [44] to test the presence of linkage disequilibrium between loci. The frequency of null alleles at each locus and for each population was estimated using FreeNA [46]. Signatures for recent demographic events such as bottlenecks among populations were assessed using the program Bottleneck [47]. Deviations from mutation drift equilibrium implemented in this program were computed using the Stepwise Mutation Model (SMM, 95%) and the Two-Phase Model (TPM, 5%). Significance of deviations were tested using the Wilcoxon test, following recommendations of authors for microsatellite analysis [47].

Population differentiation and clustering

Population differentiation was estimated using the unbiased Pairwise F_{st} implemented in the program FreeNA [46]. We used the *ENA* method to correct the potential positive bias induced by the occurrence of null alleles on F_{st} estimation. Tests for population differentiation significance were made using FSTAT v. 2.9.3.2 [48] based on 325,000 permutations (level for significant differentiation set to $p < 0.0001$). The estimation of unbiased genetic distance based on allele frequency requires a minimum population size. An approximate number of 20 individuals per population is usually advised for structured populations ($F_{st} > 0.05$) [43]. Due to sampling difficulties, nine scattered populations were represented by a smaller number of individuals ($N < 16$). These populations were included in analysis for their contribution to the spatial coverage, bearing in mind that population genetic estimates of differentiation have to be carefully considered in these cases. Isolation by distance between populations was tested through a Mantel test implemented in Genepop 4.2 [44]. Correlations were computed between linear geographic distances and pairwise genetic differentiation ($F_{st} / 1 - F_{st}$) for 10,000 permutations. Isolation by distance was assessed on the whole dataset and within each side of the mountain range.

The genetic clustering of populations was computed using the Bayesian approach provided by the software STRUCTURE 2.3.4 [28]. STRUCTURE attributes to each individual a likelihood of membership to a predefined number of clusters, minimizing deviation from linkage DIS-equilibrium between loci. Analyses were performed using the “correlated allele frequencies” without “sampling populations as prior”. We used the “admixture model” due to the relatively small scale of the study and because *M. galloprovincialis* has good flight performances [22], reaching up to 22.1 km per year [23]. The “optimal” number of clusters (K) was determined using the Delta-K method [49]. The genotypes were analyzed by running 10 repeats of a 100,000 burn-in period followed by 500,000 replicates of Markov Chain Monte Carlo

(MCMC), for K values ranging from 1 to 10. We used the online service STRUCTURE HARVESTER [50] to visualize the evolution of Delta-K (mean likelihood of K/ variance of likelihood for the same K) through the values of K tested. The value of K maximizing likelihood and minimizing variance was chosen as optimal for this dataset. Among the 10 repeats, the one showing the highest mean value of Ln likelihood was selected for graphical representation. For control purposes, clustering analysis was performed with and without loci exhibiting significant null allele frequency. We used ArcGis 9.3 (ESRI) to plot the membership of individuals (Qvalue) to the predefined clusters for each population.

Detection of migration

We estimated the proportion of admixture along the four transects to detect the signature of migration events across the Pyrenean chain. Admixture in populations was computed using the coalescent-based maximum Likelihood method implemented in the program LEA [51]. This method was chosen due to its broad applicability and its very low bias for ancient hybridization [52]. LEA estimates the likelihood of a hybrid population to belong to two predefined parental populations. Along each transect, we defined parental populations at the two extremities. Admixture was calculated for each population between these parental populations, based on 10 computations of 11,000 iterations.

The frequency and origin of first generation migrants was estimated using assignment tests implemented in GeneClass 2.0 [53]. We used the Bayesian method [54] to test the likelihood of each individual to belong to any of the 26 populations including the one of origin (Likelihood (L) = L population of origin / L maximum in all populations). Calculations were performed based on the Monte Carlo resampling method [55], using 10,000 simulated genotypes, with a threshold for migrants detection set to 0.001. For such analysis, we corrected null alleles in the dataset to minimize the risk of bias in the indices. The loci significantly affected (>10%) were corrected using the INA method [46] assuming a single null allele state shared by all populations.

Modeling the potential spread of infested populations of *M. galloprovincialis* across the Pyrenees

We used a mathematical model to explore the effect of physical and biological parameters on the spread of a hypothetical infested population of *M. galloprovincialis* across the Pyrenees (from Spain to France). We modeled the spread of an endemic vector species (*Monochamus sp.*) carrying an invasive organism (PWN) and able to transmit this invasive species to resident vector populations. In this case study, each partner of this PWN-beetle association is affected by the dynamics of the other. The PWN needs the vector to disperse and the vector benefits from the mortality of trees caused by the PWN to find more suitable resources and reproduce. Thus, the model has to consider both sides of the association: the pre-existence of natural population of *M. galloprovincialis* in the area of study (resident populations) and the effect of PWN outbreaks on the densities of this species. The potential spread of PWN and its damage at relatively local scale (i.e., non-human mediated dispersal) was described by a reaction-diffusion model [37, 38]. This model simulates both population dispersal and logistic growth:

$$\frac{\partial N}{\partial t} = D \left(\frac{\partial^2 N}{\partial x^2} + \frac{\partial^2 N}{\partial y^2} \right) + f(N) \quad (1)$$

$$f(N) = \epsilon N \left(1 - \frac{N}{K} \right) \quad (2)$$

where N is the population density of *M. galloprovincialis*, x and y the geographic coordinates (in km), and t the time (i.e., the year as the vector has one generation per year in the geographic area considered in this study). D is the diffusion coefficient (km^2/year). We used the value estimated in China for *Monochamus alternatus*, $D = 6.48 \text{ km}^2/\text{year}$ [37], due to the lack of accurate data to estimate this parameter more precisely in Europe. f is a logistic growth function with ε the growth rate and K the carrying capacity of the local environment. The growth rate, $\varepsilon = 2.17$, was estimated based on the ratio of host trees killed by PWN in China [37]. As there are insufficient data to estimate this parameter in Europe, we employed this estimate in our model. Values of D and ε were also used in a previous spread model applied to PWN in Europe [38]. The carrying capacity, K , is the maximum population density that the local environment can maintain. K depends on the number of declining host trees available for the development of *M. galloprovincialis*. This parameter depends whether the PWN is present or not.

Here, we explicitly describe the symbiotic relationship between the PWN and its vector considering that the carrying capacity changes at the PWN arrival. We assume that $K = K_0$ before and $K = K_i$ after PWN arrival, with a ratio between K_i and K_0 estimated from trapping PWN-infested and PWN-free areas in Portugal. The parameter K_0 was estimated using results of trapping in the Pyrenees (where the PWN is not yet present) and pine tree density nearby. In addition, we assume that the density of the resident population of the vector before the arrival of PWN nearly reaches local carrying capacity ($R = \alpha K_0$, with $\alpha = 0.9$). The pine density was estimated based on the percentage of land covered by pine species (data provided by EFI, <http://www.efi.int/>; [56], assuming that 100% cover represents 156000 trees/ km^2 (approximately the highest pine density in maritime pine forests). Given the wide range of hosts for *M. galloprovincialis*, we did not differentiate pine species and considered overall pine density when calculating the carrying capacities and resident population densities [19, 57].

To account for the invasion dynamics of infested populations (N) and the recruitment of resident populations (R), we refine the growth function (2) according to the invasion stage. More precisely, we consider the following four chronological steps, assuming that PWN arrives in year $t = n$ at a given location:

1. Before the PWN arrival ($t < n$), absence of infested vectors:
 $N = 0$ and $K = K_0$
2. PWN introduction at $t = n$, a certain amount of trees are weakened or killed by the PWN, the carrying capacity increases from K_0 to K_i . But at this stage, there is not yet recruitment of resident populations as newly infested beetles will emerge the year after:

$$f(N) = \varepsilon N \left(1 - \frac{N + R}{K_i} \right) \quad (3)$$

3. At $t = n + 1$, offspring of both infested and resident populations carries the PWN when emerging from PWN-infested trees. Thus, we consider the recruitment of resident populations:

$$f(N) = \varepsilon(N + R) \left(1 - \frac{N + R}{K_i} \right) \quad (4)$$

4. In the following years ($t > n+1$), the whole population is infested (no further recruitment):

$$f(N) = \varepsilon N \left(1 - \frac{N}{K_i}\right) \quad (5)$$

With this schematic approach, the aim is to account for the main processes and provide a preliminary analysis of the effects of the Pyrenean chain on spread of infested beetles rather than giving an accurate prediction of the spread rate. Since direct validation was impossible (because PWN is not yet present in the Pyrenees), an indirect validation of the vector spread model was done by comparing simulations to the results of the genetic study.

Description of spread simulations

The study area covers the Pyrenean chain and adjacent sides (41.52–44.22° N, -3.24–3.96° E, $\approx$ 300 km North to South of the ridgeline and $\approx$ 700 km East to West). Simulations were made over a grid of cells of 1 km x 1 km covering this area, using R language [58]. We simulated the potential spread of PWN-infested populations starting with 200 infested beetles at time $t = 0$ from the South-West of the Pyrenees. Simulations were done until all favorable areas of the grid were colonized. Areas were considered colonized as soon as the population density exceeded 1 infested insect per km². To measure the time needed to cross the Pyrenees from South to North in different points of the mountain range, we simulated the spread along the four transects (A, B, C & D). Simulation started from the Spanish extremity of transects (South of the Pyrenees) and we measured the number of generations required to reach the French extremity (North of the Pyrenees).

Effects of temperature and elevation

Temperature affects the developmental rate of *M. galloprovincialis* [59]. Since this insect has its larval development that spans both summer and winter, it is likely to be affected by winter temperatures encountered at high elevations in the Pyrenees. Winter temperatures have been shown to be good indicator for survival and development of pine associated insects with a winter larval development [60, 61]. We considered the mean minimum temperature in winter (December, January and February) over 1950–2000, called TN hereafter (downloaded from <http://www.worldclim.org>, [62]). To determine the temperature threshold for survival, we extracted TN from 353 sites where *M. galloprovincialis* was recorded throughout Europe. We then obtained the TN value below which this species was not or rarely found. As there is probably a gradual effect of low temperatures and large variability on insect survival instead of a sharp and common threshold, we tested various values (0, -2, -5, -7, and -10°C). The highest elevation at which *M. galloprovincialis* can survive and spread is closely linked to temperature. Therefore we also tested the following elevation thresholds: 1200, 1500, 1590, 1700 and 2000 m based on trapping results. In the model simulations, we considered no beetle survival and no PWN transmission below the temperature threshold and above the elevation threshold, respectively.

Pine Wilt Disease expression and host tree decline

PWN transmission to a host tree does not necessarily involve the expression of pine wilt disease (PWD) and tree decline. PWD expression depends on environmental factors and notably summer temperatures and is generally not observed in areas where the mean temperature in July is below 20°C [63]. Therefore, in this case, there is no additional resource for the vector and the

carrying capacity remains the same before and after PWN invasion ($K_i = K_0$). These areas are clearly unfavorable for the natural expansion of the nematode compared with warmer areas where PWN cause outbreaks of vector density. In another set of simulations, we therefore considered that infested PWN populations can spread only where the mean temperature of July is above 20°C. Since temperature is predicted to increase by 2.8–4.0°C in the Pyrenees over the next century [64], the area where they can spread could be modified. To explore the effects of climate change in the near future, we used putative values of +1 and +2°C in the simulations.

Results

Genetic analysis

The number of alleles per locus ranged from 2 to 18, with a mean value of 7.23 over the 26 populations. The average allelic richness of populations over all loci ranged from 2.51 to 3.72 and the average private allelic richness ranged from 0 to 0.21 (Table 1). Both allelic and private allelic richness were significantly higher in populations of the Southern side of the Pyrenean chain. F_{is} calculated over populations indicates a slight excess of homozygotes; values ranged from -0.016 to 0.303 for an average value of 0.167 in the Northern side and 0.102 in the South of the mountain range (Table 1). Significant deviations from Hardy Weinberg Equilibrium after FDR correction were detected for the loci Mon27, Mon30 and Mon35. FreeNA identified significant null alleles frequency (>10%) for Mon01 (92%), Mon27 (96%), Mon30 (50%) and Mon35 (38%). No linkage disequilibrium was detected among the 12 loci. Deviation from mutation drift equilibrium tested through Bottleneck showed that only one population exhibited a significant signature of recent bottleneck (Beriain, heterozygosity excess: $p = 0.032$). Heterozygosity deficit was detected in two populations (Falces and Oloron, heterozygosity deficiency: $p = 0.028$ and 0.038 respectively).

Population genetic structure

Pairwise F_{st} computed under the *ENA* correction of the 26 populations was higher than 0.05 for 86% of population pairs, but differentiation was significant ($p < 0.0001$) only for 28.8% of pairs (S1 Table). We observed a ratio of significant pairwise differentiation that was higher between populations of the two sides of the mountain range (72.6%) than between populations of the same side (3.4%). The differentiation was generally moderate to low, the mean pairwise F_{st} over the whole dataset was 0.0657 (SD: 0.0423). Mean pairwise F_{st} among populations of the same side was 0.0375 (SD: 0.0263), while this distance was 0.0922 (SD: 0.0372) between populations on the two sides. The maximum differentiation ($F_{st} = 0.185$) was observed between Castillonroy (Spain) and Garros (France). Along transects, we did not observe significant pairwise F_{st} between populations of the same side (except between Falces and Beriain, Fig 2). Isolation by distance was significant for the whole dataset ($p = 0.009$), between populations within the Northern side ($p = 0.037$) and within the Southern side ($p = 0.003$).

STRUCTURE identified an optimum of two clusters in the dataset through the Delta-K method ($\Delta K K2 = 1606.68$, $K3 = 10.70$; Fig 1A). We found two distinct genetic clusters with a clear split between populations on each side of the mountain range (Fig 1B). However, some scattered individuals were not assigned to the same cluster within their populations of origin. We did not observe distinct variation in assignment to clusters among the gradients of elevation. Assignment of individuals and populations to one of the clusters was less marked on the Western side of the chain. Analysis using the parameter “No admixture model” gave similar results. Individual assignment (Q values) with and without null alleles did not give strong differences in the patterns observed (mean difference in cluster assignment over all individuals was 2.6%, SD = 3.5%).

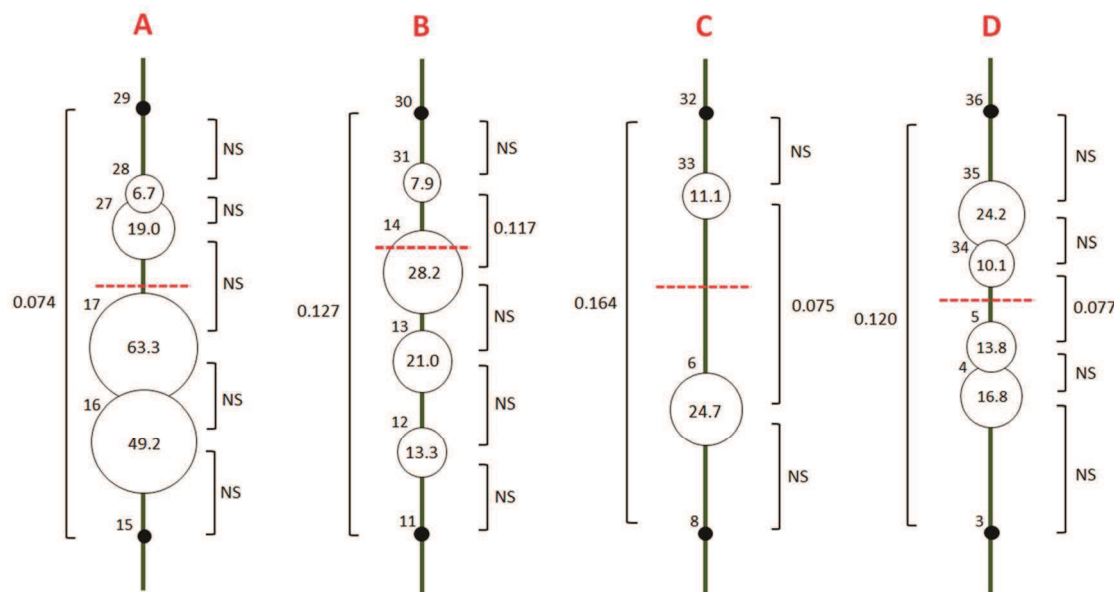


Fig 2. Pairwise F_{st} and likelihood of hybrid populations along the 4 transects (green line). Numbers on the top left of each circle refer to population code (see Fig 1). Values attached to brackets refer to the unbiased estimates of pairwise F_{st} calculated under *ENA* correction (NS = non-significant). The red dotted line shows the placement of the highest ridgeline. Values in the circles and their relative size show the percentage of likelihood to belong to the parent population of the opposite of the ridgeline. Populations used as parents to determine likelihood of hybrid populations are represented as black spots.

doi:10.1371/journal.pone.0134126.g002

Admixture and migration

The rate of admixture of populations obtained with LEA analysis ranged from 6.7% to 63.3% (Fig 2). Multiple iterations gave stable results (average SD over populations = 0.5%, maximal SD = 1.1%). Admixture was higher in populations of the South hillside for transects A, B and C, suggesting an asymmetric migration of individuals, from France to Spain on the Western side of the Pyrenees. Transects A and B showed a negative relation between admixture coefficient and distance to the ridgeline. A total of 40 first generation migrants were detected among the 26 populations tested. Events of migration were mainly observed within each hillside, and particularly between neighbor populations (Fig 3). When considering the two genetic clusters defined with STRUCTURE, migration between clusters was more often observed among populations of the western hillside (Falces, Marcalain, Itziar, Irun, Etchalar). It should be noted that the origin of migrants computed by GeneClass is an estimation of the most likely population of origin which themselves may be poorly differentiated from each other within hillsides. Thus, these results show a picture of general patterns of migration within and between clusters rather than an accurate prediction of origin of migrants. Few first generation migrants were detected between populations separated by high elevations (Prades-Castellbell, Jaca-Font froide) and we did not detect a specific signature of migration along road axes in the highest parts of the mountain ranges.

Resident populations and potential spread of PWN-infested populations

Based on the carrying capacity (K_0) estimated across the study area, we found that *M. galloprovincialis* is expected to be widely distributed over the Pyrenees (Fig 4D). The highest K_0 was found along the Mediterranean and the Atlantic coasts, and on the Spanish side of the

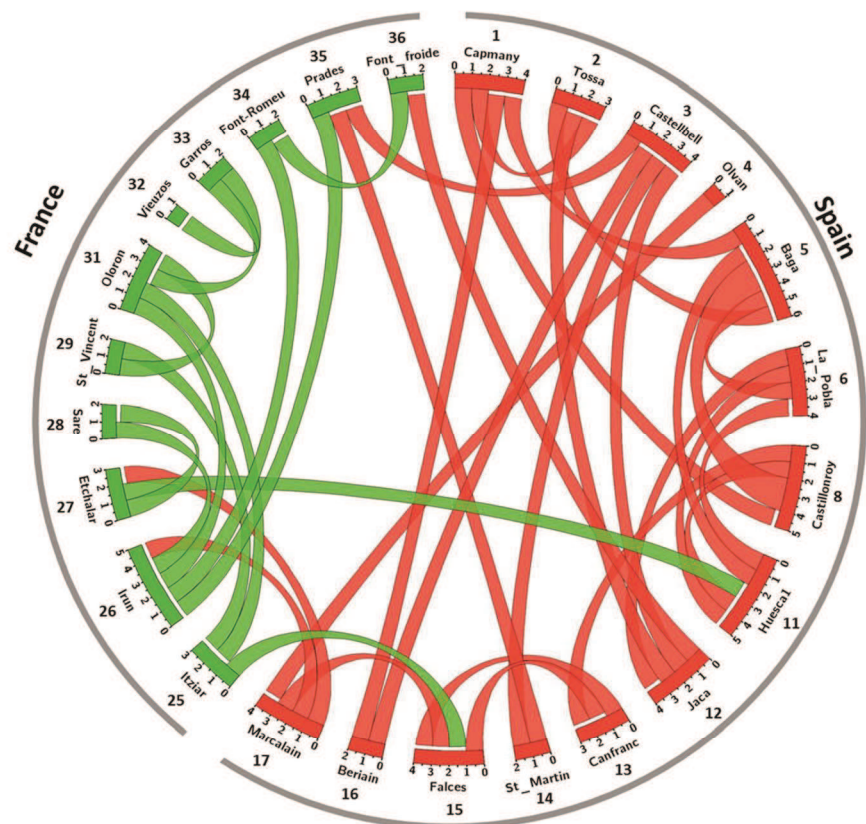


Fig 3. Origin of the 40 first generation migrants detected among the 26 populations tested. Populations are represented on the circle by arcs. Colors refer to the assignments of populations to the two clusters identified by STRUCTURE (red = cluster of the southern side and green = cluster of the northern side of the mountain range). Ribbons represent migrants of first generation. The attached end indicates the population of "arrival" and the detached end to the most likely population of origin of the migrant.

doi:10.1371/journal.pone.0134126.g003

Pyrenees. Conversely, low K_0 was observed in an area located away from the coast on the French Pyrenean side. Based on trapping, we found an average carrying capacity of $K_0 = 0.027$ insect / tree (S2 Table) and a ratio between K_i and K_0 of $K_i \approx 4 K_0$ (See S3 Table for details). Results of simulations are represented by areas colonized by intervals of 10 successive generations (Figs 5 and 6). When considering no thresholds of temperature and elevation, model simulations of the spread of infested populations of *M. galloprovincialis* showed a wave of colonization which expanded over all suitable habitats in Spain and France.

Thresholds of temperature and elevation

The temperature distribution over the 353 sampling sites where *M. galloprovincialis* was recorded in Europe showed that this species can be found over a large range of TN temperatures (S2 Fig). The species was sampled in areas with a TN ranging from -10.0 to $+9.8^\circ\text{C}$; however, 50% of sampling sites were between -2.0 and $+2.0^\circ\text{C}$. Only 6% of the sites were below -5.0°C and 2% below -7.0°C . Thus, the thresholds of -5.0 and -7.0°C for beetle survival tested

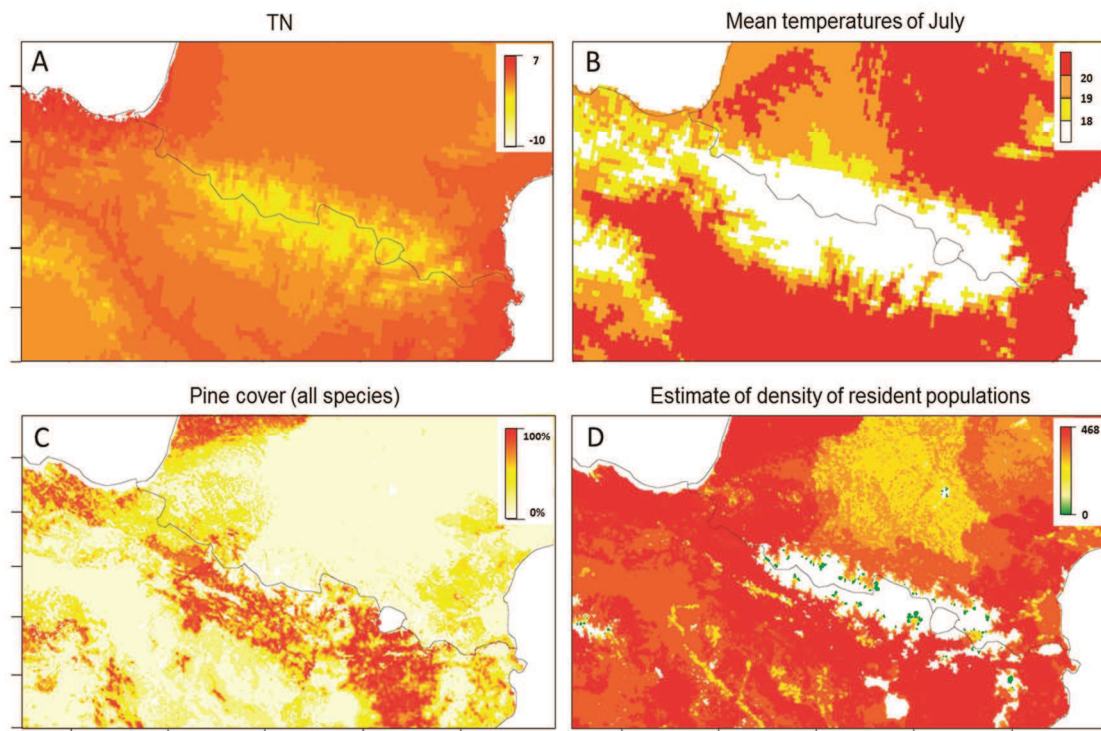


Fig 4. Temperatures, pine cover and resident populations in the Pyrenees. A) Mean of minimum temperature (°C) in winter (December, January and February; called TN) over 1950–2000. B) Mean temperatures (°C) in July over 1950–2000. C) Pine cover (%). D) Projection of the distribution and density of resident populations of *M. galloprovincialis* over the Pyrenean chain. Maximal elevation threshold for the beetle survival set to 1590 m for this representation. Scale refers to the population density (number of individuals per km²).

doi:10.1371/journal.pone.0134126.g004

seem consistent with literature and trapping records, as very few insects were recorded below these temperatures. In addition, the lowest TN where *M. galloprovincialis* was collected was -7°C (Col de la Pierre St Martin, transect B, Table 1). Population density in this site was very low (only 10 specimens collected over the 3 months of trapping, against an average of 131 on the other sites, S2 Table). We thus consider that this value gives a consistent approximation of temperature threshold for beetle survival. Regarding elevation, we recorded *M. galloprovincialis* at a maximal elevation of 1590 m (Col de la Pierre St Martin). Given the low density encountered at this site, we also considered this elevation as the most likely threshold for beetle survival. The threshold of -7°C and 1590 m defined a discontinuous barrier of 1456 km² in length and a continuous barrier of 9156 km² respectively along the Pyrenean chain (Fig 5). These barriers did not stop the simulated spread of infested populations and moderately affected the spread rate, due to the presence of corridors of suitable habitats located on each edge of the Pyrenean chain bounding the Atlantic and Mediterranean coasts.

Over all the simulations, none of the tested thresholds of temperature and elevation were predicted to completely stop the spread of infested vector populations across the Pyrenees. However, these thresholds affected the number of generations (i.e., years) needed to reach the opposite end of each transect (Table 2). The spread rate was not affected by changes in minimal temperature and maximal elevation for beetle survival along the most western transect (A)

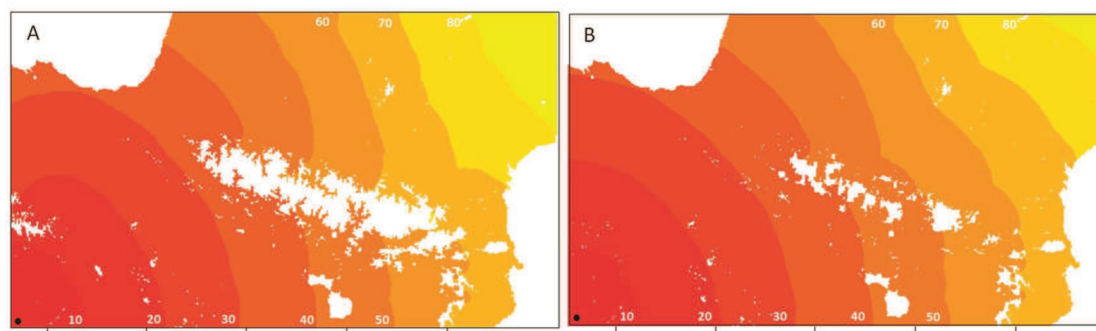


Fig 5. Simulations resulting from the spread model. Figures show the potential spread of infested populations of *M. galloprovincialis* over the Pyrenees based on pine density and limits of survival of this species. A) Simulated spread when considering a maximal elevation of 1590 m for beetle survival. B) Simulated spread when considering a minimal temperature (TN) of -7°C for beetle survival. Simulations started from Spain (black circle) and were run until complete spread over the area. Each color panel represents the area colonized by the PWN in 10 years.

doi:10.1371/journal.pone.0134126.g005

where elevation is lower (maximal elevation of approximately 800 m along the transect A versus 2200, 2800 and 2300 m for transects B, C and D respectively). This dynamic was moderately affected along the most Eastern transect (D) (5% and 7% of variation among the altitudes and temperature thresholds tested). The effect of changes in temperature and elevation was higher along transects B and C, located across the most elevated part of the ridgeline (21 and 19% for altitudes and 50 and 60% for temperatures).

When testing the temperature threshold for PWD expression (20°C), and potential effect of climate warming of +1°C and +2°C, an unfavorable area of 88,281, 58,402 and 42,173 km² were determined over the Pyrenees (Fig 4B). Potential spread of infested populations was very different under warming scenarios (Fig 6). The temperature threshold of 20°C created a wide unfavorable area for spread of PWN on the western site of the Pyrenees, but 1°C and 2°C increases in temperature reduced considerably this area. All the scenarios tested predicted a corridor suitable for PWD expression and spread of infested *M. galloprovincialis* along the Mediterranean hillside of the Pyrenean chain.

Discussion

New phoretic combinations resulting from introductions of alien species have been documented for several insect-pathogen relationships. They involve either an endemic pathogen spread by an introduced insect [65, 66], or conversely an exotic pathogen spread by a native insect [67]. Phoretic associations leading to major environmental disturbances have been well studied [68, 69]. However, few studies have addressed the question of their dispersal in a heterogeneous landscape.

In Europe, *M. galloprovincialis* is already associated with the endemic, non-pathogenic species *Bursaphelenchus mucronatus* [70]. A novel phoretic association with *B. xylophilus* would improve the fitness of *M. galloprovincialis* by providing temporally more declining trees for development. To assess the potential spread of the PWN, both partners of this nematode-beetle association and their synergic effects must be considered. In this study, we combined population genetic and modeling approaches to explore the effects of the Pyrenean chain on the spread of PWN-infested populations of *M. galloprovincialis*. We show that the Pyrenees represents a partial barrier to northward/southward migrations of beetle populations and we identified potential corridors for the natural spread of the PWN. This integrative study provides the first analysis of the potential dispersal of the PWN-vector association at local geographic scale.

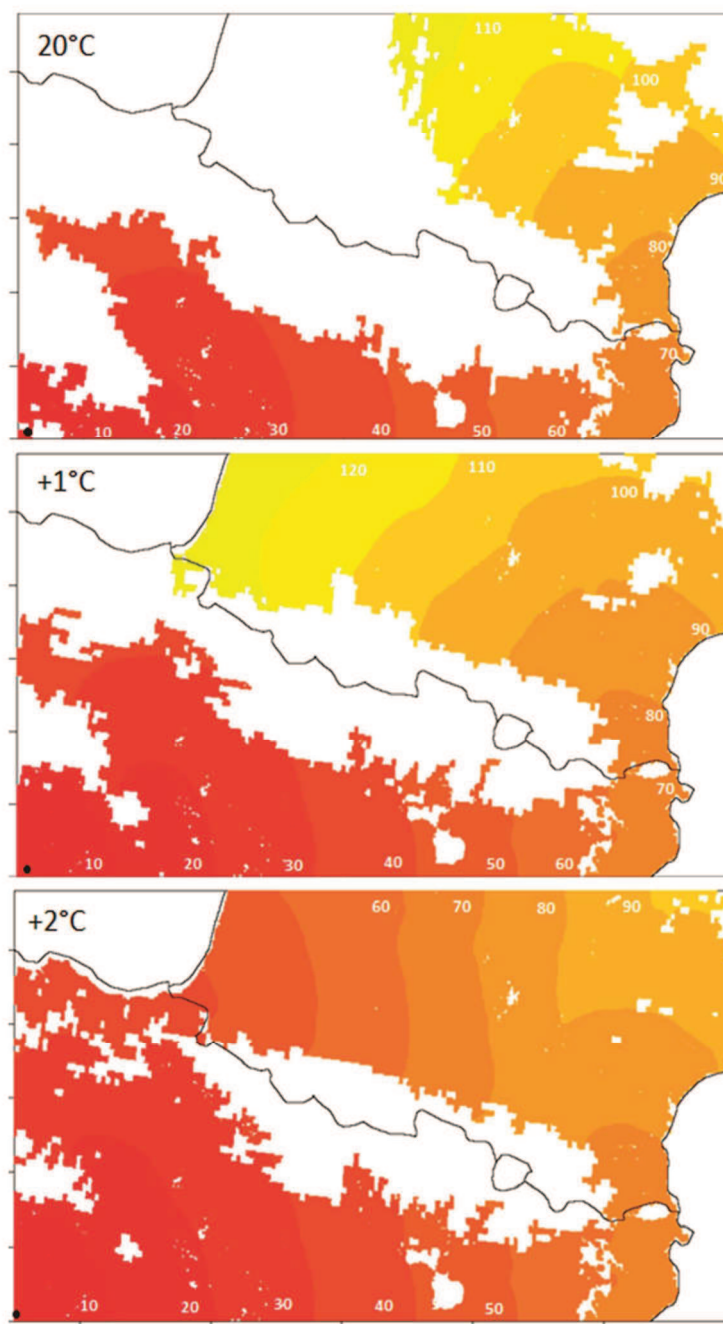


Fig 6. Simulated spread of PWN-infested populations assuming no spread of PWD infested beetles where the disease does not express. Maps represent the population density of infested populations for various temperature thresholds for PWD expression (mean July temperatures of 20°C and + 1°C and +2°C of climate warming). Simulations started from Spain (black circle) and were run until complete spread over the area. Each color panel represents the area colonized by the PWN in 10 years.

doi:10.1371/journal.pone.0134126.g006

Table 2. Simulations of the temporal dynamic of spread of the PWN-infested populations of *M. galloprovincialis* along the four transects (A, B, C & D).

Max. alt. (m)	1200	1500	1580	1700	2000	%
A	23	23	23	23	23	0
B	24	21	20	20	19	21
C	52	44	43	41	36	30.8
D	29	29	28	28	28	3.4
Min. temp. (°C)	0	-2	-5	-7	-10	%
A	23	23	23	23	23	0
B	38	23	20	19	19	50
C	64	51	39	26	25	60
D	29	28	27	27	27	6.9

Values represents the number of generations required to spread from the Spanish extremity to the French extremity of the transects. Max. alt. and Min. temp. represent respectively the maximal elevation and minimum temperatures (TN) for beetle survival tested. Percentages refer to the extent of change in number of generations between the minimum and maximum temperature and altitude tested.

doi:10.1371/journal.pone.0134126.t002

in Europe and highlights priority zones for future pest management programs along the Pyrenean barrier.

Genetic structure and differentiation of populations

Our population genetic analyses provide evidence that the Pyrenean chain has acted and continues to act as a barrier to dispersal of *M. galloprovincialis*. We found population differentiation (pairwise F_{ST}) systematically higher and more significant between populations from each side of the mountain ridge than within hillsides. In addition, the clustering method clearly identified two genetic lineages distinct to the northern and southern sides respectively, forming a zone of separation along the ridgeline of the mountain range. This result is in agreement with the genetic differentiation observed for *M. alternatus* across a mountain range in Japan [71]. From an evolutionary perspective, such a genetic structure suggests the presence of a secondary contact zone of two differentiated lineages of *M. galloprovincialis* along the Pyrenees, consistent with the phylogeographic patterns documented for several native species in Europe including trees [72, 73], vertebrates [74] and insects [75]. In our case, the area studied should be expanded to a wider phylogeographic context to validate this scenario.

The overall genetic differentiation of the populations was low to moderate but consistent with intraspecific population differentiation encountered at local scale for flying beetles [71, 76, 77, 78]. We found a significant isolation by distance correlation in the studied area, which is compatible with results found for *M. alternatus* in Japan [71]. The low genetic differentiation within hillsides suggests intensive gene flow and subsequently rather high dispersal abilities for *M. galloprovincialis*. This corroborates the important dispersal distances estimated in laboratory conditions [22] and in the field [23] for this species. This result is also consistent with observation made on the congeneric species *M. alternatus* for which low population differentiation was found between populations in lowland valleys [71]. Our results tend to show that *M. galloprovincialis* does not encounter strong barriers to dispersal in lowland valleys to the North and South of the Pyrenees. We found no signal of population differentiation within the altitudinal gradients of hillsides. Altitudinal gradients of genetic differentiation have been encountered in the salamander *Ambystoma macrodactylum*, based on microsatellite markers [79] and in the alpine butterfly *Lycaena tityrus*, based on genes under selection [80]. Our observation suggest no clear differentiation between populations in high and low elevation areas, and subsequently

no differentiation between insects feeding on thermophile Mediterranean pines (*P. halepensis*, *P. pinaster*) and the ones feeding on pines encountered at altitude (*P. sylvestris*, *P. uncinata*). These results are consistent with feeding experiments showing that this species may develop on almost all pine species encountered in this zone [57]. This analysis did not provide evidence to suggest that changes of pine species along hillsides of the Pyrenees drive divergence and gene flow in this species. Absence of differentiated lineage in an elevated area can be a sign of recent expansion of the species in altitude. We did not observe recent bottlenecks among populations, indicating that the genetic patterns observed are better explained by long term processes than from recent spatial expansion in these areas.

Genetically differentiated lineages of *M. galloprovincialis* associated with elevation or specific host pines are important to consider as they may show different responses once in contact with the PWN. The absence of structured populations with altitude did not support the existence of the former altitudinal subspecies *M. galloprovincialis pistor* (Germar, 1818) [81] characterized by darker teguments. For the Pyrenean populations analyzed in this study, our results are in agreement with morphological analysis [82] suggesting no evidence of a genetic base for the validity of this former subspecies. As a result, management does not need to take into account multiple lineages for *M. galloprovincialis* in the Pyrenean area.

Corridors and barriers to dispersal of *M. galloprovincialis*

The presence of two genetically differentiated lineages on each hillside enables detection of migration and long-distance dispersal events across the mountain range. We found higher gene flow on the western and eastern sides of the mountain range, and especially along the most western transect (A). This is probably due to the moderate elevation and the presence of a continuous pine tree cover encountered in this area. Bayesian assignment based on individuals also argues for a higher admixture rate among populations of transect A. The western and eastern extremities of the Pyrenean chain are known corridors for lineage expansion of several species during recent post-glacial re-colonization events [83, 84, 85]. Similarly, distribution of mitochondrial haplotypes of the pine processionary moth (*Thaumetopoea pityocampa*) suggests a recolonisation route of this species from Spain to France through the western side of the mountain range [86]. The pine processionary moth develops on a large range of pines species. Thus this indicates that the western side of the mountain range was a continuous corridor of pines at least temporarily over the current interglacial period (Holocene, see Bucci et al. 2007 [87] for distribution of genetic clusters of *P. pinaster* in this area). In the case of *M. galloprovincialis*, however, although migration was detected in this area, our results show the presence of a separation zone, as observed in the grasshopper *Chorthippus parallelus* [75]. The maintaining of this zone despite favorable habitats and gene flow is unexpected and may be related to the effect of isolation by monopolization or by isolation by local adaptation of the two lineages to local environmental conditions [88].

Long-distance dispersal events, caused by human transportation are critical parameters, because they can significantly accelerate the spread of PWN [37]. Such transportations may disperse beetles bypassing unfavorable areas (i.e. high elevations) that would naturally stop the migration. The Pyrenees is the only land connection between the Iberian Peninsula and the rest of Europe, and the road traffic along the border between Spain and France is intense. Around 120,000 vehicles cross the border each day along 27 axes. The majority of traffic (67%) occurs in the coastal zones [89]. The traffic in elevated areas is very low and scattered, except on the Eastern side of the mountain range (transect D), in Andorra and Bourg-Madame [89]. This traffic may contribute significantly to long-distance dispersal of the beetle and we searched for genetic signatures of such events. However, we did not find evidence of long-

distance dispersal along specific road axes in the high mountains. This suggests that human transportation of *M. galloprovincialis* along these axes is absent or not frequent enough to be detected using the methods employed in this study. Conversely, we observed higher migration events in the western and eastern hillsides of the Pyrenees where most of the traffic is concentrated. Estimating the relative importance of human mediated versus natural dispersal as determinants of genetic patterns is challenging. In this study, we found that individuals disperse more from North to South of the ridgeline in the western hillside. This situation contradicts the patterns of northward natural expansion of Spanish populations observed for other species [85, 86, 90] in this area. This case may be related to intensive human mediated dispersal, caused by the transport of a massive amount of pine logs to Spain after windstorms that occurred in Southwestern France [91]. Such hypothesis should be considered with care given the lack of data regarding the temporal scale at which the genetic patterns observed were formed.

Estimation of gene flow assessed in this study provides an overview of effective dispersal of *M. galloprovincialis*. It informs about the distance between the place of birth and reproduction of individuals over generations. This provides a general picture of migration patterns of the beetle in the Pyrenees, but does not necessarily relate to effective dispersal of PWN which can be transmitted without reproduction of the beetle through maturation feeding. The modeling approach allowed us to assess potential expansion of the PWN including an estimate of effective dispersal of the nematode.

Estimated density of resident populations

The resident population density derived from the pine tree density suggests that *M. galloprovincialis* is present in all areas within the mountain range, at least at low population density. This widespread presence indicates that suitable host plants for *M. galloprovincialis* are universally available in this area and that habitat connectivity is important. Such conclusions are supported by trapping observations, showing that this species is present in almost all sites sampled including the ones with very low pine density, and where trees are present as very scattered pine stands (Oloron, Vieuzos). In addition, over the study area, all pines species were found to host *M. galloprovincialis*. We even trapped specimens in an isolated plot of *P. pinea* in Spain (Huesca 1) which is described as unsuitable for its larval development [57]. These specimens may either be locally adapted to this host, or may come from distant location and other pine species, pulled into the pheromone traps. Thus, according to our observations and simulations and the widespread distribution of suitable pine tree host species, the distribution of the vector beetle is not a limiting factor to PWN spread in most of the studied area.

Potential spread of PWN across the Pyrenean chain

The model used in this study to simulate the potential spread of infested populations of *M. galloprovincialis* was refined from a previous model of PWN spread [37, 38]. Our model took into account the chronological steps in recruitment of resident populations and the increase in population density due to tree mortality in PWN-infested zones. This allowed us to consider more accurately interactions between the two species in the simulations of dispersal of PWN infested populations. From sensitivity analysis of the model, D (diffusion coefficient) appears to have the greatest effect on results compared to the growth rate (E) and the carrying capacity (K_0). Simulated spread supports the genetic structure and migration patterns observed for this species. Thresholds of temperature and elevation tested always defined at least a moderate barrier to dispersal at high elevation within the mountain range but identified corridors of dispersal on the Eastern and Western sides of the ridge line. Thus, according to these results based on

biological and physical parameters of survival of *M. galloprovincialis*, the Pyrenean chain constitutes only a partial barrier for the spread of PWN-infested populations.

When considering the temperature threshold for PWD expression (mean July temperatures above 20°C), high elevations and the western side of the Pyrenean chain formed potential wider barriers to the spread of PWN-infested populations. This result suggests a higher potential of temperatures threshold for PWD expression to slow the spread of PWN across the mountain range compared with thresholds of beetle survival. This is in agreement with results found in China, where simulations predicted an effect of elevation on the spread of PWD [37, 92], but in contrast to the very high mountains in China, the Pyrenees (lower elevation) would have a weaker effect in case of climate warming. To confirm these results, there is a need to refine precisely this threshold in Europe and to quantify the spread of the PWN in areas where the PWD does not express. In addition, it should be noted that temperatures used in this study are based on records from recent decades (1950–2000) and that given the current climate change, the scenarios considering temperature increase of +1 and +2°C are probably more realistic. Current development of models for PWD expression in Europe will help to refine these predictions [93].

Future climate change and particularly temperature rise may affect the spread of the PWN by providing new areas suitable for PWD development. In this study, a moderate temperature rise of +1 and +2°C reduced considerably the potentially unsuitable area for PWD expression and connected suitable area on the western side of the mountain range. This indicates that even a moderate climate change may affect significantly the range expansion of PWD by reducing the extent of potential barriers as pointed by Roques et al. 2014 [92]. Temperature increase may also widen the range of habitat of *M. galloprovincialis* in elevated areas. Due to uncertainties on the temperatures of survival threshold of this species, it was not possible to formally test the effect of climate change on spread of infested beetles based only on beetle survival.

Our simulations are based on existing knowledge on *M. galloprovincialis*, the PWN and their interactions. However, several parameters such as PWN transmission efficiency, the role of a sister species found in elevated areas in PWN transmission (*M. sutor*), the dispersal behavior of infested beetles and environmental factors affecting PWD expression are still lacking accurate data. Future studies will have to address these questions to refine the spread model and better predict invasion routes of this pest in Europe. Long-distance jumps resulting from human transportation is an important parameter to be considered in the potential spread of PWN. Although spread models taking into account human mediated dispersal have been applied to Europe, long-distance dispersal of PWN in Europe is still poorly understood and lacks observed data [38]. Lastly, early detection and eradication of new outbreaks of PWN-infested populations may strongly affect PWN spread. Clear-cut belts may also create artificial barriers to the spread of this pest. These control measures and the efficiency of their application are obviously critical parameters to consider in the potential range expansion of the PWN in Europe.

Conclusion

This study aimed at evaluating the effect of an altitudinal barrier on the natural spread of PWN infested populations of *M. galloprovincialis*. Based on gene flow estimates of the beetle and on the simulations resulting from a refined spread model, we showed that the Pyrenean chain constitutes a partial barrier to the natural spread of this alien from Spain to the rest of Europe. Our results suggest that the western and eastern parts of this mountain range will not constitute effective barriers to prevent the spread from Spain to France, although the western side appears less favorable to PWD expression under current climate conditions. These areas should be

considered as priority zones for pest monitoring and management programs. Barriers to dispersal of *M. galloprovincialis* and spread of the PWN highlighted here are potentially relevant to others mountainous systems in Europe (Alps, Carpathians) exhibiting similar biological and physiological constraints.

Supporting Information

S1 Fig. Location of sampling sites and road axes in the Pyrenees.
(EPS)

S2 Fig. Distribution of the mean of minimal temperature in winter over 1950–2000 (TN) for 353 sampling sites of *M. galloprovincialis* in Europe.
(EPS)

S1 Table. Unbiased pairwise F_{st} estimates of the 26 populations considered.
(XLSX)

S2 Table. Estimation of the carrying capacity (K). The number of insects per tree was evaluated based on density of individuals trapped and the number of trees in the area surrounding the trap. For this calculation, we consider only localities for which traps were set during the same duration (4 months). Traps were set in dense and large pine plots approximating a cover of 100%. Assuming that 100% pine cover represents a number of 156000 trees/km², we found a number of 4898 trees in a surface of radius of 100 m around the sampled sites (estimated area of attraction of the traps). We then report the number of individuals trapped to the number of trees estimated in the surface influenced by the traps. We found an average value of 0.027 insect per tree. Sensitivity analysis (see appendix 3) shows that variation of K has very low impact on model results. Subsequently, uncertainties on estimation of this parameter are not likely to affect strongly model results.
(XLSX)

S3 Table. Estimating the ratio between K_0 and K_i . Traps (Multifunnel and specific attractant volatile) were set in PWN free and PWN-infested pine stands (*Pinus pinaster*) in Portugal and Spain during July 2013. We counted specimens caught during 2 weeks of trapping. We estimated the population density as on average four times higher in PWN-infested areas than in PWN free areas ($K_i \approx 4K_0$).
(XLSX)

S4 Table. Sensitivity analysis. The model sensitivity to the main parameters (D , ε and K_i) was assessed by varying the parameters' value by +/- 10% from the baseline values previously given. The effect of these changes was quantified in terms of infested area (km²) and percentage of change compared to the baseline settings. Simulations were done from $t = 1$ to 20 years, and from 3 different starting points (P1: 41.68° N, -2.94° E; P2: 41.68° N, 0.81° E and P3: 41.68° N, 2.73° E) to assess a possible effect of the starting point on parameter's sensitivity. Among the 3 parameters tested, variations in diffusion coefficient (D) involved the largest changes in area colonized. Others parameters were affected by less than 2%.
(XLSX)

S5 Table. Microsatellite dataset.
(TXT)

Acknowledgments

We would like to thank Jacques Garcia for trapping survey assistance and Juan Alberto Pajares Alonso for providing valuable samples from Spain. We thank Jelena Jovic and two anonymous reviewers for their constructive comments on early versions of this article. We also thank Hugh Evans for English checking and helpful comments.

Author Contributions

Conceived and designed the experiments: JH AR AB CR GR. Performed the experiments: JH AB. Analyzed the data: JH CR GR. Contributed reagents/materials/analysis tools: JH AB CR GR. Wrote the paper: JH AR CR GR.

References

1. Roques A, Kenis M, Lees D, Lopez-Vaamonde C, Rabitsch W, Rasplus J-Y, et al. Alien terrestrial arthropods of Europe. *BioRisk*. 2010; 4: 1–1028.
2. Santini A, Ghelardini L, De Pace C, Desprez-Loustau ML, Capretti P, Chandelier A, et al. Biogeographical patterns and determinants of invasion by forest pathogens in Europe. *New Phytol*. 2013; 197: 238–250. doi: [10.1111/j.1469-8137.2012.04364.x](https://doi.org/10.1111/j.1469-8137.2012.04364.x) PMID: [23057437](https://pubmed.ncbi.nlm.nih.gov/23057437/)
3. Eschen R, Roques A, Santini A. Taxonomic dissimilarity in patterns of interception and establishment of alien arthropods, nematodes and pathogens affecting woody plants in Europe. *Divers Distrib*. 2014; 1–10.
4. Blackburn TM, Essl F, Evans T, Hulme PE, Jeschke JM, Kuhn I, et al. A Unified Classification of Alien Species Based on the Magnitude of their Environmental Impacts. *PLoS Biol*. 2014; 12(5). doi: [10.1371/journal.pbio.1001850](https://doi.org/10.1371/journal.pbio.1001850)
5. Evans HF, Mc Namara DG, Braasch H, Chadoeuf J, Magnusson C. Pest risk analysis (PRA) for the territories of the European Union (as PRA area) on *Bursaphelenchus xylophilus* and its vectors in the genus *Monochamus*. *Bull OEPP*. 26. 1996; 199–249.
6. Wiens J. Ecological heterogeneity: an ontogeny of concepts and approaches. In: *The Ecological Consequences of Environmental Heterogeneity* (eds Hutchings M, John E, Stewart A). Blackwell Science Ltd, Oxford. 2000; 9–31.
7. Zeller KA, McGarigal K, Whiteley AR. Estimating landscape resistance to movement: a review. *Landscape Ecol*. 2012; 27: 777–797.
8. Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol*. 1998; 7: 453–464. PMID: [9628000](https://pubmed.ncbi.nlm.nih.gov/9628000/)
9. Hewill GM. Post-glacial re-colonization of European biota. *Biol J Linn Soc Lond*. 1999; 68: 87–112.
10. Schmitt T. Biogeographical and evolutionary importance of the European high mountain systems. *Front Zool*. 2009; 6: 9. doi: [10.1186/1742-9994-6-9](https://doi.org/10.1186/1742-9994-6-9) PMID: [19480666](https://pubmed.ncbi.nlm.nih.gov/19480666/)
11. Schmitt T. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Front Zool*. 2007; 4: 11. PMID: [17439649](https://pubmed.ncbi.nlm.nih.gov/17439649/)
12. Kuroda K. Physiological incidences related to symptom development and wilting mechanism. In: Zhao BG, Futai K, Sutherland JR, and Takeuchi Y (eds) *Pine wilt disease*, Springer Tokyo. 2008; 204–222.
13. Bergdahl DR. Impact of the pine wood nematode in North America: Present and future. *J Nematol*. 1988; 20: 206–265.
14. Mamiya Y. History of Pine Wilt Disease in Japan. *J Nematol*. 1988; 20: 219–226. PMID: [19290205](https://pubmed.ncbi.nlm.nih.gov/19290205/)
15. Mota M, Braasch H, Bravo M.A, Penas AC, Burgermeister W, Metge K, et al. First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. *Nematology*. 1999; 1: 727–734.
16. Robertson L, Cobacho AS, Escuer M, Santiago Merino R, Esparrago G, Abellera A, et al. Incidence of the pinewood nematode *Bursaphelenchus xylophilus* Steiner & Buhner, 1934 (Nickle, 1970) in Spain. *Nematology*. 2011; 13: 755–757.
17. Linit MJ. Nematode-vector relationships in the pine wilt system. *J Nematol*. 1988; 20: 227–235.
18. Sousa E, Bravo MA, Pires J, Naves P, Penas AC. *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae) associated with *Monochamus galloprovincialis* (Coleoptera: Cerambycidae) in Portugal. *Nematology*. 2001; 3: 89–91.
19. Hellrigl KG. La bionomie des espèces de *Monochamus* (Coleoptera, Cerambycidae) et leur importance pour la sylviculture et l'économie du bois. *Redia*. 1971; 52: 367–511.

20. Naves PM, Camacho S, de Sousa E, Quartau JA. Transmission of the pine wood nematode *Bursaphelenchus xylophilus* through oviposition activity of *Monochamus galloprovincialis* (Coleoptera: Cerambycidae). *Entomol Fenn*. 2007; 18: 193–198.
21. Schröder T, McNamara D, Gaar V. Guidance on sampling to detect pine wood nematode *Bursaphelenchus xylophilus* in trees, wood and insects. *Bulletin OEPP/EPPO Bulletin*. 2009; 39: 179–188.
22. David G, Giffard B, Piou D, Jactel H. Dispersal capacity of *Monochamus galloprovincialis*, the European vector of the pine wood nematode, on flight mills. *J Appl Entomol*. 2013; 138: 566–576.
23. Mas H, Hernandez R, Villaroya G, Sanchez G, Pérez-Laorga E, Gonzalez E, et al. Dispersal behavior and long distance flight capacity of *Monochamus galloprovincialis* (Olivier 1795). In: Schröder T (ed.), *Pine Wilt Disease Conference*. pp. 22, Braunschweig, ISSN: 1866-590X. 2013.
24. Yamamura K, Mitsunobu K, Norio A, Futochi K, Yasutsune S. Estimation of dispersal distance by mark-recapture experiments using traps: correction of bias caused by the artificial removal by traps. *Popul Ecol*. 2003; 45: 149–155.
25. Selkoe KA, Toonen RJ. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol Lett*. 2006; 9: 615–629. PMID: [16643306](#)
26. Chikhi L, Bruford MW, Beaumont MA. Estimation of Admixture Proportions: A Likelihood-Based Approach Using Markov Chain Monte Carlo. *Genetics*. 2001; 158: 1347–1362. PMID: [11454781](#)
27. Cornuet JM, Piry S, Gordon L, Estoup A, Solignac M. New Methods Employing Multilocus Genotypes to Select or Exclude Populations as Origins of Individuals. *Genetics*. 1999; 153: 1989–2000. PMID: [10581301](#)
28. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155: 945–959. PMID: [10835412](#)
29. Broquet T, Petit EJ. Molecular Estimation of Dispersal for Ecology and Population Genetics. *Annu Rev Ecol Syst*. 2009; 40: 193–216.
30. Shigesada N, Kawasaki K. Biological invasions: theory and practice. In *Oxford series in ecology and evolution*, Oxford University Press, Oxford; 1997.
31. Pitt JPW, Worner SP, Suarez AV. Predicting Argentine and spread over the heterogeneous landscape using a spatially explicit stochastic model. *Ecol Appl*. 2009; 19: 1176–1186. PMID: [19688925](#)
32. Carrasco LR, Mumford JD, MacLeod A, Harwood T, Grabenweger G, Leach AW, et al. Unveiling human-assisted dispersal mechanisms in invasive alien insects: Integration of spatial stochastic simulation and phenology models. *Ecol Modell*. 2010; 221: 2068–2075.
33. Robinet C, Kehlenbeck H, Kriticos DJ, Baker RHA, Battisti A, Brunel S, et al. A suite of models to support quantitative assessment of spread in pest risk analysis. *PLoS ONE*. 2012; 7(10).
34. Yoshimura A, Kawasaki K, Takasu F, Togashi K, Futai K, Shigesada N. Modeling the spread of pine wilt disease caused by nematodes with pine sawyers as vector. *Ecology*. 1999; 80: 1691–1702.
35. Takasu F, Yamamoto N, Kawasaki K, Togashi K, Kishi Y, Shigesada N. Modeling the expansion of an introduced tree disease. *Biol Invasions*. 2000; 2: 141–150.
36. Togashi K, Shigesada N. Spread of the pinewood nematode vectored by the Japanese pine sawyer: modeling and analytical approaches. *Popul Ecol*. 2006; 48: 271–283.
37. Robinet C, Roques A, Pan HY, Fang GF, Ye JR, Zhang YZ, et al. Role of Human-Mediated Dispersal in the Spread of the Pinewood Nematode in China. *PLoS ONE*. 2009; 4(2). doi: [10.1371/journal.pone.0004646](#)
38. Robinet C, Van Opstal N, Baker R, Roques A. Applying a spread model to identify the entry points from which the pine wood nematode, the vector of pine wilt disease, would spread most rapidly across Europe. *Biol Invasions*. 2011; 13: 2981–2995.
39. EPPO Reporting Service No. 02, Paris 2014-02-01. Available: <http://archives.eppo.int/EPPOReporting/2014/Rse-1402.pdf>.
40. Soliman T, Mourits MCM, van der Werf W, Hengeveld GM, Robinet C, Lansink AGJMO. Framework for Modeling Economic Impacts of Invasive Species, Applied to Pine Wood Nematode in Europe. *PLoS ONE*. 2012; 7(9). doi: [10.1371/journal.pone.0045505](#)
41. Haran J, Roux-Morabito G. Development of 12 microsatellites loci for the longhorn beetle *Monochamus galloprovincialis* (Coleoptera Cerambycidae), vector of the Pine Wood Nematode in Europe. *Conserv Genet Resour*. 2014; 6: 975–977.
42. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*. 2012; 28: 2537–2539. PMID: [22820204](#)
43. Kalinowski S. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes*. 2005; 5: 187–189.

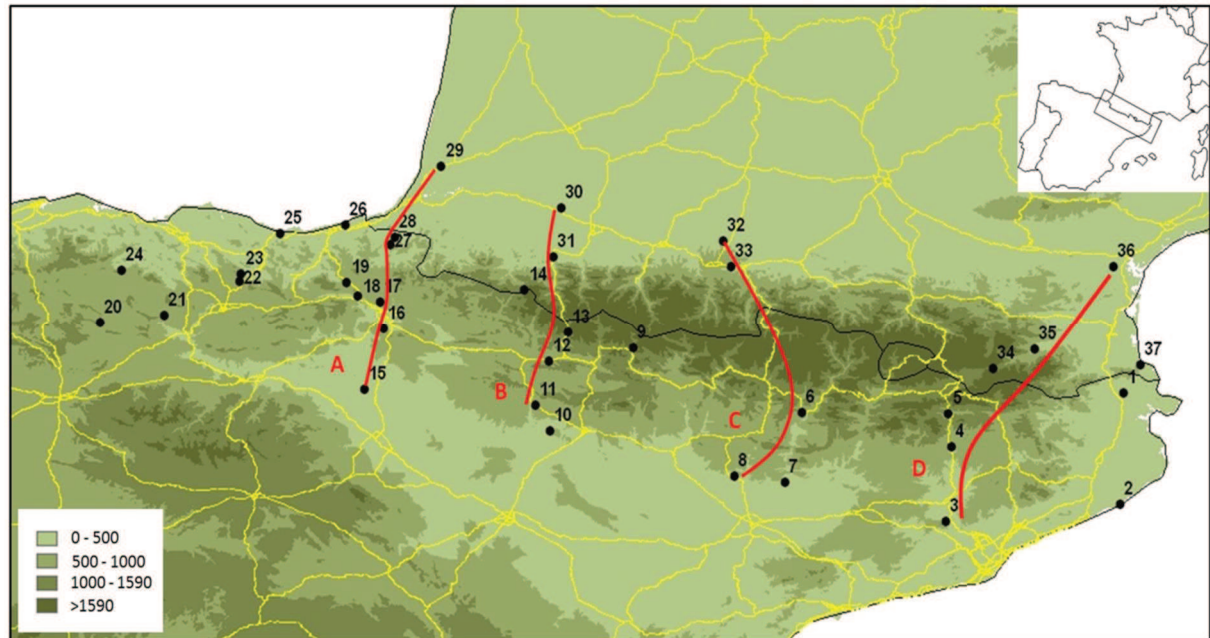
44. Raymond M, Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Heredity*. 1995; 86: 248–249.
45. Storey JD. A direct approach to false discovery rates. *J R Stat Soc B*. 2002; 64: 479–498.
46. Chapuis MP, Estoup A. Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol*. 2007; 24: 621–631. PMID: [17150975](#)
47. Piry S, Luikart G, Cornuet JM. BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *J Hered*. 1999; 90: 502–503.
48. Goudet J. FSTAT (Version 1.2): A computer program to calculate F-statistics. *J Hered*. 1995; 86: 485–486.
49. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol*. 2005; 14: 2611–2620. PMID: [15969739](#)
50. Earl DA, Von Holdt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*. 2012; 4: 359–361.
51. Langella O, Chicki L, Beaumont MA. LEA (likelihood-based estimation of admixture): a program to estimate simultaneously admixture and time since the admixture event. *Mol Ecol Notes*. 2001; 1: 357–358.
52. Choisy M, Franck P, Cornuet JM. Estimating admixture proportions with microsatellites: comparison of methods based on simulated data. *Mol Ecol*. 2004; 13: 955–968. PMID: [15012768](#)
53. Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudoin L, Estoup A. GENECLASS2: A software for genetic assignment and first-generation migrant detection. *J Hered*. 2004; 95: 536–539. PMID: [15475402](#)
54. Rannala B, Mountain JL. Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci USA*. 1997; 94: 9197–9201. PMID: [9256459](#)
55. Paetkau D, Slade R, Burden M, Estoup A. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Mol Ecol*. 2004; 13: 55–65. PMID: [14653788](#)
56. Tröltzsch K, Van Brusselen J, Schuck A. Spatial occurrence of the major tree species group in Europe derived from multiple data sources. *For Ecol Manage*. 2009; 257: 294–302.
57. Naves P, Sousa E, Quartau J. Feeding and oviposition preferences of *Monochamus galloprovincialis* for some conifers under laboratory conditions. *Entomol Exp Appl*. 2006; 120: 99–104.
58. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-90005107-0. 2009. URL <http://www.R-project.org>.
59. Naves P, de Sousa E. Threshold temperatures and degree-day estimates for development of post-dormancy larvae of *Monochamus galloprovincialis* (Coleoptera: Cerambycidae). *J Pest Sci*. 2009; 82: 1–6.
60. Battisti A, Stastny M, Netherer S, Robinet C, Schopf A, Roques A, et al. Expansion of geographic range in the pine processionary moth caused by increased winter temperatures. *Ecol Appl*. 2005; 15: 2084–2096.
61. Robinet C, Baier P, Pennerstorfer J, Schopf A, Roques A. Modelling the effects of climate change on the potential feeding activity of *Thaumetopoea pityocampa* (Den. & Schiff.) (Lep., Notodontidae) in France. *Glob Ecol Biogeogr*. 2007; 16: 460–471.
62. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Clim*. 2005; 25: 1965–1978.
63. Rutherford TA, Webster JM. Distribution of pine wilt disease with respect to temperature in North America, Japan and Europe. *Can J For Res*. 1987; 17: 1050–1059.
64. López-Moreno JI, Goyette S, Beniston M. Climate change prediction over complex areas: spatial variability of uncertainties and predictions over the Pyrenees from a set of regional climate models. *Int J Climatol*. 2008; 28: 1535–1550.
65. Lu M, Wingfield MJ, Gillette NE, Mori SR, Sun JH. Complex interactions among host pines and fungi vectored by an invasive bark beetle. *New Phytol*. 2010; 187: 859–866. doi: [10.1111/j.1469-8137.2010.03316.x](#) PMID: [20546136](#)
66. Luchi N, Mancini V, Feducci M, Santini A, Capretti P. *Leptoglossus occidentalis* and *Diplodia pinea*: a new insect-fungus association in Mediterranean forests. *For Pathol*. 2011. doi: [10.1111/j.1439-0329.2011.00750.x](#)
67. Brasier CM. Recent genetic changes in the *Ophiostoma ulmi* population: the threat to the future of the elm. In *Populations of plant pathogens: Their dynamics and Genetics*. Edited by Wolfe MS, Caten CE. Oxford: Blackwell Publ. 1987; 213–226.
68. Zhao BG. Pine wilt disease in China. In: Zhao B.G., Futai K., Sutherland J.R., and Takeuchi Y. (eds) *Pine wilt disease*, Springer Tokyo. 2008; 18–25.

69. Potter C, Harwood T, Knight J, Tomlinson I. Learning from history, predicting the future: the UK Dutch elm disease outbreak in relation to contemporary tree disease threats. *Philos Trans R Soc Lond B Biol Sci*. 2011; 366: 1966–1974. doi: [10.1098/rstb.2010.0395](https://doi.org/10.1098/rstb.2010.0395) PMID: [21624917](https://pubmed.ncbi.nlm.nih.gov/21624917/)
70. Vincent B, Koutroumpa F, Altemayer V, Roux-Morabito G, Gevar J, Martin C, et al. Occurrence of *Bursaphelenchus mucronatus* (Nematoda: Aphelenchoididae) in France and association with *Monochamus galloprovincialis* (Coleoptera: Cerambycidae). *Ann For Sci*. 2008; 111. doi: [10.1051/forest:2007083](https://doi.org/10.1051/forest:2007083)
71. Shoda-Kagaya E. Genetic differentiation of the pine wilt disease vector *Monochamus alternatus* (Coleoptera: Cerambycidae) over a mountain range—revealed from microsatellite DNA markers. *Bull Entomol Res*. 2007; 97: 167–174. doi: [10.1017/S000748530700483X](https://doi.org/10.1017/S000748530700483X) PMID: [17411479](https://pubmed.ncbi.nlm.nih.gov/17411479/)
72. Desmasure B, Comps B, Petit RJ. Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L) in Europe. *Evolution*. 1996; 50: 2515–2520. doi: [10.2307/2410719](https://doi.org/10.2307/2410719)
73. King RA, Ferris C. Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Mol Ecol*. 1998; 7: 1151–1161.
74. Wallis GP, Arntzen JW. Mitochondrial-DNA variation in the crested newt superspecies—Limits cytoplasmic gene flow among species. *Evolution*. 1989; 43: 88–104. doi: [10.2307/2409166](https://doi.org/10.2307/2409166)
75. Cooper SJB, Hewitt GM. Nuclear DNA sequence divergence between parapatric subspecies of the grasshopper *Chorthippus parallelus*. *Insect Mol Biol*. 1993; 4: 49–60.
76. Coates BS, Sumerford DV, Miller NJ, Kim KS, Sappington TW. Comparative Performance of Single Nucleotide Polymorphism and Microsatellite Markers for Population Genetic Analysis. *J Hered*. 2009; 100: 556–564. doi: [10.1093/jhered/esp028](https://doi.org/10.1093/jhered/esp028) PMID: [19525239](https://pubmed.ncbi.nlm.nih.gov/19525239/)
77. Choi SK, Kim KS, Lee H, Adamczyk JJ, Greenberg SM. Temporal Changes in Genetic Variation of Boll Weevil (Coleoptera: Curculionidae) Populations, and Implications for Population Assignment in Eradication Zones. *Ann Entomol Soc Am*. 2011; 104: 816–825. doi: [10.1603/AN11012](https://doi.org/10.1603/AN11012)
78. Schrey NM, Schrey AW, Heist EJ, Reeve JD. Genetic heterogeneity in a cyclical forest pest, the southern pine beetle, *Dendroctonus frontalis*, is differentiated into east and west groups in the southeastern United States. *J Insect Sci*. 2011; 11. doi: [10.1673/031.011.11001](https://doi.org/10.1673/031.011.11001)
79. Giordano AR, Ridenhour BJ, Storfer A. The influence of altitude and topography on genetic structure in the long-toed salamander (*Ambystoma macrodactylum*). *Mol Ecol*. 2007; 16: 1625–1637. doi: [10.1111/j.1365-294X.2006.03223.x](https://doi.org/10.1111/j.1365-294X.2006.03223.x) PMID: [17402978](https://pubmed.ncbi.nlm.nih.gov/17402978/)
80. Karl I, Fischer K. Altitudinal and environmental variation in lifespan in the Copper butterfly *Lycaena tityrus*. *Funct Ecol*. 2009; 23: 1132–1138. doi: [10.1111/j.1365-2435.2009.01607.x](https://doi.org/10.1111/j.1365-2435.2009.01607.x)
81. Koutroumpa FA, Rougon D, Bertheau C, Lieutier F, Roux-Morabito G. Evolutionary relationships within European *Monochamus* (Coleoptera: Cerambycidae) highlight the role of altitude in species delineation. *Biol J Linn Soc Lond*. 2013; 109: 354–376.
82. Wallin H, Schroeder M, Kvamme T. A review of the European species of *Monochamus* Dejean, 1821 (Coleoptera, Cerambycidae)—with a description of the genitalia characters. *Nor J Entomol*. 2013; 60: 11–38.
83. Taberlet P, Fumagalli L, Hausser J. Chromosomal versus mitochondrial-DNA evolution—Tracking the evolutionary history of the southwestern European populations of the *Sorex-araneus* group (Mammalia, Insectivora). *Evolution*. 1994; 48: 623–636. doi: [10.2307/2410474](https://doi.org/10.2307/2410474)
84. Santucci F, Emerson BC, Hewitt GM. Mitochondrial DNA phylogeography of European hedgehogs. *Mol Ecol*. 1998; 7: 1163–1172. PMID: [9734073](https://pubmed.ncbi.nlm.nih.gov/9734073/)
85. Miguel I, Iriondo M, Garnery L, Sheppard WS, Estomba A. Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and post-glacial re-colonization routes in the western Europe. *Apidologie*. 2007; 38: 141–155. doi: [10.1051/apido:2007007](https://doi.org/10.1051/apido:2007007)
86. Rousselet J, Zhao RX, Argal D, Simonato M, Battisti A, Roques A, et al. The role of topography in structuring the demographic history of the pine processionary moth, *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae). *J Biogeogr*. 2010; 37: 1478–1490. doi: [10.1111/j.1365-2699.2010.02289.x](https://doi.org/10.1111/j.1365-2699.2010.02289.x)
87. Bucci G, Gonzalez-Martinez SC, Le Provost G, Plomion C, Ribeiro MM. Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Mol Ecol*. 2007; 16: 2137–2153. doi: [10.1111/j.1365-294X.2007.03275.x](https://doi.org/10.1111/j.1365-294X.2007.03275.x) PMID: [17498237](https://pubmed.ncbi.nlm.nih.gov/17498237/)
88. Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Mol Ecol*. 2013; 22: 5983–5999. doi: [10.1111/mec.12561](https://doi.org/10.1111/mec.12561) PMID: [24128305](https://pubmed.ncbi.nlm.nih.gov/24128305/)
89. DREAL Midi-Pyrénées et Secrétaria General de Transportes. Observatoire franco-espagnol des trafics dans les Pyrénées—supplément au Document n° 6; 2013.
90. Joger U, Fritz U, Guicking D, Kalyabina-Hauf S, Nagy ZT. Phylogeography of western Palearctic reptiles—Spatial and temporal speciation patterns. *Zool Anz*. 2007; 246: 293–313.

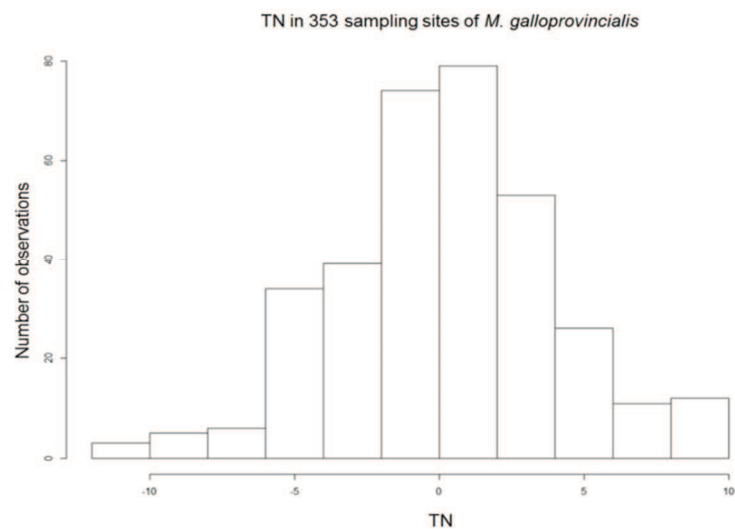
91. López S, Goldarazena A. Flight dynamics and abundance of *Ips sexdentatus* (Coleoptera: Curculionidae: Scolytinae) in different sawmills from northern Spain: Differences between local *Pinus radiata* (Pinales: Pinaceae) and southern France incoming *P. pinaster* timber. *Psyche*. 2012; 145930: 1–6.
92. Roques A, Zhao L, Sun J, Robinet C. Pine wood nematode, pine wilt disease, vector beetle and pine tree: how a multiplayer system could reply to climate change? In: Björkman C. & Niemelä P. (ed.), *Climate Change and Insect Pests*. CABI Editions; 2014. in press.
93. Gruffudd H, Evans HF, Jenkins T. Using an evapo-transpiration model to predict the current and future range and severity of pine wilt disease caused by pine wood nematode, *Bursaphelenchus xylophilus* in Europe. In: Schröder T.(ed.), *Pine Wilt Disease Conference*. 82–83, Braunschweig, ISSN: 1866-590X. 2013.

Supporting Information

S1 Figure: Location of sampling sites and road axes in the Pyrenees.



S2 Figure: Distribution of the mean of minimal temperature in winter over 1950–2000 (TN) for 353 sampling sites of *M. galloprovincialis* in Europe.



S1 Table: Unbiased pairwise F_{st} estimates of the 26 populations considered. F_{st} estimates calculated under ENA correction. (NS = non-significant).

	Code	1	2	3	4	5	6	8	11	12	13	14	15	16	17	25	26	27	28	29	30	31	32	33	34	35	36
Capmany	1		NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS	*	*	*	*	*	*	*	*	*
Tossa	2	0.0070		NS	NS	NS	NS	NS	NS	NS	NS	*	*	NS	NS	NS	*	NS	*	*	*	*	*	*	*	*	*
Castellbell	3	0.0165	0.0264		NS	NS	NS	NS	NS	NS	NS	NS	*	NS	NS	NS	*	NS	*	*	*	*	*	*	*	*	*
Olvan	4	0.0508	0.0492	0.0249		NS	NS	NS	NS	NS	NS	NS	*	NS	NS	NS	*	NS	*	*	*	*	*	*	*	*	*
Baga	5	0.0293	0.0388	0.0022	0.0297		NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	*	NS	*	NS	*	*	*	*	*	*	NS
La Pobla	6	0.0219	0.0273	0.0053	0.0301	-0.0059		NS	NS	NS	NS	NS	*	NS	NS	NS	*	NS	*	*	*	*	*	*	*	*	NS
Castillonroy	8	0.0130	0.0233	0.0124	0.0351	0.0159	0.0211		NS	NS	NS	NS	NS	NS	NS	NS	*	NS	*	*	*	NS	*	*	*	*	*
Huesca 2	11	0.0035	0.0287	0.0050	0.0424	0.0013	0.0111	0.0080		NS	NS	NS	*	*	NS	NS	*	*	*	*	*	*	*	*	*	*	*
Jaca	12	0.0146	0.0302	0.0109	0.0500	0.0200	0.0223	0.0163	-0.0065		NS	NS	*	NS	NS	NS	*	*	*	*	*	*	*	*	*	*	*
Canfranc	13	0.0069	0.0349	0.0104	0.0559	0.0272	0.0101	0.0315	0.0103	0.0093		NS	NS	NS	NS	NS	*	NS	*	*	*	*	*	*	*	*	*
St Martin	14	0.0247	0.0562	0.0228	0.0426	-0.0007	-0.0049	0.0373	-0.0003	0.0233	0.0172		NS	NS	NS	NS	*	NS	*	*	*	*	*	*	*	*	*
Falces	15	0.0372	0.0614	0.0413	0.0675	0.0391	0.0339	0.0461	0.0293	0.0488	0.0251	0.0372		NS	NS	NS	NS	NS	*	*	*	*	*	*	*	*	*
Beriain	16	0.0600	0.0621	0.0368	0.0787	0.0441	0.0413	0.0918	0.0472	0.0446	0.0366	0.0459	0.0539		NS	NS	NS	NS	*	*	*	NS	*	NS	*	NS	NS
Marcalain	17	0.0630	0.0720	0.0193	0.0629	0.0213	0.0068	0.0735	0.0372	0.0541	0.0351	0.0228	0.0395	0.0024		NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Itziar	25	0.0947	0.1036	0.0542	0.0758	0.0463	0.0331	0.1007	0.0675	0.0890	0.0631	0.0468	0.0662	0.0760	0.0154		NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Irun	26	0.1194	0.1412	0.0678	0.1076	0.0743	0.0633	0.1411	0.1089	0.1187	0.0914	0.0946	0.0962	0.0670	0.0045	0.0360		NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
Etchalar	27	0.1200	0.1168	0.0412	0.0601	0.0436	0.0346	0.1211	0.0707	0.0907	0.0769	0.0609	0.0858	0.0651	0.0023	0.0343	0.0300		NS	NS	NS	NS	NS	NS	NS	NS	NS
Sare	28	0.1294	0.1236	0.0768	0.0968	0.0833	0.0631	0.1576	0.1167	0.1315	0.1038	0.0885	0.0978	0.0520	0.0133	0.0447	0.0226	0.0212		NS	NS	NS	NS	NS	NS	NS	NS
St Vincent	29	0.1233	0.1259	0.0744	0.0862	0.0746	0.0509	0.1375	0.1081	0.1193	0.0940	0.0855	0.0741	0.0521	0.0060	0.0347	0.0094	0.0255	-0.0077		NS	NS	NS	NS	NS	NS	NS
Orthez	30	0.1495	0.1352	0.0911	0.0974	0.0872	0.0672	0.1695	0.1271	0.1505	0.1268	0.0982	0.0999	0.0953	0.0429	0.0553	0.0520	0.0251	-0.0022	0.0087		NS	NS	NS	NS	NS	NS
Oloron	31	0.1328	0.1434	0.0769	0.0911	0.0898	0.0787	0.1538	0.1180	0.1354	0.1048	0.1173	0.0914	0.0917	0.0389	0.0571	0.0265	0.0264	0.0159	0.0096	0.0258		NS	NS	NS	NS	NS
Vieuzos	32	0.1487	0.1415	0.0892	0.0927	0.0984	0.0767	0.1640	0.1329	0.1383	0.1205	0.1093	0.1127	0.0846	0.0213	0.0537	0.0250	0.0319	0.0032	0.0042	0.0269	0.0181		NS	*	*	NS
Garros	33	0.1558	0.1470	0.0998	0.1138	0.1014	0.0757	0.1853	0.1360	0.1488	0.1203	0.1123	0.1027	0.0806	0.0466	0.0558	0.0407	0.0359	-0.0008	-0.0019	0.0015	0.0246	0.0303		NS	NS	NS
FontRomeu	34	0.0998	0.1013	0.0853	0.0835	0.0772	0.0675	0.1343	0.0918	0.0923	0.0936	0.0683	0.0820	0.0538	0.0458	0.0766	0.0798	0.0835	0.0350	0.0439	0.0556	0.0621	0.0529	0.0562		NS	NS
Prades	35	0.0758	0.0824	0.0467	0.0714	0.0443	0.0322	0.0970	0.0683	0.0768	0.0600	0.0498	0.0509	0.0212	0.0003	0.0405	0.0286	0.0332	0.0073	0.0041	0.0146	0.0316	0.0323	0.0184	0.0107		NS
FontFroide	36	0.1156	0.1201	0.0674	0.0829	0.0693	0.0525	0.1265	0.0890	0.0991	0.0797	0.0716	0.0691	0.0444	0.0006	0.0173	0.0175	0.0132	0.0107	-0.0020	0.0348	0.0170	0.0134	0.0223	0.0516	0.0169	

S2 Table: Estimation of the carrying capacity (K). Nb. refers to the number of individuals trapped.

Location	Nb.	Insect / tree
Olvan	38	0.007
Baga	19	0.003
La Pobla de S.	704	0.143
Camarassa	38	0.007
Huesca 1	258	0.052
Huesca 2	20	0.004
Jaca	192	0.039
Canfranc	247	0.050
Col St-Martin	10	0.002
Orthez	142	0.029
Oloron	16	0.003
Garros	115	0.023
Font-Romeu	20	0.004
Prades	19	0.004
Average	131.285	0.026

S3 Table: Estimating the ratio between K_0 and K_i . Nb. refers to the number of individuals trapped, Long. and Lat. refer to the geographic coordinates

Location	Country	PNW present	Nb.	Long.	Lat.
Ourense	Spain	No	23	-7.54	42.01
Redondelo	Portugal	No	21	-7.55	41.71
Castro Daire	Portugal	Yes	77	-7.92	40.92
Coimbra	Portugal	Yes	91	-8.52	40.08
Tondela	Portugal	Yes	72	-8.09	40.49

S4 Table: Sensitivity analysis. “Area colonized” represents the total area infested by infested-beetle populations using baseline values for the parameters at time $t = 20$ generations. Sites P1, P2 and P3 are the starting points used for the simulations. Values represent the percentage of change in the extent of colonized area (km^2).

Starting point		P1	P2	P3
Area colonized (Km^2)		12556	23750	13089
D (6.48 km^2/year)	-10%	-11.23%	-7.31%	-8.01%
	+10%	9.82%	6.95%	7.59%
E (2.17)	-10%	-0.12%	-0.22%	-0.12%
	+10%	0.12%	0.05%	0.19%
K_0(0.3 insect/tree)	-10%	-1.55%	-1.02	-0.95%
	+10%	1.16%	0.61%	0.78%

Chapitre V

Discussion générale & perspectives

Chapitre V : Discussion générale et perspectives

L'objectif général de cette thèse était d'estimer le potentiel de dispersion d'une espèce invasive impliquée dans une association phorétique nouvellement formée avec une espèce native. Pour cela, nous nous sommes focalisés sur le cas de l'invasion du PWN en Europe et de son association avec *M. galloprovincialis*. Les études menées ont permis (i) de caractériser la structure génétique et l'histoire évolutive de l'insecte vecteur en Europe ; (ii) de mettre en évidence les barrières à la dispersion de ce vecteur et enfin, (iii) d'estimer le potentiel dispersif de populations de vecteurs infestés par le nématode. Dans ce dernier chapitre, les résultats obtenus au travers de ces différentes études seront discutés de manière synthétique.

V.1 – Potentiel dispersif du nématode à travers son association avec *M. galloprovincialis*.

V.1.1- Phylogéographie de *M. galloprovincialis* : une espèce structurée par les glaciations et les facteurs climatiques

Une structure génétique affectée par les glaciations

Au cours de la reconstruction phylogéographique de *M. galloprovincialis* (article n°1), les génotypes microsatellites de 26 populations européennes ont été analysés afin d'en évaluer la structure génétique à large échelle. Le regroupement des populations en 8 clusters a mis en évidence l'existence de 5 lignées principales en Europe continentale. Ces lignées se distribuent globalement de la manière suivante : Portugal/Galice ; Espagne ; ouest et nord de la France ; Italie et sud-est de la France ; Balkans /Europe centrale/Scandinavie, et se regroupent en deux lignées majeures (Péninsule Ibérique ; Europe centrale et Europe de l'est). Ces lignées continentales forment des zones de contact au niveau des principaux reliefs d'Europe (Pyrénées, Alpes) et au niveau de la frontière Allemagne – France

pour les deux lignées rencontrées dans les zones les plus septentrionales de l'aire de distribution (nord de la France et Balkans). Ce type de patron génétique est très fréquemment rencontré chez les organismes des régions tempérées d'Europe et correspond à la signature de contractions et d'expansions d'aires de distribution sous l'effet des oscillations glaciaires du quaternaire (Taberlet 1998, Hewitt 2000, Schmitt et al. 2007). Les seules zones favorables au maintien de l'espèce durant les maxima glaciaires sont appelées refuges glaciaires. Les populations de ces zones géographiques conservent en général une diversité génétique importante, résultant du nombre élevé de mutations qui peut apparaître au niveau intraspécifique sur de longues périodes de temps. À l'inverse, les populations des zones recolonisées montrent en général une diversité génétique faible, en raison des multiples goulots d'étranglement qui se produisent lors d'un processus d'expansion, et qui conduisent à une perte d'allèles ou d'haplotypes (Hewitt 1999). Nous avons observé les richesses alléliques les plus élevées au sein de plusieurs populations du pourtour méditerranéen et du centre de l'Espagne. Ces zones correspondent à des refuges glaciaires présumés pour les essences de pins sur lesquels cette espèce se développe (Bucci et al. 2007, Fady 2012). Il est donc fortement probable que *M. galloprovincialis* ait subsisté aux derniers maxima glaciaires au niveau de ces refuges, à partir desquels cette espèce a pu recoloniser l'Europe, et former sa distribution actuelle.

Une origine principalement climatique de la distribution des lignées intraspécifiques

Nous nous sommes intéressés à l'origine de la distribution de ces lignées génétiques car, chez *M. galloprovincialis*, cette distribution peut être conditionnée par une préférence pour une essence hôte en particulier, au-delà du simple fait de l'histoire évolutive et de la recolonisation postglaciaire. En effet, les tests de choix d'hôtes réalisés en conditions contrôlées ont montré qu'il existait une préférence de cette espèce pour certaines essences pour la prise alimentaire et pour l'oviposition (Naves et al. 2006). Ce point est important à explorer dans le contexte de dispersion du PWN car l'existence d'une telle préférence pourrait fortement affecter le comportement de dispersion de cette espèce. Les résultats de l'analyse de redondance ont montré que les paramètres environnementaux (températures, précipitations et essence hôte) expliquent une large proportion de la part explicable de la variabilité génétique des populations. Ces résultats ont également montré que les moyennes

annuelles de précipitations constituent le paramètre qui explique le mieux la distribution de cette variabilité, devant l'effet de l'essence hôte. La distribution des lignées de *M. galloprovincialis* est donc fortement influencée par l'environnement, en particulier les précipitations. En revanche, l'essence de pins ne semble pas jouer un rôle prépondérant, et le faible signal observé pourrait être simplement dû à un effet de covariance avec d'autres variables car le climat explique également la distribution des essences de pins. Les tests de préférence ne reflètent peut-être qu'un choix fait par cette espèce si plusieurs essences sont disponibles en milieu contrôlé. Il est donc fort probable que, dans le milieu naturel, cette espèce s'accommode de l'essence dont elle dispose parmi celles qui lui sont favorables (*P. halepensis*, *P. pinaster*, *P. nigra*, *P. sylvestris*). Ce résultat suggère que la dispersion de *M. galloprovincialis* est peu susceptible d'être affectée par les variations de distribution des différentes essences de pins hôtes.

Un fort effet des précipitations peut paraître surprenant pour une espèce phytophage, et il est possible que ce paramètre soit lié à un autre facteur dont la nature n'est pas apparente et qui conditionne réellement la dispersion. Cependant plusieurs études tendent à montrer que les précipitations peuvent jouer un rôle important chez les espèces du genre *Monochamus*. Il a, par exemple, été démontré que l'humidité relative affecte la longévité chez *M. alternatus* (Kong et al. 2006), qui est significativement plus longue à 90% qu'à 40% d'humidité relative pour une température constante de 25 °C. Dans une étude sur *Monochamus leuconotus* (Pascoe), attaquant les plantations de café, la modélisation de la distribution de l'espèce a montré que les variables liées aux précipitations sont les plus importantes (devant les températures) pour expliquer cette distribution (Kutywayo et al. 2013).

L'effet le plus visible des précipitations dans la distribution des lignées est observé au niveau des zones de contact entre les clusters ibériques est et ouest, et entre l'Espagne et la France au niveau du littoral atlantique (résumées sur la Figure 8), zones pour lesquelles il n'existe pourtant pas de barrières physiques majeures à la dispersion, tel qu'un relief très important ou une absence de pins-hôtes. L'isolement par l'environnement (Isolation By Environment, IBE) est un concept qui propose que les individus dispersent préférentiellement au sein des conditions auxquelles ils ont pu localement s'adapter (Wang & Bradburd 2014). L'IBE est très fréquent dans les études de patrons de flux de

gènes (Sexton et al. 2014), et représente une hypothèse intéressante pour expliquer le maintien des zones de contact en l'absence de barrière autre qu'une variation de régime de précipitations.

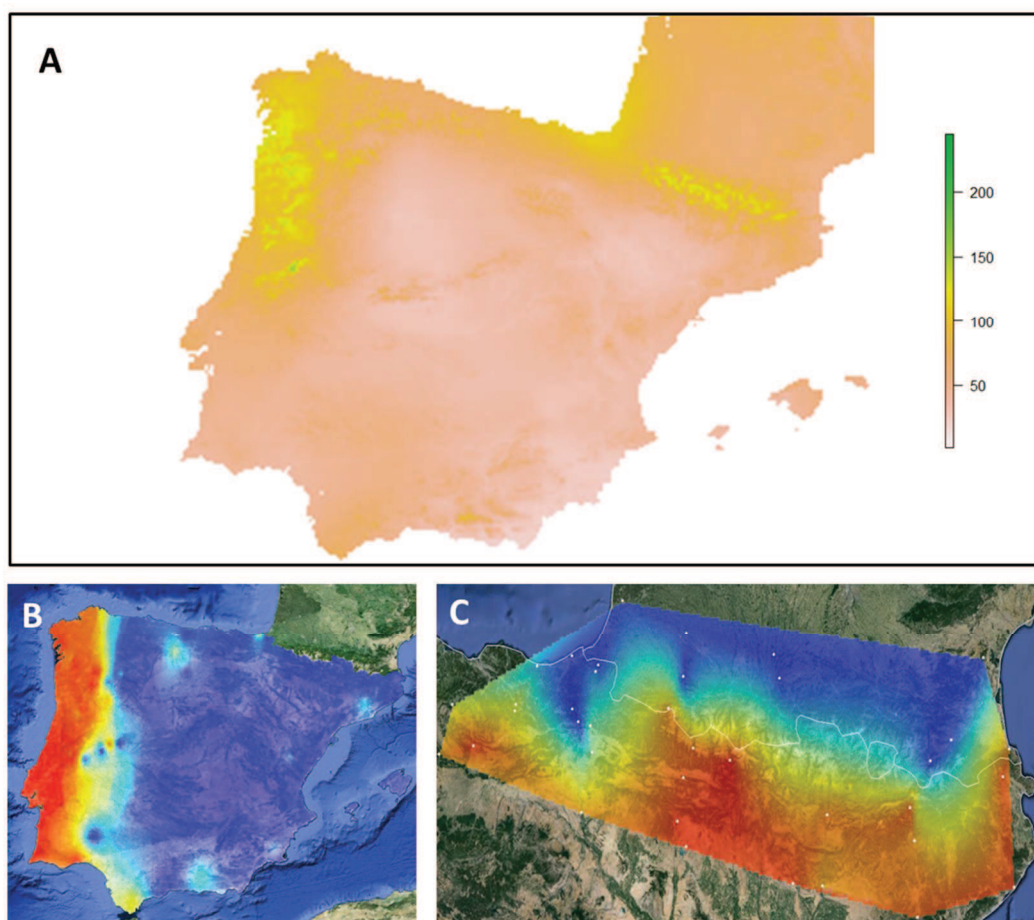


Figure 8: Précipitations et distribution des lignées génétiques de *M. galloprovincialis* identifiées par regroupement bayésien. **A :** Précipitations moyennes annuelles (moyenne des moyennes mensuelles ; en millimètres) dans la péninsule ibérique et dans le sud-ouest de la France. **B :** Interpolation de distribution des deux lignées génétiques en péninsule ibérique (Rouge : lignée ibérique ouest, Bleu : lignée ibérique est). **C :** Même interpolation de distribution au niveau de la barrière pyrénéenne (Rouge : lignée ibérique, Bleu : lignée française).

Actuellement, la dispersion du PWN est principalement observée au sein du cluster ibérique de l'ouest (Portugal/Galice), malgré une incursion en Espagne dans la province de Salamanque (EPPO 2014). Il est difficile de déterminer si cette distribution résulte ou non de l'IBE chez *M. galloprovincialis* et d'une dispersion réduite de ce vecteur entre deux zones différenciées climatiquement. En effet, la distribution actuelle du PWN résulte de sa dispersion naturelle réalisée par le vecteur, mais également par les transports humains. Il faut également noter que cette distribution peut venir de variations dans l'expression de la PWD. En effet, de manière générale, les facteurs climatiques sont connus pour affecter l'expansion spatiale des espèces invasives (Zenni & Hoban 2015), et les précipitations, par exemple, déterminent fortement le degré de développement de la PWD (Braasch & Enzian 2004) et le risque d'observer d'importants dommages (Lawson & Sathyapala 2008). Une dispersion limitée de *M. galloprovincialis* entre des environnements contrastés pourrait représenter une barrière ou tout au moins un frein à la dispersion du PWN. Il est en revanche peu probable que la différenciation génétique observée entre les lignées du vecteur constitue une réelle barrière à la dispersion, puisque le PWN s'est révélé capable de s'associer à plusieurs espèces vectrices au sein du genre *Monochamus* (*M. carolinensis*, *M. alternatus*, *M. galloprovincialis*).

Clarification des contours de l'espèce

L'étude phylogéographique a également permis de préciser les contours taxonomiques de *M. galloprovincialis*. En particulier cette étude n'a pas confirmé l'existence de la sous-espèce *M. galloprovincialis pistos* (Germar, 1818). Cette forme mélanique est observée dans les zones d'altitude et de latitude élevées bénéficiant de conditions climatiques plutôt fraîches (Koutroumpa et al. 2013). La structure génétique obtenue avec nos marqueurs microsatellites montre que les individus de cette forme collectés en altitude ne divergent pas génétiquement des populations voisines rencontrées en plaine, et ce, en plusieurs sites de l'aire de distribution de cette espèce (Suède, Alpes, Pyrénées). Il n'existe donc a priori pas de base génétique pour cette forme, telle qu'elle a été décrite dans la littérature. Ces résultats font écho aux analyses morphologiques, qui indiquent qu'aucun caractère stable ne permet de distinguer cette sous-espèce de la forme typique *M. galloprovincialis*.

galloprovincialis (Koutroumpa et al. 2013, Wallin et al. 2013). Cette forme semble donc être simplement un variant chromatique dont l'existence pourrait être liée à l'effet des conditions d'altitude sur la mélanisation de la cuticule. Il faut d'ailleurs noter que le degré de mélanisation de *M. galloprovincialis* a déjà naturellement tendance à augmenter au cours de la vie de l'adulte (G. David com. Pers.).

L'existence d'hybrides interspécifiques entre *M. galloprovincialis* et *M. sutor* pourrait être envisagée car ces espèces vivent en sympatrie, partagent certaines essences hôtes et sont proches génétiquement (Césari et al. 2005, Koutroumpa et al. 2013). Cette hypothèse pourrait également expliquer la difficulté d'identifier des caractères morphologiques stables pour différencier ces deux espèces, en particulier chez certains individus atypiques (Koutroumpa et al. 2013). Au cours des tests de cross priming des 12 loci microsatellites (Annexe n°2), un certain nombre de loci spécifiques à *M. galloprovincialis* ont été amplifiés avec succès chez *M. sutor*, mettant en évidence des allèles spécifiques à chacune des deux espèces. A travers les différentes études de cette thèse, aucun génotype hybride n'a été observé. Etant donné le nombre de spécimens impliqués (près de 2200 au total) et la couverture géographique de l'ensemble cumulé des études – incluant plusieurs zones de sympatrie entre les espèces – on peut conclure que ces deux espèces ne s'hybrident a priori pas en conditions naturelles. Ce résultat conforte d'ailleurs les hypothèses d'incompatibilité de reproduction issues de l'étude morphologique des pièces génitales (Koutroumpa et al. 2013). Ces résultats montrent donc que *M. galloprovincialis* est une espèce aux contours bien délimités, sans introgression visible avec son espèce sœur, et ne présentant pas de lignée différenciée associée aux régions fraîches engagée dans un processus avancé de spéciation.

Une structure génétique favorable pour détecter les événements de dispersion à longue distance

Comme cela a été évoqué en introduction, l'identification des voies potentielles d'invasion d'un pathogène impliqué dans une APNF est facilitée par la mise en évidence des voies et modes de dispersion (naturelle ou non) du vecteur avant l'invasion. En particulier, il est essentiel de mettre en

évidence l'existence de déplacements à longue distance du vecteur (résultant de transport accidentels par l'homme) car de tels sauts peuvent contribuer significativement à l'expansion spatiale du phoronte. Il existe un important potentiel de dispersion du PWN en Europe par transport longue distance de *M. galloprovincialis* porteur de ce nématode. En effet, les larves et jeunes adultes de ce coléoptère peuvent être transportés dans le bois de pins dont les rondins et les produits dérivés (planches, palettes) sont intensivement échangés. Par exemple, un adulte a récemment été intercepté dans un bois de palette à destination de la Nouvelle-Calédonie (Mille et al. 2015). La longévité importante des adultes (2-3 mois) favorise également leur survie durant ces sauts à longue distance. Enfin, faut noter que de tels transports ont été montrés en proportion importante chez l'espèce *M. alternatus* en Chine (Hu et al. 2013). En Europe, la dispersion longue distance de *M. galloprovincialis* reste très peu connue, en partie en raison du caractère imprévisible de ces événements.

La reconstruction de la phylogéographie de *M. galloprovincialis* a permis d'identifier plusieurs lignées génétiques différenciées et structurées spatialement en Europe. Cette distribution fournit une base favorable pour tester l'existence d'événements de dispersion à longue distance par transport accidentel d'origine anthropique. En effet, les individus peuvent facilement être assignés à un des clusters par tests d'assignation, pour tracer leur origine. Ceci est particulièrement intéressant dans les cas où des individus seraient interceptés dans des zones sensibles au sein de marchandises difficiles à tracer (Mille et al. 2015). De tels événements on été recherchés dans deux zones « sensibles » que représentent les ports et les scieries (Irun et Tarascon, Article n°1), mais aucun phénomène d'introduction d'individus provenant de localité distante n'a pu être mis en évidence. En revanche, l'assignation a permis d'établir que les populations de *M. galloprovincialis* introduites à Madère sont originaires de la zone où s'étend le regroupement génétique ouest ibérique (Portugal et Galice). *M. galloprovincialis* a été mentionné pour la première fois en 1992 sur cette ile, et une origine espagnole de cette introduction avait été initialement suggérée, sans localité précise (Erber & Franquinho Aguiar 1996).

Les zones de contact entre clusters constituent des zones particulièrement adaptées à l'identification de ces transports à plus petite échelle, car la différenciation génétique rend l'identification des migrants plus aisée. L'étude des patrons de migration au niveau des Pyrénées

(article n°3) a permis de mettre en évidence le nombre et la localisation de tels événements entre versants, malgré certaines limites de la méthode employée (discutés dans l'article n°3).

V.1.2- Les barrières et les corridors à la dispersion naturelle de *M. galloprovincialis*.

Des flux de gènes importants à faible échelle

Plusieurs analyses réalisées au cours cette thèse ont mis en évidence un faible niveau de structuration génétique des populations de *M. galloprovincialis* à échelle locale. Ceci a été montré au niveau des Pyrénées (Article n°3), où les populations d'un même versant ne montrent globalement pas de différenciation génétique significative dans une aire de l'ordre de la centaine de km. Une différenciation à ces échelles aurait pu être attendue, car il existe de fortes variations environnementales entre les altitudes basses et élevées de chaque versant. Cette observation est confirmée par le variogramme réalisé sur les agrégats de la péninsule ibérique (Article n°2), qui montre une forte autocorrélation spatiale des génotypes (absence d'indépendance) pour des distances inférieures 190 km.

Ce faible niveau de différenciation suggère un important flux de gènes entre populations et l'absence de barrières à la dispersion à ces échelles dans les zones sans barrière géographique majeure. Cette observation indique également qu'il existe des taux et des distances de dispersion relativement élevés chez *M. galloprovincialis*. Des résultats similaires ont été obtenus chez l'espèce vectrice asiatique (*M. alternatus*), pour laquelle une absence de différenciation a été montrée entre populations issues de localités distantes de moins de 100 km (Kawai et al. 2006, Shoda-Kagaya 2007). Ces résultats sont à relier aux estimations de distance de dispersion réalisées pour *M. galloprovincialis* en milieu naturel et en conditions contrôlées. David et al. (2014) ont par exemple montré que cette espèce parcourait en moyenne 16 km sur sa durée de vie, en manège de vol, avec des maxima de 45 km pour les femelles et 63 km pour les mâles. Les études de capture-marquage-recapture ont montré que les

distances parcourues par certains individus pouvaient atteindre 13 km, avec des distances journalières allant jusqu'à 1500 m (Mas et al. 2013).

Les espèces du genre *Monochamus* vectrices du nématode du pin et en particulier *M. galloprovincialis* semblent donc être très peu structurées génétiquement à des échelles spatiales faibles (<100 km). Cette conclusion rejoint celle de plusieurs études portant sur d'autres coléoptères voiliers inféodés aux pins (*Dendroctonus frontalis*, *D. ponderosae*, Schrey et al. 2008, Janes et al. 2015). Par exemple, aucune structure génétique n'est observée chez *D. frontalis* aux échelles inférieures à 500 km, et il a été proposé que les individus de cette espèce forment une seule population panmictique à ces échelles (Schrey et al. 2008). Cette absence de structure semble donc être un trait également partagé par plusieurs espèces de coléoptères xylophages. Il peut être objecté que le faible niveau de structure génétique que nous avons observé soit dû à un manque de variabilité au sein des marqueurs microsatellites utilisés et non à l'intensité des flux de gènes chez l'espèce étudiée. Cependant, on notera qu'une absence de structure a également été observée à faible échelle chez *D. ponderosae* à partir de marqueurs nucléaires de type SNP (Single Nucleotide Polymorphism, Janes et al. 2015), ces marqueurs étant généralement reconnus pour être plus résolutifs.

L'absence de structure génétique à faible échelle chez *M. galloprovincialis* est intéressante en soit car elle signifie que les flux de gènes – et donc la dispersion – sont importants à cette échelle, ce qui est a priori favorable à l'expansion naturelle du nématode du pin. En revanche cette absence de structure rend difficile la recherche de phénomènes de dispersion en relation avec les structures fines du paysage, et cela aura des répercussions sur le dispositif expérimental des études recherchant un effet à petite échelle des paramètres environnementaux sur la dispersion (cf. section V.1.2).

Les zones d'altitude représentent des barrières à la dispersion de *M. galloprovincialis*

L'effet de l'altitude et des conditions associées sur la dispersion de *M. galloprovincialis* a été mis en évidence au niveau des Pyrénées (article n°3). La rupture spatiale dans la distribution des clusters génétiques, la différenciation génétique et la différence de niveau d'admixture entre les populations des deux versants sont des arguments en faveur d'un effet de barrière de ce relief sur les

flux de gènes. Ces résultats sont en accord avec les observations faites chez *M. alternatus* (Shoda-Kagaya 2007), qui ont montré une différenciation génétique significative des populations de part et d'autre de la chaîne Ohu (Tohoku district), alors que les populations au sein de chaque versant n'ont pas montré de telle différenciation. Ces premiers résultats restent relativement descriptifs et il est difficile d'en extraire le paramètre qui conditionne réellement cette rupture dans les flux de gènes. En dehors de l'effet simple de la topographie, la dispersion peut être affectée par les températures froides ou les variations de densités en pins associées aux zones d'altitude. La dispersion peut également être affectée – comme cela a été évoqué précédemment – par le phénomène d'isolement par adaptation (IBE), en particulier au niveau des Pyrénées qui correspondent à une zone de contact entre lignées évolutives (Chapitre II et IV).

L'analyse de génétique spatialisée (article n°2) a également montré un effet négatif des reliefs sur la dispersion de *M. galloprovincialis* dans une grande partie de la péninsule ibérique, ce qui permet de généraliser les conclusions obtenues au niveau des Pyrénées seules. Cet effet des reliefs est donc une constante chez cette espèce et ne correspond pas au seul effet de l'adaptation locale de deux lignées évolutives en contact (IBE), comme cela aurait pu être observé au niveau des Pyrénées. L'analyse de génétique spatialisée a également permis de montrer que l'absence de pin n'est pas le facteur qui explique cet effet des reliefs. En effet, si cette situation est rencontrée aux altitudes les plus importantes des Pyrénées, les pins sont en revanche continuellement distribués sur une grande partie des reliefs de péninsule ibérique où cet effet a également été détecté.

Enfin, nous avons observé un effet significatif des températures minimales en tant que barrières à la dispersion de *M. galloprovincialis*, et cet effet a été observé dans les mêmes zones que celles où l'effet de l'élévation (topographie) est détecté de manière significative. Bien que n'ayant pas observé de covariance entre l'élévation et la température sur l'ensemble de l'aire d'étude, il est évident que ces deux paramètres sont covariants à échelle locale. La température affecte le développement et la survie larvaire de cet insecte, (Tomminen 1993 Naves et al. 2007b, Koutroumpa et al. 2008, Naves & Sousa 2009) ainsi que l'activité de vol des adultes de *M. galloprovincialis* (Hernandez et al. 2011). La topographie affecte également le comportement de vol de cet insecte. En effet, Torrez-Vila et al. (2015) ont montré à travers une étude de capture-marquage-recapture que les adultes ont une tendance

à voler vers le bas d'une pente. La température et la topographie sont donc deux paramètres covariants et tous deux susceptibles d'expliquer l'effet « barrière » observé au niveau des reliefs. L'analyse des coefficients de commonalité a permis de montrer que les températures minimales constituent le paramètre qui explique le plus cet effet des reliefs, l'effet de la topographie (effet unique) contribuant de manière négligeable à la variance expliquée par le modèle de régression. Cependant, il est possible qu'il existe un effet additif de ces deux paramètres, car leur effet cumulé montre la meilleure contribution à la variance expliquée parmi toutes les combinaisons de paramètres.

Un effet barrière des fortes densités en pins sur la dispersion de *M. galloprovincialis*

Les pins constituent une ressource obligatoire pour *M. galloprovincialis* ; leur distribution et leur densité sont donc susceptibles d'affecter la dispersion de cet insecte, à la fois des immatures pour la nutrition de maturation et des femelles accouplées pour la recherche d'un site d'oviposition. Il a été montré dans les précédentes études que *M. galloprovincialis* n'était pas structuré génétiquement selon l'essence hôte, tout du moins au sein de la péninsule ibérique. L'effet de la densité sur la dispersion a donc été testé sur l'ensemble des pins, sans prendre en compte les variations spatiales de distribution et de densité de chacune des essences sur lesquelles cette espèce se développe. Pour étudier cet effet sur la dispersion, deux hypothèses mutuellement exclusives ont été considérées. La première hypothèse supposait un effet barrière des fortes densités forestières, et par conséquent un effet corridor des zones ouvertes. La seconde hypothèse impliquant un effet corridor des fortes densités en pins et un effet barrière des densités forestières faibles. L'étude de génétique spatialisée (article n°2) montre que la première hypothèse explique le mieux la structure génétique observée chez *M. galloprovincialis*. Il apparaît donc que cette espèce disperse peu lorsque la densité en pin est forte et que la ressource est abondante. Ces résultats sont appuyés par les observations du comportement de vol des adultes *in natura* (Bonifacio 2009, Torres-vila et al. 2015). Ce comportement phylopatrique en situation de ressource disponible n'est d'ailleurs pas spécifique à cette espèce puisque qu'il est rencontré chez la processionnaire du pin (*Thaumetopoea pityocampa*), un lépidoptère également oligophage sur pins (Démolin 1969).

La prévalence de cette hypothèse n'est pas intuitive a priori car *M. galloprovincialis* est généralement supposé être une espèce voilière, susceptible de se déplacer sur de grandes distances à la recherche d'arbres sénescents pour se développer. Cette ressource peut être relativement rare et inciter à la dispersion au sein des pinèdes. Cependant, bien que capable de se développer dans les arbres sénescents, il faut noter que le développement de cette espèce peut se faire dans les branches mortes retrouvées naturellement sur les pins vivants (phénomène d'auto élagage). *M. galloprovincialis* est par ailleurs rencontré en densité très élevée au sein de peuplements de pins sans dépérissement apparent (Landes, G. David com. Pers.). Les branches mortes constituent une ressource beaucoup mieux répartie dans l'espace et le temps qu'un arbre sénescent entier et il est probable que cette disponibilité en ressource explique le comportement phylopatrique mis en évidence ici. Par opposition, ces résultats suggèrent que les zones montrant une faible densité de pins ont tendance à favoriser la dispersion de *M. galloprovincialis*. Cette information est plus évidente à justifier car il semble a priori cohérent qu'une faible densité d'arbres hôtes pousse les adultes à voler plus pour accéder à leur ressource (Torres-Vila et al. 2015). Sur ce point, ces observations font également écho au comportement de dispersion de la processionnaire du pin. En effet, le suivi du front d'expansion de cette espèce vers le nord a permis de montrer que les faibles densités en pins rencontrées dans certaines zones (*i.e.* la Beauce, région Centre, France) ne constituent pas une barrière mais au contraire ont tendance à favoriser sa dispersion (Rossi et al. 2015).

Replacés dans un contexte de lutte contre l'expansion du PWN, ces résultats suggèrent de considérer avec précaution l'utilisation de zones de coupes franches autour des zones infectées. En effet, selon ces hypothèses, une zone ouverte ou à faible densité en pins est plus propice à la dispersion qu'un boisement dense, et est donc moins à même de limiter l'expansion des adultes porteurs du PWN.

L'étude de génétique spatialisée a également montré que la seconde hypothèse (les pins sont des corridors à la dispersion) expliquait de manière significative la structure génétique de *M. galloprovincialis* dans certaines zones. Ce signal antagoniste avec la première hypothèse a été observé à une fréquence bien plus faible, et est structuré géographiquement (moitié sud de la péninsule

ibérique), ce qui suppose qu'il n'est a priori pas issu d'artefacts liés aux erreurs de type I (faux positifs), ces biais pouvant subsister lors de l'utilisation de tests de Mantel partiels (Guillot & Rousset 2013). Ce résultat implique donc que les densités les plus importantes de pins peuvent localement constituer des corridors à la dispersion de *M. galloprovincialis*. Deux hypothèses non exclusives peuvent expliquer un tel signal contradictoire : (i) Dans les régions où le taux de boisement est faible, les densités en pins les plus fortes peuvent correspondre à des peuplements relativement épars dans un environnement où les pins sont totalement absents. Dans de tels cas, les densités de pins les plus fortes pour l'analyse sont en réalité des densités faibles si on les compare avec ceux rencontrés dans d'autres zones, et représentent de surcroît la seule ressource disponible. Dans ces conditions les pins peuvent donc constituer un corridor à la dispersion. (ii) L'existence d'un tel signal peut également trouver son origine dans la signature laissée par des phénomènes de migration de masse ou de recolonisation de certaines zones par *M. galloprovincialis*. En effet, les activités anthropiques ont fortement influencé la distribution et la densité des pins en péninsule ibérique à des échelles temporelles variables (Ruiz-Benito et al. 2012, Abel - schaad et al 2014, Lopez-Merino et al. 2014, Serra et al. 2014). Il est possible que les déconnexions et reconnexions du paysage résultant de ces activités aient généré des phénomènes de recolonisation le long de couloirs favorables.

L'intérêt de considérer simultanément des échelles et des localités multiples pour inférer l'effet des paramètres environnementaux sur la dispersion.

L'étude de génétique spatialisée (article n°2) a été conduite à plusieurs échelles et en plusieurs localités. Cette nouvelle approche basée sur des répliques d'aires d'étude a été initialement entreprise à but exploratoire car l'échelle des flux de gènes et l'échelle de l'effet des paramètres environnementaux sur ces flux n'étaient pas connues chez *M. galloprovincialis*. Si cette approche a en effet permis de lever ces incertitudes, son application a également apporté une contribution à la compréhension des facteurs affectant l'inférence dans les analyses de génétique spatialisée.

L'analyse à échelles multiples a tout d'abord permis de montrer un fort effet de l'échelle de l'étude sur la détection de l'effet des paramètres environnementaux sur la dispersion pour la majorité des paramètres testés (températures hivernales, élévation et pins en tant que barrières à la dispersion, ainsi que les distances euclidiennes simples). Globalement, une relation positive a été observée entre l'échelle et le nombre d'aires montrant un effet significatif. Un tel résultat est logique vu la très faible structuration génétique rencontrée aux échelles les plus faibles chez *M. galloprovincialis* et chez les *Monochamus* en général. Étendre la zone d'étude permet donc d'inclure des populations structurées et de mieux inférer l'effet des paramètres environnementaux. Sur ce point, les résultats obtenus rejoignent ceux de Dreier et al. (2014), montrant l'absence d'effet du paysage – à bien plus faible échelle certes (<10 km) – chez plusieurs espèces voilières, faute de différenciation génétique. Ces résultats sont bien sûr également dus au fait que les grandes aires incluent plus d'individus dans l'analyse de Mantel partiel, ce qui favorise le support statistique du test lorsque le signal existe, mais également, comme cela a été montré, au fait qu'une plus grande étendue couvre plus de variabilité dans les paramètres environnementaux testés et donc plus de contraste susceptibles d'affecter la dispersion. Cet effet échelle-dépendant de l'inférence en génétique spatialisée a été documenté chez plusieurs espèces (Murphy et al. 2010, Angelone et al. 2011, Keller et al. 2013). Celui-ci s'est révélé central pour étudier l'effet des paramètres environnementaux sur la dispersion des espèces à fort pouvoir dispersif tel que le Caribou (*Rangifer tarandus caribou*, Galpern et al. 2012). L'étude relative à l'article n°2 constitue donc un nouvel exemple de l'importance de l'échelle, qui reste généralement peu prise en compte en génétique spatialisée (Zeller et al. 2012).

Cette étude s'est également basée sur des réplifications spatiales d'aires d'étude. L'utilisation de ces réplifications a mis en évidence une hétérogénéité spatiale dans la détection de l'effet des paramètres environnementaux sur la dispersion à échelle fixe. Les résultats ont montré que cette variation dans l'inférence n'était pas seulement causée par une différence dans la distribution des paramètres testés, ou par des déséquilibres d'échantillonnage. Un tel résultat est inattendu, car il est généralement admis qu'une espèce a une réponse globalement similaire vis-à-vis d'un même paramètre environnemental au sein d'une même région. Pour expliquer de tels résultats, nos hypothèses reposent sur des arguments relatifs aux modifications récentes du paysage et à l'histoire

évolutive de l'espèce en péninsule ibérique. Il faut noter tout d'abord que l'effet des différentes variables sur la dispersion est en général moins détecté dans les zones où la structure génétique de *M. galloprovincialis* est la plus homogène (sud et est de la péninsule). On peut donc supposer que l'effet des paramètres environnementaux sur la dispersion ne soit pas détecté faute de mise en place d'une structure génétique relative à ces paramètres. Une telle situation peut dériver des modifications récentes du paysage – qui sont répandues en Espagne, comme cela a été évoqué plus haut – qui ont pu conduire à des migrations récentes de populations et conduire à une faible différenciation génétique des populations. La seconde explication de ce phénomène est de l'ordre de l'artefact, et correspond au déséquilibre dans les analyses causé par l'inclusion de deux lignées différenciées génétiquement dans une même aire d'étude. En effet, l'inclusion de deux lignées évolutives dans une aire d'étude a pour conséquence de cacher une différenciation génétique plus fine issue de l'effet des paramètres environnementaux. Ce cas a été rencontré de manière évidente au niveau de la zone de contact entre les deux clusters de péninsule ibérique (Portugal / Espagne). Les aires d'étude couvrant les deux lignées n'ont pas montré d'effet des paramètres testés tandis qu'un tel effet a été détecté lorsque l'une des deux lignées a été retirée.

De telles situations invitent à bien étudier la structure génétique d'une espèce dans son ensemble avant de choisir une aire d'étude pour inférer l'effet des paramètres environnementaux sur les flux de gènes. Ceci est particulièrement vrai chez les espèces à l'histoire évolutive complexe, et si le cas présenté ici est relativement simple (deux clusters bien définis), il n'est pas exclu qu'il existe une structure historique plus diffuse pouvant également déséquilibrer les analyses de génétique spatialisée en d'autres points de la péninsule ibérique (Pflüger & Balkenhol 2014).

L'utilisation de répliques d'échelles et de localités en analyse de génétique spatialisée présente également l'avantage de contourner certains biais relatifs à l'utilisation des tests de Mantel et de Mantel partiels. En effet, bien que très largement utilisés dans les approches de génétique spatialisée (Cushman et al. 2006, Cushman & Landguth 2010, Galpern et al. 2012, Castillo et al. 2014), les tests de Mantel ont été critiqués pour leur risque élevé d'erreur de type I (faux positifs, Guillot & Rousset 2013). La dimension quantitative d'une approche par répliques permet d'atténuer le risque de faux positifs car l'observation d'un signal structuré géographiquement – et donc rencontré

en plusieurs aires d'étude – montre a priori que le signal observé n'est pas issu d'un artefact ponctuel. La validité des tests de Mantel a également été critiquée en cas de covariance des variables environnementales testées (Guillot & Rousset 2013). La covariance entre variables est très répandue dans l'environnement et les problèmes qu'elle implique dans l'inférence faite en génétique spatialisée ne concernent pas uniquement les tests de Mantel. En couplant l'approche par tests de Mantel avec l'analyse de commonalité, il a été possible d'identifier les facteurs affectant réellement les flux de gènes et d'écarter ceux issus de la covariance avec ces facteurs. Enfin, l'utilisation de réplifications a permis de montrer l'existence d'un effet conflictuel entre deux hypothèses de résistance des paramètres environnementaux mutuellement exclusives. Nous avons observé que cet effet conflictuel était structuré géographiquement et justifiable biologiquement (effet de la densité en pins sur la dispersion). Un tel cas n'a, à notre connaissance, jamais été mis en évidence.

Les réplifications spatiales sont très rarement considérées en génétique spatialisée, et lorsqu'elles sont utilisées, un nombre relativement limité d'aires d'études est pris en compte (Short Bull et al. 2011, Zeller et al. 2012). Nos résultats montrent, sur un nombre important de réplifications, que leur utilisation est pourtant essentielle. En effet, les réplifications permettent d'inclure une dimension quantitative dans les résultats issus des analyses de génétique spatialisée et ce point est fondamental pour généraliser l'effet des paramètres environnementaux sur la dispersion. Inversement, ces résultats suggèrent qu'une étude de génétique spatialisée basée sur une seule zone et à une échelle unique risque de conduire à des conclusions fausses ou partielles. Cette conclusion constitue probablement l'apport majeur de l'article n° 2.

L'approche exposée ici est particulièrement adaptée aux espèces dont l'échelle des flux de gènes et de l'histoire évolutive n'est pas connue a priori et contribue à donner une image globale de l'effet des paramètres environnementaux plutôt qu'une information ponctuelle dépendant du dispositif expérimental. Cette approche pourrait donc être adaptée aux espèces voilières au fort pouvoir dispersif, qui constituent une large part des espèces vectrices chez les insectes.

Dans cette section, l'existence de barrières à la dispersion de *M. galloprovincialis* a été mise en évidence sur une base moléculaire. Ce point est primordial dans l'évaluation du potentiel de dispersion

du PWN. Cependant, les résultats de cette approche génétique donnent une estimation de la dispersion efficace du vecteur, qui ne correspond pas nécessairement à celle du PWN. De plus, les patrons génétiques à partir desquels les flux de gènes sont inférés peuvent s'être mis en place dans un environnement ancien, qui ne correspond pas nécessairement à celui rencontré actuellement. Un modèle mathématique a donc été utilisé pour simuler la dispersion de populations de vecteurs infestés par le PWN dans le paysage hétérogène actuel. Cette approche vient en complément des précédentes estimations et permet d'affiner les prédictions de l'expansion potentielle du PWN en Europe.

V.1.3- Synergie de l'association phorétique et dispersion du nématode

Les modèles mathématiques simulant la dispersion des populations invasives sont largement utilisés pour estimer le potentiel d'invasion et prédire les zones à risque (Shigesada & Kawasaki 1997, Pitt et al. 2009, Carrasco et al. 2010, Robinet et al. 2012). Cependant, ces modèles ont rarement été appliqués dans les cas d'associations phorétiques en intégrant la complexité de ce phénomène.

Le modèle utilisé au cours de ma thèse présente la structure d'un modèle de réaction-diffusion déjà appliqué pour simuler la dispersion du PWN (Robinet et al. 2009, 2011). Il s'en distingue par trois points dans son application, pour intégrer l'effet synergique de l'association nouvellement formée avec la prise en compte : i) de la préexistence et de la densité du vecteur dans la dispersion du nématode, ii) de l'augmentation de la densité de vecteurs liée à l'invasion du PWN, et iii) de la latence dans le processus de recrutement des populations résidentes de vecteurs suite à l'arrivée du PWN. Il s'agit de la première application d'un modèle de dispersion naturelle du PWN à une échelle régionale.

Une distribution continue et une densité importante de vecteurs

Le premier obstacle potentiel à la dispersion naturelle du PWN est la disponibilité en arbres hôtes et en vecteurs. La distribution simulée des populations résidentes de vecteurs (article n°3) –

basée sur la densité en pins et les densités estimées de populations associées – suggère que *M. galloprovincialis* a une distribution quasi-continue en plaine, bien que sa densité puisse varier. Les piégeages réalisés au cours des différentes études vont dans le sens de ce résultat : le vecteur a été capturé dans toutes les zones échantillonnées, même dans les zones à très faible densité forestière, où les pins forment des patches dispersés, mais également dans les peuplements de pins décrits comme non favorables à son développement (*P. pinea*, Population Huesca, sud des Pyrénées).

Les simulations de dispersion de populations de vecteurs infestées par le PWN en plaine n'ont pas mis en évidence de barrière à la dispersion liée à une absence de vecteur. D'après les simulations de dispersion et les piégeages, il semble donc que la connectivité du paysage soit importante et que la disponibilité en vecteur et en ressource ne soit pas un frein à la progression naturelle du PWN, tout du moins dans la zone d'étude. Ces résultats contredisent les simulations de Yoshimura et al. (1999) qui ont montré l'existence d'un seuil de densité en pins sous lequel la PWD ne parvient pas à s'établir. Même s'il est question de dispersion du PWN dans les deux cas, il est difficile de discuter ces différences car la zone géographique, l'échelle spatiale de l'étude (et donc la résolution des données utilisées), le vecteur et les pins impliqués ne sont pas les mêmes.

Il peut être objecté que cette absence de frein soit due à une estimation trop élevée des densités de populations résidentes de vecteurs (que nous avons basé sur les piégeages faits dans la zone pyrénéenne). Notre estimation était de 42.12 adultes par hectare, pour un couvert forestier de 100 % (correspondant à 156000 arbres par km², $K_0 = 0.027$ adulte / arbre). Des estimations récentes en milieu naturel dans une autre région d'Espagne (Extremadure) ont donné des densités de vecteurs variant de 96 à 474 adultes par hectares, avec des formats de pièges et des attractifs phéromonaux équivalents à ceux utilisés pour notre estimation (Torres-vila et al. 2015). Ces résultats tendent à montrer que nos mesures sous-estiment plus qu'elles ne surestiment cette densité. Les simulations ne résultent donc a priori pas d'une surestimation de la densité de population résidente dans notre modèle.

Les Pyrénées, une barrière partielle à l'expansion du PWN

Les simulations de dispersion de vecteurs infestés par le PWN à travers les Pyrénées ont été réalisées avec de multiples seuils de survie du vecteur et de développement de la PWD. Cette approche à seuils multiples a été appliquée pour tester plusieurs scénarii, mais également pour prendre en compte les incertitudes qui persistent concernant les niveaux réels de survie du coléoptère dans cette zone. Le principal facteur testé correspond à l'effet des températures froides d'altitude (températures hivernales et élévation, les deux étant étroitement liées). Nous nous sommes focalisés sur ce paramètre car *M. galloprovincialis* réalise la fin de son développement larvaire en hiver, mais également car ces températures se sont révélées être de bons indicateurs de survie chez d'autres insectes tels que la processionnaire du pin, dont les phases larvaires se développent également en hiver et sur les pins. (Battisti et al. 2005, Robinet et al. 2007).

Malgré l'amplitude des différents seuils de survie de *M. galloprovincialis* testés (température hivernale et élévation), les simulations de dispersion basées sur cette survie ont donné des prédictions globalement similaires. Les régions les plus élevées constituent systématiquement des barrières à la dispersion des vecteurs infestés par le PWN, bien que l'étendue de ces barrières varie fortement en fonction du seuil considéré. Les zones côtières, de part et d'autre du relief, constituent en revanche systématiquement des corridors à la dispersion du PWN. Il faut noter que les barrières basées sur les seuils de survie de *M. galloprovincialis* restent globalement assez minces, car le vecteur est rencontré jusqu'à des altitudes assez élevées (1590 m). De plus, notre étude à l'échelle européenne montre que cette espèce semble être capable de se maintenir dans des zones où les températures hivernales peuvent être très basses. La disponibilité en vecteurs dans les zones froides d'altitude ne semble donc pas constituer un frein majeur à la progression du PWN.

L'utilisation de seuils de survie pour tester l'existence de barrières peut être critiquée. En effet la limite de survie d'une espèce en altitude n'est jamais une limite franche mais plutôt un gradient qui peut varier spatialement et temporellement. Les données de survie de *M. galloprovincialis* à différentes températures sont actuellement insuffisantes pour définir un seuil précis. Il a donc été choisi de faire une évaluation empirique de seuil de survie, basée sur les observations de piégeage

réalisées sur la zone d'étude et incluant les données de présence de l'espèce à large échelle géographique.

Les simulations ont montré que le seuil d'expression de la maladie (une température moyenne de 20°C en juillet) définit des barrières potentiellement beaucoup plus larges à la dispersion du PWN que celles basées sur la survie du vecteur seul. Ces zones défavorables à l'expression de la PWD font même disparaître le corridor de l'extrémité ouest du relief sous les températures actuelles. Le potentiel de dispersion du PWN dans les zones où la maladie ne s'exprime pas n'est pas connu. Cependant, il est très probable que ces zones soient défavorables à sa progression. En effet, si les arbres ne déclinent pas, on s'attend à ce que les vecteurs ne pullulent pas autant que dans les zones où la maladie s'exprime. Dans l'hypothèse où ces zones seraient effectivement défavorables à l'expansion du PWN, nos simulations prédisent un fort potentiel des températures limites d'expression de la PWD pour limiter la progression du PWN niveau des Pyrénées.

Collectivement, les résultats des simulations issues de notre modèle suggèrent que les régions élevées des Pyrénées représentent un frein potentiel à l'expansion de la PWD. Ces résultats corroborent ceux des simulations faites en Chine, qui prédisent également un effet de l'altitude sur l'expansion de cette maladie (Robinet et al. 2009, Roques et al. 2015). Cette information est une indication importante pour la mise en place de stratégies de détection précoce et de contrôle de l'expansion du PWN, et une telle conclusion est potentiellement applicable à d'autres massifs montagneux en Europe où les conditions biologiques et climatiques présentent des similarités.

Réponse des partenaires de l'association aux changements climatiques

Les modifications climatiques actuellement observées se traduisent par une élévation globale des températures (IPCC, 2014). Cet effet affecte particulièrement des régions d'altitude et les espèces qui s'y trouvent, ainsi que les interactions interspécifiques (Guralnick 2007, Jamieson et al. 2012). Au cours des différentes études de cette thèse, nous avons montré que les reliefs et les températures basses qui leur sont associées représentent une barrière naturelle à la dispersion de *M.*

galloprovincialis et probablement la principale barrière à l'expansion de la PWD. Une élévation des températures est donc susceptible d'atténuer ou de faire disparaître ces barrières. Cet effet potentiel est particulièrement visible au niveau du seuil d'expression de la maladie. Nos simulations ont montré que l'étendue de la zone défavorable à l'expression de la maladie au sein de la chaîne pyrénéenne variait très fortement avec une augmentation modérée des températures. Une élévation moyenne de 1°C pourrait par exemple entraîner la connexion d'aires favorables à l'expression de la maladie au niveau de l'extrémité ouest des Pyrénées. D'après nos simulations, l'élévation des températures sous l'effet du changement climatique est donc a priori favorable à l'expansion du PWN. Ce constat est en accord avec l'évolution de la distribution des dommages liés à la PWD en Asie. En effet, le réchauffement actuel est supposé être responsable de l'expression de la PWD au sein de zones froides historiquement supposées être sans risque au Japon et en Chine (Nakamura et al. 2013, Huang et al. 2013). Les simulations de développement de la PWD montrent également qu'un réchauffement de 2°C au niveau de latitudes actuellement non-favorables à l'expression des symptômes de la maladie, rendrait ces zones favorables (Belgique, Evans 2014). La dispersion du PWN peut également être influencée par les effets indirects du changement climatique. L'augmentation de la fréquence des feux et des tempêtes, consécutifs à l'élévation des températures, peut générer de gros volumes de bois mort et favoriser les pullulations de vecteurs, comme cela a été montré pour *M. scutellatus* en Amérique du nord (Gandhi et al. 2007, Bélanger et al. 2013). De même, les périodes de sécheresse entraînent une augmentation des dépérissements de pins sensibles à la maladie (James et al. 2006). De tels effets n'ont pas été intégrés dans notre modèle faute de données, mais ils constituent des éléments potentiellement favorables à l'expansion du PWN.

Complémentarité des approches de modélisation et des observations génétiques

L'intérêt de combiner une étude de génétique des populations et de modélisation était de faire une validation croisée des résultats à partir d'approches complémentaires ayant chacune des avantages et des limites dans l'estimation du potentiel de dispersion du PWN par son vecteur. L'analyse basée sur une estimation moléculaire de la dispersion a permis d'observer la dispersion effective de *M.*

galloprovincialis dans le paysage et d'identifier les couloirs de dispersion de ce vecteur. Cependant, cette approche ne nous a pas permis d'estimer une mesure précise des distances de dispersion car il est difficile de savoir à quelle échelle de temps ces patrons se sont mis en place (voir Bermond et al. 2013 pour un contre-exemple) et si ces flux de gènes observés se sont mis en place dans l'environnement actuel. De plus, elle n'intègre pas l'évolution de la dispersion du vecteur en situation de synergie avec le PWN, l'estimation des flux de gènes ayant été faite sur une zone où le PWN n'est pas encore présent. La modélisation de la dispersion est une approche qui nous a permis de tester la dispersion dans des conditions actuelles de climat et de distribution des paramètres environnementaux, et notre modèle affiné a permis de prendre en compte la synergie vecteur-PWN. Cependant, le modèle utilisé décrit le processus de dispersion sans prendre en compte l'effet de certains paramètres environnementaux sur la dispersion (qui ne sont pas encore suffisamment connus). Ces paramètres, tels que l'isolement par adaptation à l'environnement local, peuvent être essentiels dans les processus de dispersion (Pflüger & Balkenhol 2014).

Au cours de l'étude sur la barrière pyrénéenne (Article n°3), nous avons observé une forte cohérence entre les résultats issus des deux approches, à la fois concernant les corridors (forte dispersion en plaine) que les barrières (zones élevées) à la dispersion. Cette combinaison d'approches permet donc de renforcer nos conclusions sur le potentiel dispersif des populations de vecteur infestées par le PWN à travers ce relief.

V.1.4 - Perspectives d'études sur le modèle nématode-*Monochamus*

Comme cela a été évoqué en introduction, la dispersion des espèces impliquées dans une APNF est un phénomène complexe qui implique de multiples facteurs. Malgré le nombre important de connaissances acquises sur le modèle *Monochamus*-PWN, plusieurs inconnues subsistent et restent à explorer pour mieux anticiper l'invasion du PWN en Europe.

La première de ces inconnues est le rôle potentiel des autres espèces du genre *Monochamus* dans la dispersion du PWN, et en particulier de *M. sutor*, l'espèce la plus proche de *M. galloprovincialis*. Cette espèce est absente de la péninsule ibérique (à l'exception des Pyrénées) et son potentiel de transport du nématode n'a donc pas pu être observé à ce jour. *M. sutor* est largement distribué dans les zones fraîches et humides d'Europe (altitude > 800 m), et pourrait constituer un relais pour la dispersion du nématode dans les zones les plus élevées (Koutroumpa et al. 2013). Cependant, plusieurs éléments laissent supposer un rôle secondaire de cette espèce dans la dispersion du nématode. Tout d'abord, cette espèce se maintient dans des zones froides qui correspondent justement aux zones défavorables à l'expression de la PWD. Dans ces aires, les dépérissements de pins et l'effet synergique sur la densité de *Monochamus* ne sont donc pas attendus, pour chacune des deux espèces. Il faut également noter que *M. sutor* a une gamme d'hôte plus large que *M. galloprovincialis*. On le retrouve sur *Picea* spp., *Larix* spp. et *Pinus* spp. (Koutroumpa et al. 2013, Evans et al. 1996), ce qui diminue a priori la fréquence de la transmission *Pinus-Pinus*, les autres essences (*Picea* spp., *Larix* spp.) étant moins favorables, voir défavorables au développement nématode (Evan et al. 1996).

Un second point important dans l'estimation du potentiel de dispersion du PWN est l'efficacité de sa transmission par *M. galloprovincialis*. Or, ce point reste très mal connu, on ignore par exemple la proportion d'adultes infestés lorsqu'ils émergent d'un bois infesté par le PWN. Des analyses préliminaires montrent que leur fréquence peut varier de 12 à 54 % (Sousa et al. 2001, com. Pers. G. David). Des connaissances plus fines sont également requises sur l'efficacité de la transmission du nématode par transmission primaire et secondaire et leur importance réciproque, ainsi que l'effet de la charge en nématode sur cette efficacité. Enfin, l'information qui fait le plus défaut aujourd'hui est sans doute la capacité du PWN à être dispersé dans des zones où la PWD ne s'exprime pas.

Il faut également noter l'existence de biais potentiels dans l'évaluation de la densité de vecteurs, en particulier dans les captures de *M. galloprovincialis* à partir de piégeages utilisant l'attractif phéromonal spécifique Galloprotect (SEDQ, Spain). En effet, plusieurs observations

indiquent que les différentes lignées ne répondraient pas de manière similaire à cette phéromone. Bien que non formellement testé, il semble que les individus d'Europe de l'est soient piégés en densité bien moindre que les individus d'Europe de l'ouest, en conditions similaires. Par exemple, un piégeage intensif de plus de 100 pièges mis en place en Belgique sur trois ans n'a permis de collecter que sept individus (Jean-Claude Grégoire com. Pers.). A l'inverse, au cours de nos échantillonnages, en particulier en France et en Espagne, nous avons capturé plusieurs dizaines d'individus par piège au cours d'une seule saison de collecte. Cette différence de réponse est peu susceptible d'avoir impacté les études de cette thèse car nous nous sommes principalement focalisés sur la structure génétique des individus capturés. En revanche, les études futures évaluant les densités de population de vecteurs sur la base de piégeage doivent intégrer cette éventualité, en particulier si des localités distantes sont considérées. Les études évaluant la structure génétique du vecteur à échelle locale au niveau des zones de contact entre lignées divergentes devront également prendre en compte cet aspect.

Concernant les facteurs pouvant influencer la dispersion de *M. galloprovincialis*, il a été montré que plusieurs espèces de *Monochamus* américains localisaient leur hôte grâce aux phéromones d'agrégation émis par les espèces de scolytes sympatriques à ces espèces (Allison et al. 2001). Cette dimension n'a pas été prise en compte dans nos travaux faute de données similaires chez les *Monochamus* Européens. Cependant, ce phénomène pourrait affecter positivement la dispersion de *M. galloprovincialis* au-delà des distances mesurées actuellement, en particulier vers les zones où les densités de populations de scolytes associés aux pins explosent. Dans le cadre de l'invasion du nématode du pin, un tel phénomène pourrait également augmenter l'attractivité des arbres infestés et contribuer au recrutement des vecteurs sur un large périmètre.

Enfin, l'ensemble de cette thèse s'est essentiellement focalisée sur la dispersion naturelle du PWN à travers son APNF avec *M. galloprovincialis*. Or, il faut noter que les transports par l'homme de matériaux infestés par le PWN avec ou sans son vecteur contribuent également à l'expansion du PWN. Les transports sur de longues distances peuvent grandement accélérer cette expansion (Robinet et al. 2009). Bien que globalement corrélés aux activités anthropiques en Chine (Robinet et al. 2009), ces événements restent très stochastiques et difficiles à intégrer dans les simulations en Europe, où le peu de recul sur l'invasion (nématode découvert en 1999 et très peu d'événements de dispersion à

longue distance clairement identifiés) ne permet pas pour le moment de faire de projections. De plus, plusieurs inconnues sur l'effet réel des mesures de contrôle persistent, ce qui rend les prévisions d'expansion difficiles à partir des données issues de l'expansion au Portugal. Des modèles sont en cours de développement pour intégrer ces paramètres aux simulations de dispersion du PWN en Europe (Evans 2014).

V.2 - Synthèse et implications pour les stratégies de lutte contre le PWN

Cette thèse a permis de contribuer significativement à la connaissance de la structure génétique et de la dispersion de *M. galloprovincialis* ainsi qu'à l'estimation du potentiel d'expansion du nématode du pin à travers son association phorétique avec ce vecteur. Les différentes études réalisées ont permis de clarifier le contour de *M. galloprovincialis* et de montrer l'absence d'hybridation avec l'espèce sœur *M. sutor*. Ces études ont également permis de mettre en évidence une structure phylogéographique de type recolonisation postglaciaire chez cette espèce, et une distribution des lignées évolutives liée principalement au régime de précipitations, au-delà de l'effet des oscillations glaciaires du quaternaire. Ce dernier point semble suggérer, par extension, qu'il n'existe pas de spécialisation pour une essence hôte de pins. L'étude des flux de gènes a permis de montrer l'existence de barrières à la dispersion de *M. galloprovincialis*, principalement au niveau des reliefs, et suggèrent que les températures basses qui leur sont associées expliquent un tel effet. Nous avons également montré que les fortes densités en pins représentaient des barrières à la dispersion, et inversement, que les zones ouvertes avec des pins épars pouvaient constituer des corridors à cette dispersion. Enfin, le modèle affiné simulant l'expansion naturelle du nématode au travers de son association avec *M. galloprovincialis* suggère que celui-ci rencontre peu d'obstacle en plaine, mais peut être affecté par les reliefs. Ces derniers résultats sont largement corroborés par les résultats de génétique de *M. galloprovincialis*. Enfin, les simulations réalisées prédisent un fort effet des

températures basses sur l'expansion du PWN, et par conséquent un risque élevé de voir ces barrières réduites par le réchauffement climatique.

Ces résultats ont plusieurs implications pour le contrôle de l'expansion du nématode en Europe. Ils indiquent que les reliefs dans leur ensemble ont un bon potentiel pour freiner l'expansion naturelle de la PWD sous le climat actuel. Cette information est d'autant plus intéressante qu'il existe plusieurs zones de reliefs entre les sites actuellement colonisés par le PWN et le reste de l'Europe (reliefs espagnols, Pyrénées, Massif central, Alpes). Ces reliefs constituent donc des barrières potentielles à prendre en compte dans la lutte contre ce ravageur. Inversement, les couloirs de migrations potentiels du nématode mis en évidence de part et d'autre des Pyrénées peuvent orienter les plans de détection précoce en renforçant les dispositifs de piégeage sur ces couloirs. Les résultats de cette thèse suggèrent que les coupes rases de pins employées pour confiner le PWN pourraient avoir tendance à favoriser la dispersion des *M. galloprovincialis* et donc potentiellement celle du PWN par le biais d'individus infectés. Cette information fait écho aux résultats d'une étude récente montrant l'inefficacité de ce type de mesure de contrôle (Jactel et al. 2015). Cependant, nos résultats doivent être considérés avec précaution car l'échelle de l'étude de génétique spatialisée dont sont issues ces conclusions (>200 km) est plus grande que celle des plans de gestion (quelques km). En revanche, ces indications invitent à considérer une solution inverse : pas de coupe rase et simple retrait des arbres infectés dans les zones touchées, comme cela a récemment été proposé (Jactel et al. 2015). Enfin, cette thèse a montré la possibilité d'assigner les individus de vecteurs à une des lignées européennes. Ces assignations peuvent permettre de retracer les événements de dispersion à longue distance, d'identifier les points sensibles d'introduction de vecteurs originaires de zones géographiques distantes et ainsi de mieux prévoir la dispersion du PWN par transport artificiel de vecteurs infectés.

V.3 - Contribution du modèle à la connaissance des associations phorétiques

Des outils adaptés à l'estimation du potentiel dispersif des espèces impliquées dans une APNF

Les travaux réalisés au cours de cette thèse ont apporté un ensemble de points fondamentaux relatifs aux d'outils disponibles pour l'étude de la dispersion des espèces impliquées dans une association phorétique nouvellement formée. Tout d'abord, le développement d'une approche de génétique spatialisée à échelles et localités multiples est un outil dont l'application a permis de contourner certains biais dans l'estimation de l'effet des paramètres environnementaux sur la dispersion des espèces à fort pouvoir dispersif et à histoire évolutive complexe. Cette approche a en particulier permis de montrer que certaines espèces peuvent ne pas être structurées en fonction des paramètres testés de manière homogène à l'échelle de leur aire de distribution. Cette méthode a également permis d'observer un effet négatif des clusters différenciés dans les analyses de génétique spatialisée. Ce dernier point n'ayant, à notre connaissance, pas été exploré pour ce type d'approche. Cette méthode a donc des applications potentielles pour estimer la dispersion chez les vecteurs impliqués dans une APNF de type vecteur natif – pathogène introduit, car ces vecteurs sont en général des espèces à forte aptitude à la dispersion (Robinet et al. 2009, Potter et al. 2011). En revanche cette technique doit être adaptée aux cas inverses (vecteur introduit – pathogène natif) pour lesquels la structure génétique du vecteur dans sa zone d'introduction peut être insuffisante pour inférer un effet des paramètres environnementaux sur la dispersion. Dans de telles situations, il est probablement plus approprié d'utiliser cette technique dans la zone native du vecteur (où une structure génétique a pu se mettre en place) et d'étudier l'expansion possible de ce vecteur dans sa zone d'introduction par approche de modélisation de niche potentielle (He et al. 2015).

Au cours de cette thèse, nous avons également développé un modèle d'expansion prenant en compte les interactions vecteur – pathogène, et en particulier l'effet de synergie de ce couple. Ce modèle constitue une base qui peut être adaptée à d'autres modèles d'APNF, car les autres cas connus ont des

fonctionnements très similaires (dépérissements des arbres et explosions démographiques de vecteurs propices à la dispersion du pathogène, Mamiya & Eda 1972, Potter et al. 2011, Yan et al. 2005). Toutefois ce modèle doit être affiné pour intégrer la complexité de certaines interactions, en particulier lorsque l'effet de synergie implique d'autres espèces endémiques que le vecteur lui-même. Dans certaines situations, en effet, la dispersion du pathogène est due à l'action conjuguée de deux insectes natifs. C'est le cas du champignon invasif *Seiridium cardinale* (Coelomycetes) induisant de sévères dépérissements aux peuplements de *Cupressus sempervirens* dans le centre de l'Italie. Ce champignon est transmis par hémiptère vecteur *Orsillus maculatus* (Lygaeidae) uniquement lorsque les cônes de l'arbre sont préalablement attaqués par *Megastigmus wachtli* (Hymenoptera Torymidae) (Battisti et al. 2000). Dans de telles situations, le modèle doit prendre en compte la préexistence des deux insectes natifs et les différences dans l'effet de synergie, car ces espèces se sont pas toutes favorisées par l'arrivée du pathogène (seul *Orsillus maculatus* dans ce cas).

Collectivement, les approches développées dans cette thèse apportent donc une contribution à la « boîte à outils » disponible pour estimer le potentiel dispersif des espèces impliquées dans une APNF. Cependant, elles doivent être déclinées suivant la multitude et la complexité des situations générées par ce phénomène.

L'association *PWN-Monochamus* : un bon modèle pour étudier les APNF ?

Plusieurs éléments font émerger l'association *PWN-Monochamus* comme système modèle pour étudier les associations phorétiques nouvellement formées de type mutualiste, autant sur un plan appliqué que fondamental. La première raison est que cette association a généré des impacts considérables, autant sur les écosystèmes que sur les activités humaines associées au commerce et à la production de bois (Evans et al. 1996). En conséquence, l'invasion a été relativement bien documentée dès le début, et l'observation relativement facile des dépérissements causés par la PWD a facilité cette collecte de données de distribution. Nous possédons donc un historique relativement détaillé de cette invasion sur les différents continents. La dimension des impacts de cette association a également

permis de mobiliser un effort de recherches important sur les différents aspects de cette association. Cet effort, fédéré par des programmes menés à échelle européenne (PHRAME, REPHRAME), a conduit à la constitution d'un corpus de connaissances qui représente une base solide pour étudier ces systèmes complexes. En conséquence, nous disposons d'un ensemble d'éléments appliqués, dont plusieurs relatifs à cette thèse, pour mieux gérer et prévoir les impacts d'autres associations.

Au-delà de ces considérations pratiques, ce système ouvre également des pistes de recherche fondamentales particulièrement stimulantes. En effet, l'invasion du PWN entraîne une situation originale : l'acquisition soudaine par une population de vecteur d'une valeur sélective (fitness) supérieure. Ceci est dû au fait que l'arrivée du nématode permet aux vecteurs d'accéder à un grand volume de ressources, par affaiblissement des arbres qu'ils infestent. Cette situation est particulière car cette valeur sélective supérieure ne correspond pas à un avantage sélectif ayant une base génétique. Elle tranche donc avec les situations mieux étudiées impliquant la dispersion d'une population porteuse d'un gène procurant un avantage sélectif, comme par exemple d'un gène de résistance à un insecticide (Steinbach et al. 2015). Cette situation représente un cadre idéal pour tester plusieurs hypothèses évolutives relatives aux espèces endémiques. En particulier, il serait intéressant de tester l'effet de l'augmentation démographique du vecteur sur sa structure génétique dans un contexte théorique de « vague de pullulation », où nous avons donc la possibilité d'estimer la structure avant et après l'arrivée du PWN et de la PWD. L'hypothèse sous-jacente étant qu'une forte densité puisse augmenter la proportion d'individus dispersant à longue distance et favoriser le brassage des génotypes. Une telle hypothèse est d'ailleurs suggérée pour expliquer l'absence d'une structure génétique forte chez *M. alternatus*, suite à l'invasion du PWN au Japon (Kawai et al. 2006). Cette situation serait particulièrement intéressante à observer chez *M. galloprovincialis* au niveau des zones de contact entre lignées différenciées et adaptées à des conditions environnementales particulières. La structure serait-elle conservée après passage de la PWD (maintien des génotypes de vecteurs adaptés à l'environnement) ou serait-elle modifiée en profondeur par ce phénomène (primauté du fitness de l'association avec le PWN sur l'adaptation aux conditions locales) ? Ces questions sont fondamentales car elles touchent au moteur même de l'évolution : comment le changement de niveau de ressource et

la dynamique d'une population qu'elle induit peut affecter la structure génétique et la trajectoire évolutive de l'espèce entière.

Du point de vue des associations phorétiques, il serait également intéressant de tester l'existence de phénomènes de sélection de certains génotypes de *Monochamus* adaptés au transport du PWN. En effet, l'invasion du PWN recrée les conditions primordiales d'une interaction vecteur-pathogène et permet donc de tester les phénomènes évolutifs qui ont conduit aux relations phorétiques, plus stables, que nous rencontrons aujourd'hui chez de nombreux insectes forestiers. L'association PWN-*Monochamus* est d'autant plus adaptée à l'étude de ces phénomènes qu'il est possible de travailler avec des répliques de l'invasion, en Europe et en Asie, ce qui permet de tester ces hypothèses en s'affranchissant de la dimension stochastique et unique de chaque phénomène d'invasion (Melbourne & Hastings 2009).

Malgré l'intérêt des questions précitées, ces aspects n'ont pas pu être abordés dans le cadre de l'invasion européenne du PWN. En effet, les conditions n'y sont pas réunies pour tester ces hypothèses, tout du moins pour le moment. Ceci est dû au fait que l'invasion au Portugal est relativement récente, couvre une surface assez faible et a été très fortement affectée par les transports humains et les plans d'éradication. Tester l'effet d'une « vague » de pullulation est difficile à mettre en place dans ces conditions. De ce point de vue, la situation en Asie est différente et probablement plus adaptée à ces questions, car l'invasion du PWN y est plus ancienne et couvre actuellement une plus grande aire. Une autre difficulté de la situation européenne réside dans le fait que les populations du cluster portugais (où le PWN se trouve actuellement) sont peu structurées génétiquement. Cette homogénéité est probablement due à l'expansion géographique des populations qu'a entraîné l'afforestation du pays au cours des siècles derniers (Mendes et al. 2004). Un cluster très homogène génétiquement est peu propice à une étude visant à évaluer les modifications de structure génétique de ses populations. Il est probable que les marqueurs microsatellite utilisés dans cette thèse soient d'une résolution insuffisante pour étudier ces phénomènes micro-évolutifs à l'échelle du cluster ibérique de l'ouest. Des marqueurs plus nombreux et plus résolutifs, tels que les SNP (Single Nucleotide Polymorphism), seront probablement plus adaptés et sont actuellement en cours de développement.

Cette thèse constitue donc une base pour aborder ces questions dans le futur en Asie, ou en Europe si l'évolution de la distribution du PWN permet la mise en place de telles études.

Vers une émergence des APNF ?

Malgré la relative singularité des APNF, plusieurs éléments indiquent que la fréquence de ces associations pourrait s'intensifier à l'avenir. D'une part, une large part des invasions biologiques concernent les insectes, et en particulier les familles des coléoptères, des hémiptères et des hyménoptères (Roques et al. 2010). Or, ces familles contiennent une part significative d'espèces associées aux plantes et vectrices de pathogènes à l'état naturel, et sont donc naturellement susceptibles de former de nouvelles associations. D'autre part, on assiste actuellement à une augmentation d'introduction de pathogènes (Eschen et al. 2015), dont l'établissement et les dommages sont favorisés par les changements climatiques (La Porta et al. 2008). De manière générale, on observe actuellement une augmentation du nombre d'émergence de maladie d'origine fongique dans le monde, en particulier celles des plantes (Fisher et al. 2012, Wingfield et al. 2015). Enfin, l'intensité actuelle des échanges de plantes en pot entre continents (Smith et al. 2007, Santini et al. 2013) a pour effet de déplacer l'ensemble des parasites, phytophages et pathogènes qui leur sont associés, favorisant ainsi les interactions entre espèces et l'apparition de nouvelles associations. Dans cette perspective, et vu les dommages environnementaux et économiques associés aux APNF, il est indispensable de poursuivre l'étude de ces phénomènes afin de mieux prévoir et gérer leurs impacts.

Références bibliographiques

Références

- Abelleira A, Picoaga A, Mansilla JP, Aguin O, 2011. Detection of *Bursaphelenchus xylophilus*, causal agent of pine wilt disease on *Pinus pinaster* in Northwestern Spain. *Plant Disease* 95, 776.
- Abel-Scaad D, Lopez-Saez JA, Pulido F, 2014. Heathlands, fire and grazing. A paleoenvironmental view of Las Hurdes (Cáceres, Spain) history during the last 1200 years. *Forest Systems* 23, 247–258.
- Akbulut S, Linit MJ, 1999. Flight performance of *Monochamus carolinensis* (Coleoptera: Cerambycidae) with respect to nematode phoresis and beetle characteristics. *Environmental Entomology* 28, 1014–1020.
- Akbulut S, Stamps WT, 2012. Insect vectors of the pinewood nematode: a review of the biology and ecology of *Monochamus* species. *Forest Pathology* 42, 89–99.
- Allendorf FW, Lundquist LL, 2003. Introduction: population biology, evolution, and control of invasive species. *Conservation Biology* 17, 24–30.
- Allison JD, Borden JH, McIntosh RL, Grott P, and Gries R, 2001. Kairomonal responses by four *Monochamus* species (Coleoptera: Cerambycidae) to bark beetle pheromones. *Journal of Chemical Ecology* 27, 633–646.
- Angelone S, Kienast F, Holderegger R, 2011. Where movement happens: scale-dependent landscape effects on genetic differentiation in the European tree frog. *Ecography* 34, 714–722.
- Aslan CE, Sikes BA, Gedan KB, 2015. Research on mutualisms between native and non-native partners can contribute critical ecological insights. *NeoBiota* 26, 39–54.
- Avice JC, 2000. Phylogeography: The history and formation of species. Cambridge, US: Harvard University Press, pp. 447.
- Baldo L, de Queiroz A, Hedin M, Hayashi CY, Gatesy J, 2011. Nuclear-mitochondrial sequences as witnesses of past interbreeding and population diversity in the jumping bristletail *Mesomachilis*. *Molecular Biology and Evolution* 28, 195–210.
- Barbero M, Loisel R, Quézel P, Richardson DM, Romane F, 1998. Pines of the mediterranean basin. 153–170. In Richardson D.M. (ed.), *Ecology and biogeography of Pinus*. Cambridge, GB: University Press, pp. 527.
- Battisti A, Stastny M, Netherer S, Robinet C, Schopf A et al., 2005. Expansion of geographic range in the pine processionary moth caused by increased winter temperatures. *Ecological Applications* 15, 2084–2096.
- Battisti A, Cantini R, Feci E, Frigimelica G, Guido M et al., 2000. Detection and evaluation of seed damage of cypress, *Cupressus sempervirens* L., in Italy. *Seed Science and Technology* 28, 199–208.
- Bélanger S, Baucé E, Berthiaume R, Long B, Labrie J et al., 2013. Effect of temperature and tree species on damage progression caused by whitespotted sawyer (Coleoptera: Cerambycidae) larvae in recently burned logs. *Journal of Economic Entomology* 106, 1331–1338.
- Bensasson D, Zhang D, Hartl DL, Hewitt GM, 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution* 16, 314–321.
- Berec L, Angulo E, Courchamp F, 2007. Multiple Allee effects and population management. *Trends in Ecology and Evolution* 22, 185–191.
- Bermond G, Blin A, Vercken E, Ravigne V, Rieux A et al., 2013. Estimation of the dispersal of a major pest of maize by cline analysis of a temporary contact zone between two invasive outbreaks. *Molecular Ecology* 22, 5368–5381.
- Bertheau C, Schuler H., Krumböck S, Arthofer W, Stauffer C, 2011. Hit or miss in phylogeographic analyses: the case of the cryptic NUMTs. *Molecular Ecology Resources* 11, 1056–1059.

- Boggs CL, Freeman KD, 2005. Larval food limitation in butterflies: effects on adult resource allocation and fitness. *Oecologia* 144, 353–361.
- Bonifácio L, 2009. Impacte e evolução da doença da murchidão dos pinheiros (pine Wilt Disease) na zona afectada a sul do rio Tejo. PhD thesis, Science Faculty of Lisbon University: Lisboa.
- Braasch H,ENZIAN S and al., 2004. The pine wood nematode problem in Europe - present situation and outlook. In The pine wood nematode, *Bursaphelenchus xylophilus*. Nematology Monographs and Perspectives, 77-91. In M. Mota and P. Vieira(eds), The pinewood nematode, *Bursaphelenchus Xylophilus* : proceedings of an international workshop, University of Évora, Portugal, August 20-22, 2001, Vol 1; Nematology monographs and perspectives, Boston: Brill, 2004, pp. 291.
- Broquet T, Johnson CA, Petit E, Thompson I, Burel F et al., 2006. Dispersal and genetic structure in the American marten, *Martes americana*. *Molecular Ecology* 15, 1689–1697.
- Broquet T, Petit EJ, 2009. Molecular estimation of dispersal for ecology and population genetics. *Annual Review of Ecology, Evolution and Systematics* 40, 193–216.
- Bucci G, Gonzalez-Martinez SC, Le Provost G, Plomion C, Ribeiro MM, 2007. Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Molecular Ecology* 16, 2137–2153.
- Carrasco LR, Mumford JD, MacLeod A, Harwood T, Grabenweger G et al., 2010. Unveiling human-assisted dispersal mechanisms in invasive alien insects: Integration of spatial stochastic simulation and phenology models. *Ecological Modelling* 221, 2068–2075.
- Castillo JA, Epps CW, Davis AR, Cushman SA, 2014. Landscape effects on gene flow for a climate-sensitive montane species, the American pika. *Molecular Ecology* 23, 843–856.
- Cesari, M., Marescalchi, O., Francardi, V., Mantovani, B., 2005. Taxonomy and phylogeny of European *Monochamus* species: first molecular and karyological data. *Journal of Zoological Systematics and Evolutionary Research* 43, 1–7.
- Choisy M, Franck P, Cornuet JM, 2004. Estimating admixture proportions with microsatellites: comparison of methods based on simulated data. *Molecular Ecology* 13, 955–968.
- Ciosi M, Miller NJ, Kim KS, Giordano R, Estoup A 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.
- Corander J, Siren J, Arjas E, 2008. Bayesian spatial modeling of genetic population structure. *Computational Statistics* 23, 111–129.
- Courchamp F, Chapuis JL, Pascal L, 2003. Mammal invaders on islands: impact, control and control impact. *Biological Reviews* 78, 347–383.
- Cushman SA, Landguth E, 2010. Scale dependent inference in landscape genetics. *Landscape Ecology* 25, 967–979.
- Cushman SA, McKelvey KS, Hayden J, Schwartz MK, 2006. Geneflow in complex landscapes: testing multiple models with causal modeling. *American Naturalist* 168, 486–499.
- Cushman SA, Shirk AJ, Landguth E, 2013. Landscape genetics and limiting factors. *Conservation Genetics* 14, 263–274.
- Dakin EE, Avise JC, 2004. Microsatellite null alleles in parentage analysis. *Heredity* 93, 504–509.
- David G, 2014. Etude des capacités de dispersion de *Monochamus galloprovincialis* vecteur du nématode du pin *Bursaphelenchus xylophilus*. Thèse de doctorat, Université de Bordeaux.
- David G, Giffard B, Piou D, Jactel H, 2014. Dispersal capacity of *Monochamus galloprovincialis*, the European vector of the pine wood nematode, on flight mills. *Journal of Applied Entomology* 138, 566–576.
- David G, Giffard B, van Halder I, Piou D, Jactel H, 2015. Energy allocation during the maturation of adults in a long-lived insect: implications for dispersal and reproduction. *Bulletin of Entomological Research*, Online first. Doi:10.1017/S0007485315000553.

- Démolin G, 1969. Comportement des adultes de *Thaumetopoea pityocampa* Schiff. Dispersion spatiale, importance écologique. *Annales des Sciences Forestières* 26, 89–102.
- Dreier S, Redhead JW, Warren IA, Bourke AF, Heard MS et al., 2014. Fine-scale spatial genetic structure of common and declining bumble bees across an agricultural landscape. *Molecular Ecology* 23, 3384–3395.
- Ducan RP, Blackburn TM, Sol D, 2003. The ecology of bird introductions. *Annual Review of Ecology and Systematics* 34, 71–98.
- Edwards JS, 2006. The central nervous control of insect flight. *Journal of Experimental Biology* 209, 4411–4413.
- Ellstrand NC, Elam DR, 1993. Population genetic consequence of small population-size - implication for plant conservation. *Annual Review of Ecology and Systematics* 24, 217–242.
- EPPO, 2014 EPPO Reporting Service. No. 02, Paris 2014-02-1. Paris FRA: European and Mediterranean Plant protection Organization, pp. 18.
<http://archives.eppo.int/EPPORreporting/2014/Rse-1402.pdf> [consulté le 25/09/2015].
- EPPO, 2010. Reporting Service. No. 10, Paris 2012-10-1. Paris FRA: European and Mediterranean Plant protection Organization, pp. 19.
<http://archives.eppo.int/EPPORreporting/2012/Rse-1210.pdf>. [consulté le 25/09/2015].
- Erber D, Franquinho Aguiar AM, 1996. New and remarkable species of the coleopterous fauna of Madeira. *Boletim do Museu Municipal do Funchal (Historia Natural)* 48, 41–62.
- Eschen R, Roques A, Santini A, 2015. Taxonomic dissimilarity in patterns of interception and establishment of alien arthropods, nematodes and pathogens affecting woody plants in Europe. *Diversity and Distributions* 21, 36–45.
- Etxebeste I, Alvarez G, Sanchez-Husillos E, Pajares JA, 2013. *Monochamus galloprovincialis* dispersal and the effective sampling area of operational traps. *Berichte aus dem Julius-Kuehn-Institut*, 169, 18-19. In *Pine Wilt Disease Conference 2013*. Schröder, T. (ed.), *Scientific Conference IUFRO*, unit 7.02.10 and FP7 EU-Research Project REPHRAME. 15th to 18th Oct. 2013, Braunschweig GER: Berichte aus dem Julius Kühn-Institut.
- 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.
- Evans HF, 2014. Rapport final du projet Européen REPHRAME, pp. 57.
<http://www.rephrame.eu/pwn.php>.
- Evans HF, McNamara DG, Braasch H, Chadoeuf J, Magnusson C, 1996. Pest risk analysis (PRA) for the territories of the European Union (as PRA area) on *Bursaphelenchus xylophilus* and its vectors in the genus *Monochamus*. *Bulletin OEPP* 26, 199–249.
- Fady B, 2012. Biogeography of neutral genes and recent evolutionary history of pines in the mediterranean basin. *Annals of Forest Science* 69, 421–428.
- Faria JMS, Barbosa P, Bennett RN, Mota M, Figueiredo AC, 2013. Bioactivity against *Bursaphelenchus xylophilus*: Nematotoxics from essential oils, essential oils fractions and decoction waters. *Phytochemistry* 94, 220–228.
- Farr DF, Rossman AY, Palm ME, McCray, EB, 2004. Fungus-host distributions, fungal databases, systematic botany and mycology lab., ARS/USDA. <http://nt.ars-grin.gov/fungal databases/>.
- Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484, 186–194.
- Fonseca L, Cardoso JMS, Lopes A, Pestana M, Abreu F et al., 2012. The pinewood nematode, *Bursaphelenchus xylophilus*, in Madeira Island. *Helminthologia* 49, 96–103.
- Francke-Grosmann H, 1967. Ectosymbiosis in wood-inhabiting insects, 141-205. In S. M. Henry (ed.), *Symbiosis volume II: associations of invertebrates, birds, ruminants, and other biota*. New York USA: Academic press, pp. 461.

- Futai K, 2013. Pine Wood Nematode, *Bursaphelenchus xylophilus*. *Annual Review of Phytopathology* 51, 61–83.
- Gallego D, Sanchez-Garcia FJ, Mas H, Campo MT, Lencina JL, 2012. Estudio de la capacidad de vuelo a larga distancia de *Monochamus galloprovincialis* (Oliver 1795) (Coleoptera: Cerambycidae) en un mosaico agro-forestal. *Boletín de Sanidad Vegetal Plagas* 38, 109–123.
- Galpern P, Manseau M, Wilson P, 2012. Grains of connectivity: analysis at multiple spatial scales in landscape genetics. *Molecular Ecology* 21, 3996–4009.
- Galtier N, Nabholz B, Glémin S, Hurst GDD, 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology* 18, 4541–4550.
- Gandhi KJK, Gilmore DW, Katovich SA, Mattson WJ, Spence JR, 2007. Physical effects of weather disturbances on the abundance and diversity of forest insects in North American forests. *Environmental Reviews* 15, 113–152.
- Gandhi KJK, Herms DA, 2010. Direct and indirect effects of alien insect herbivores on ecological processes and interactions in forests of eastern North America. *Biological Invasions* 12, 389–405.
- Garbelotto M, Pautasso M, 2012. Impacts of exotic forest pathogens on Mediterranean ecosystems: four case studies. *European Journal of Plant Pathology* 133, 101–116.
- Guillot G, Rousset F, 2013. Dismantling the Mantel tests. *Methods in Ecology and Evolution* 4, 336–344.
- Guralnick R, 2007. Differential effects of past climate warming on mountain and flatland species distributions: a multispecies North American mammal assessment. *Global Ecology and Biogeography* 16, 14–23.
- Gurevitch J, Fox GA, Wardle GM, Inderjit, Taub D, 2011. Emergent insights from the synthesis of conceptual frameworks for biological invasions. *Ecology Letters* 14, 407–418.
- Hanks LM, 1999. Influence of the larval host plant on reproductive strategies of cerambycid beetles. *Annual Review of Entomology* 44, 483–505.
- Haran J, Roux-Morabito G, 2014. Development of 12 microsatellites loci for the longhorn beetle *Monochamus galloprovincialis* (Coleoptera Cerambycidae), vector of the pinewood nematode in Europe. *Conservation Genetics Resources* 6, 975–977.
- Haran J, Koutroumpa F, Magnoux E, Roques A, Roux G, 2015. Ghost mtDNA haplotypes generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and phylogeographic pattern. *Journal of Zoological Systematics and Evolutionary Research*. 53, 109–115.
- Haran J, Roques A, Bernard A, Robinet C, Roux G, 2015. Altitudinal barrier to the spread of an invasive species: could the pyrenean chain slow the natural spread of the pinewood nematode? *PLoS ONE* 10(7): e0134126.
- Hedin J, Ranius T, 2002. Using radio telemetry to study dispersal of the beetle *Osmoderma eremita*, an inhabitant of tree hollows. *Computers and Electronics in Agriculture* 35, 171–180.
- Hellrigl KG, 1971. La bionomie des espèces de *Monochamus* (Coleoptera, Cerambycidae) et leur importance pour la sylviculture et l'économie du bois. *Redia* 52, 367–511.
- Hernandez R, Ortiz A, Pérez V, Gil JM, Sanchez G, 2011. *Monochamus galloprovincialis* (Olivier, 1795) (Coleoptera: Cerambycidae), comportamiento y distancias de vuelo. *Boletín de Sanidad Vegetal Plagas* 37, 79–96.
- He SY, Ge XZ, Wang T, Wen JB, Zong SX, 2015. Areas of potential suitability and survival of *Dendroctonus valens* in china under extreme climate warming scenario. *Bulletin of Entomological Research* 105, 477–84.
- Hewitt GM, 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405, 907–913.
- Hewitt GM, 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 68, 87–112.

- Himler AG, Adachi-Hagimori T, Bergen JE, Kozuch A, Kelly SE et al., 2011. Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias. *Science* 332, 254–256.
- Hlaing T, Tun-Lin W, Somboon P, Socheat D, Setha T, et al., 2009. Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. *BMC Genetics* 10, pp. 12.
- Hofstetter RW, Klepzig KD, Moser JC, Ayres MP, 2006. Seasonal dynamics of mites and fungi and their interaction with Southern Pine Beetle. *Environmental Entomology* 35, 22–30.
- Huang R, Shi J, Luo Y, 2013. Cold-tolerance and adaption of pinewood nematode in China, *Berichte aus dem Julius-Kuehn-Institut*, 169: 84-85. In *Pine Wilt Disease Conference 2013*. Schröder, T. (ed.), *Scientific Conference IUFRO*, unit 7.02.10 and FP7 EU-Research Project REPHRAME. 15th to 18th Oct. 2013, Braunschweig GER: Berichte aus dem Julius Kühn-Institut.
- Hughes J, Dorn S, 2002. Sexual differences in the flight performance of the oriental fruit moth, *Cydia molesta*. *Entomologia Experimentalis et Applicata* 103, 171–182.
- Hu S-j, Ning T, Fu D-y, Haack RA, Zhang Z, et al, 2013. Dispersal of the Japanese pine sawyer, *Monochamus alternatus* (Coleoptera: Cerambycidae), in mainland China as inferred from molecular data and associations to indices of human activity. *PLoS ONE* (8)2: e57568.
- Ibeas F, Gallego D, Diez JJ, Pajares JA, 2007. An operative kairomonal lure for managing pine sawyer beetle *Monochamus galloprovincialis* (Coleoptera: Cerymbyidae). *Journal of Applied Entomology* 131, 13–20.
- IPCC, 2014: *Climate Change 2014: synthesis report*. Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change [Core writing team, R.K. Pachauri and L.A. Meyer (eds.)]. Geneva, Switzerland : IPCC, pp.151.
- Jactel H, Castagnone P, Mota M, Piou D, Robinet C et al., 2015. *Demande d’avis sur la stratégie de lutte imposée par la décision d’exécution 2012/535/UE du 26 septembre 2012 relative à la mise en place de mesures d’urgence destinées à prévenir la propagation dans l’Union Européenne de Bursaphelenchus xylophilus*. Saisine 2014-SA-0103 nématode du pin, Rapport d’expertise collective. ANSES, 51 pp.
- James R, Tisserat N, Todd T, 2006. Prevention of pine wilt of scots pine (*Pinus sylvestris*) with systemic abamectin injections. *Arboriculture and Urban Forestry* 32, 195–201.
- Jamieson MA, Trowbridge AM, Raffa KF, Lindroth RL, 2012. Consequences of climate warming and altered precipitation patterns for plant-insect and multitrophic interactions. *Plants Physiology* 160, 1719–1727.
- Janes JK, Roe AD, Rice AV, Gorrell JC, Coltman DW et al., 2015. Polygamy and an absence of fine-scale structure in *Dendroctonus ponderosae* (Hopk.) (Coleoptera: Curculionidae) confirmed using molecular markers. *Heredity* 1–7, doi:10.1038/hdy.2015.71.
- Jaquiéry J, Broquet T, Hirzel AH, Yearsley J, Perrin N, 2011. Inferring landscape effects on dispersal from genetic distances: how far can we go? *Molecular Ecology* 20, 692–705.
- Jiu M, Zhou X-P, Tong L, Xu J, Yang X, et al, 2007. Vector-virus mutualism accelerates population increase of an invasive whitefly. *PLoS ONE* 2(1): e182.
- Jones AG, Ardren WR, 2003. Methods of parentage analysis in natural populations. *Molecular Ecology* 12, 2511–2523.
- Kalinowski S, 2005. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5, 187–189.
- Karnosky DF, 1979. Dutch elm disease: A review of the history, environmental implications, control, and research needs. *Environmental Conservation* 6, 311–322.

- Katsuyama N, Sakurai, Tabata K, Takeda S, 1989. Effect of age of post-feeding twig on the ovarian development of Japanese pine sawyer, *Monochamus alternatus*. *Research Bulletin of the Faculty of Agriculture, Gifu University* 54, 81–89.
- Kawai M, Shoda-Kagaya E, Maehara T, Zhou ZH, Lian CL et al., 2006. Genetic structure of pine sawyer *Monochamus alternatus* (Coleoptera: Cerambycidae) populations in northeast Asia: Consequences of the spread of pine wilt disease. *Environmental Entomology* 35, 569–579.
- Keller D, Holderegger R, van Strien MJ, 2013. Spatial scale affects landscape genetic analysis of a wetland grasshopper. *Molecular Ecology* 22, 2467–2482.
- Klein EK, Lavigne C, Gouyon PH, 2006. Mixing of propagules from discrete sources at long distance: comparing a dispersal tail to an exponential. *BMC Ecology* 6, pp. 12.
- Klepzig KD, Adams AS, Handelsman J, Raffa KF, 2009. Symbioses: a key driver of insect physiological processes, ecological interactions, evolutionary diversification, and impacts on humans. *Environmental Entomology* 38, 67–77.
- Koenig WD, van Vuren D, Hooge PN, 1996. Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in Ecology and Evolution* 11, 514–517.
- Kolar CS, Lodge DM, 2001. Progress in invasion biology: predicting invaders. *Trends in Ecology and Evolution* 16, 199–204.
- Kong WN, Wang H, Li J, Zaho F, Ma RY, 2006. The effects of temperature and humidity on the longevity of *Monochamus alternatus* Hope (Coleoptera: Cerambycidae). *Journal of Shanxi Agricultural University* 294–295 [in Chinese].
- Koutroumpa FA, Lieutier F, Roux-Morabito G, 2009. Incorporation of mitochondrial fragments in the nuclear genome (Numts) of the longhorned beetle *Monochamus galloprovincialis* (Coleoptera, Cerambycidae). *Journal of Zoological Systematics and Evolutionary Research* 47, 141–148.
- Koutroumpa FA, Rougon D, Bertheau C, Lieutier F, Roux-Morabito G, 2013. Evolutionary relationships within European *Monochamus* (Coleoptera: Cerambycidae) highlight the role of altitude in species delineation. *Biological Journal of the Linnean Society of London* 109, 354–376.
- Koutroumpa FA, Vincent B, Roux-Morabito G, Martin C, Lieutier F, 2008. Fecundity and larval development of *Monochamus galloprovincialis* (Coleoptera Cerambycidae) in experimental breeding. *Annals of Forest Science* 65, 707.
- Kuroda K, 2008. Physiological incidences related to symptom development and wilting mechanism, pp. 204–222. In: Zhao, B.G., Futai, K., Sutherland, J.R., Takeuchi, Y. (Eds.), *Pine Wilt Disease*. Tokyo, JAP: Springer Japan.
- Kutywayo D, Chemura A, Kusena W, Chidoko P, Mahoya C, 2013. The impact of climate change on the potential distribution of agricultural pests: the case of the coffee white stem borer (*Monochamus leuconotus* P.) in Zimbabwe. *PLoS ONE* 8(8): e73432.
- Langella O, Chicki L, Beaumont MA, 2001. LEA (likelihood-based estimation of admixture): a program to estimate simultaneously admixture and time since the admixture event. *Molecular Ecology Notes* 1, 357–358.
- La Porta N, Capretti P, Thomsen IM, Kasanen R, Hietala AM, Von Weissenberg K, 2008. Forest pathogens with higher damage potential due to climate change in Europe. *Canadian Journal of Plant Pathology* 30, 177–195.
- Lawson SA, Sathyapala S, 2008. The Risk of Pine Wilt Disease to Australia and New Zealand. 3: 41–58. In: Mota M, Vieira.P.R. (eds), *pine wilt disease: A worldwide threat to forest ecosystems*, pp. 405.
- Liebholt AM, Tobin PC, 2008. Population ecology of insect invasions and their management. *Annual Review of Entomology* 53, 387–408.

- Linit MJ, 1988. Nematode-vector relationships in the pine wilt disease system. *Journal of Nematology* 20, 227–235.
- Linit MJ, Akbulut S, 2001. Pine wood nematode phoresis: the impact on *Monochamus carolinensis* life functions. pp. 227–237 In: M. Mota and P. Vieira (eds), *the pinewood Nematode, Bursaphelenchus Xylophilus. Proceedings of an International Workshop*, University of Évora, Portugal, 20-22 August 2001. Brill Academic Publishers, pp. 291.
- Lombaert E, Guillemaud T, Cornuet JM, Malausa T, Facon B et al., 2010. Bridgehead effect in the worldwide invasion of the biocontrol harlequin ladybird. *PLoS ONE* 5(3): e9743.
- Lopez-Merino L, Martinez Cortizas A, Reher GS, Lopez-Saez JA, Mighall TM, Bindler R, 2014. Reconstructing the impact of human activities in a NW Iberian Roman mining landscape for the last 2500 years. *Journal of Archaeological Science* 50, 208–218.
- Lu M, Wingfield MJ, Gillette NE, Mori SR, Sun JH, 2010. Complex interactions among host pines and fungi vectored by an invasive bark beetle. *New Phytologist* 187, 859–866.
- Mack RN, Simberloff D, Lonsdale WM et al., 2000. Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications* 10, 689–710.
- Mallez S, 2014. Invasion du nématode du pin, *Bursaphelenchus xylophilus*. Thèse de doctorat, Université de Nice Sophia-Antipolis, pp. 182.
- Mallez S, Castagnone C, Espada M, Vieira P, Eisenback JD, 2015. Worldwide invasion routes of the pinewood nematode: What can we infer from population genetics analyses? *Biological Invasions* 17, 1199–1213.
- Mamiya Y, 2004. Pine wilt disease in Japan. 9-20 pp. In: M. Mota and P. Vieira (eds), *the Pinewood Nematode, Bursaphelenchus Xylophilus*. Proceedings of an International Workshop, University of Évora, Portugal, 20-22 August 2001. Brill Academic Publishers, pp. 291.
- Mamiya Y, 1988. History of pine wilt disease in Japan. *Journal of Nematology* 20, 219–226.
- Mamiya Y, 1983. Pathology of the Pine Wilt Disease caused by *Bursaphelenchus xylophilus*. *Annual Review of Phytopathology* 21, 201–20.
- Mamiya Y, 1975. The life history of the pine wood nematode. *Japanese Journal of Nematology* 5, 16–25.
- Mamiya Y, Eda N, 1972. Transmission of *Bursaphelenchus lignicolus* (Nematoda : Aphelenchoididae) by *Monochamus alternatus* (Coleoptera : Cerambycidae). *Nematologia* 18, 159–162.
- Manel S, Gaggiotti OE, Waples RS, 2005. Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution* 20, 136–142.
- Manel S., Holderegger R, 2013. Ten years of landscape genetics. *Trends in Ecology and Evolution*, 28, 614–621.
- Manel S, Schwartz MK, Luikart G, Taberlet P, 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology and Evolution* 18, 189–197.
- Mas H, Hernandez R, Villaroya G, Sanchez G, Pérez-Laorga E et al., 2013. Dispersal behavior and long distance flight capacity of *Monochamus galloprovincialis* (Olivier 1795). *Berichte aus dem Julius-Kuehn-Institut*, 169: 22. In *Pine Wilt Disease Conference 2013*. Schröder, T. (ed.), *Scientific Conference IUFRO*, unit 7.02.10 and FP7 EU-Research Project REPHRAME. 15th to 18th Oct. 2013, Braunschweig GER: Berichte aus dem Julius Kühn-Institut.
- Mayer CJ, Vilcinskis A, Gross J, 2008. Pathogen-induced release of plant allomone manipulates vector insect behavior. *Journal of Chemical Ecology* 34, 1518–1522.
- Mayer F, Björklund N, Wallén J, Långström B, Cassel-Lundhagen A, 2014. Mitochondrial DNA haplotypes indicate two postglacial re-colonization routes of the spruce bark beetle *Ips typographus* through northern Europe to Scandinavia. *Journal of Zoological Systematics and Evolutionary Research* 52, 285–292.

- McKinney ML, Lockwood JL, 1999. Biotic homogenization: a few winners replacing many losers in the next mass extinction. *Trends in Ecology and Evolution* 14, 450–453.
- McLeod G, Gries R, Reuss ST, Rahe JE, McIntosh R et al., 2005. The pathogen causing Dutch Elm Disease makes host trees attract insect vectors. *Proceedings of the Royal Society B* 272, 2499–2503.
- McRae B, 2006. Isolation by resistance. *Evolution* 60, 1551–1561.
- Melbourne BA, Hastings A, 2009. Highly Variable spread rates in replicated biological invasions: fundamental limits to predictability. *Science* 325, 1536–1539.
- Mendes AMC, Feliciano D, Tavares M, Dias R, 2004. The Portuguese Forests. Working Papers de Economia (Economics Working Papers) 13, Faculdade de Economia e Gestão, Universidade Católica Portuguesa (Porto), pp. 85.
- Mille C, Cazères S, Sudre J, 2015. Interception de *Monochamus galloprovincialis* (Olivier, 1795) (Cerambycidae, Lamiinae, Monochamini) en Nouvelle-Calédonie. *Bulletin Mensuel de la Société Linnéenne de Lyon* 84, 55–58.
- Millspaugh JJ, Marzluff JM (Eds.) 2001. Radio tracking and animal populations. San Diego, US: Academic Press, pp. 474.
- Mooney HA, Cleland EE, 2001. The evolutionary impact of invasive species. *Proceedings of the National Academy of Sciences of the USA* 98, 5446–5451.
- Mota M., Braasch H., Bravo M.A., Penas A.C., Burgermeister W., et al., 1999. First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. *Nematology* 1, 727–734.
- Murphy MA, Evans JS, Cushman SA, Storfer A, 2008. Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. *Ecography* 31, 685–697.
- Nakamura K, Maehara N, Aikawa T, Ichihara Y, Kosaka H, 2013. Research project to develop strategic action plan in the Pine-wilt-disease unaffected area in northern Japan. *Berichte aus dem Julius-Kuehn-Institut*, 169: 66. In *Pine Wilt Disease Conference 2013*. Schröder, T. (ed.), *Scientific Conference IUFRO*, unit 7.02.10 and FP7 EU-Research Project REPHRAME. 15th to 18th Oct. 2013, Braunschweig GER: Berichte aus dem Julius Kühn-Institut.
- Naves P, de Sousa E, 2009. Threshold temperatures and degree-day estimates for development of post-dormancy larvae of *Monochamus galloprovincialis* (Coleoptera: Cerambycidae). *Journal of Pest Science* 82, 1–6.
- Naves P, De Sousa EM, Quartau JA, 2007b. Winter dormancy of the pine sawyer *Monochamus galloprovincialis* (Col., Cerambycidae) in Portugal. *Journal of Applied Entomology* 131, 669–673.
- Naves PM, Camacho S, de Sousa E, Quartau JA, 2007. Transmission of the pine wood nematode *Bursaphelenchus xylophilus* through oviposition activity of *Monochamus galloprovincialis* (Coleoptera : Cerambycidae). *Entomologica Fennica* 18, 193–198.
- Naves P, Sousa E, Quartau J, 2006. Feeding and oviposition preferences of *Monochamus galloprovincialis* for some conifers under laboratory conditions. *Entomologica Experimentalis et Applicata* 120, 99–104.
- Naves P, Vieira M, Sousa E, 2013. New Strategies for pine wilt disease (PWD) management in Portugal: preventive methods to reduce the spread of the disease to new areas. *Berichte aus dem Julius-Kuehn-Institut*, 169: 68. In *Pine Wilt Disease Conference 2013*. Schröder, T. (ed.), *Scientific Conference IUFRO*, unit 7.02.10 and FP7 EU-Research Project REPHRAME. 15th to 18th Oct. 2013, Braunschweig GER: Berichte aus dem Julius Kühn-Institut.
- Olden JD, LeRoy Poff N, Douglas MR, Douglas ME, Fausch KD, 2004. Ecological and evolutionary consequences of biotic homogenization. *Trends in Ecology and Evolution* 19, 18–24.

- Paetkau D, Slade R, Burden M, Estoup A, 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* 13, 55–65.
- Paupy C, Delatte H, Bagny L, Corbel V, Fontenille D, 2009. *Aedes albopictus*, an arbovirus vector: from the darkness to the light. *Microbes and Infection* 11, 1177–1185.
- Peakall R, Smouse PE, 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28, 2537–2539.
- Petersen-Silva R, Naves P, Godinho P, Sousa E, Pujade-Villar J, 2014. Distribution, hosts and parasitoids of *Monochamus galloprovincialis* (Coleoptera: Cerambycidae) in Portugal mainland. *Silva Lusitana* 22, 67–82.
- Pflüger FJ, Balkenhol N, 2014. A plea for simultaneously considering matrix quality and local environmental conditions when analysing landscape impacts on effective dispersal. *Molecular Ecology* 23, 2146–2156.
- PHRAME, Plant health risk and monitoring evaluation. <http://www.forestry.gov.uk/fr/phrame>.
- Pimentel D, McNair S, Janecka J, et al., 2001. Economic and environmental threats of alien plant, animal, and microbe invasions. *Agriculture Ecosystems and Environment* 84, 1–20.
- Piry S, Luikart G, Cornuet JM, 1999. BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* 90, 502–503.
- Pitt JPW, Worner SP, Suarez AV, 2009. Predicting Argentine ant spread over the heterogeneous landscape using a spatially explicit stochastic model. *Ecological Applications* 19, 1176–1186.
- Pluess et al., 2012a. When are eradication campaigns successful? A test of common assumptions. *Biological Invasions* 14, 1365–1378.
- Pluess et al., 2012b. Which factors affect the success or failure of eradication campaigns against alien species? *PLoS ONE* (7)10: e48157.
- Potter C, Harwood T, Knight J, Tomlinson I, 2011. Learning from history, predicting the future: the UK Dutch Elm Disease outbreak in relation to contemporary tree disease threats. *Philosophical Transactions of the Royal Society B* 366, 1966–1974.
- Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Prunier JG, Colyn M, Legendre X, Nimon KF, Flamand MC, 2014. Multicollinearity in spatial genetics: separating the wheat from the chaff using commonality analyses. *Molecular Ecology* 24, 263–283.
- Pysek P, Sadlo J, Mandak B, Jarosik V, 2003. Czech alien flora and the historical pattern of its formation: what came first to Central Europe? *Oecologia* 135, 122–130.
- Ranius T, 2006. Measuring the dispersal of saproxylic insects: a key characteristic for their conservation. *Population Ecology* 48, 177–188.
- Rannala B, Mountain JL, 1997. Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the USA* 94, 9197–9201.
- REPHRAME, Development of improved methods for detection, control and eradication of pine wood nematode in support of EU Plant Health Policy. <http://www.rephrame.eu/>.
- Rhymer JM, Simberloff D, 1996. Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* 27, 93–109.
- Ribeiro B, Espada M, Vu T, Nobrega F, Carrasquinho I, 2012. Pine wilt disease: detection of the pinewood nematode (*Bursaphelenchus xylophilus*) as a tool for a pine breeding programme. *Forest Pathology* 42, 521–525.

- Richardson DM, Pyšek P, Carlton JT, 2011. A compendium of essential concepts and terminology in invasion ecology. 409–420. In: Richardson, D.M. (ed.), *Fifty years of invasion ecology. The legacy of Charles Elton*. Oxford, ENG: Wiley-Blackwell, pp. 432.
- Robertson L, Cobacho AS, Escuer M, Esparrago G et al., 2011. Incidence of the pinewood nematode *Bursaphelenchus xylophilus* Steiner & Buhrer, 1934 (Nickle, 1970) in Spain. *Nematology* 13, 755–757.
- Robinet C, Baier P, Pennerstorfer J, Schopf A, Roques A, 2007. Modelling the effects of climate change on the potential feeding activity of *Thaumetopoea pityocampa* (Den. & Schiff.) (Lep., Notodontidae) in France. *Global Ecology and Biogeography* 16, 460–471.
- Robinet C, Imbert CE, Rousselet J, Sauvard D, Garcia J et al., 2012. Human-mediated long-distance jumps of the pine processionary moth in Europe. *Biological Invasions* 14, 1557–1569.
- Robinet C, Kehlenbeck H, Kriticos DJ, Baker RHA, Battisti A et al., 2012. A suite of models to support quantitative assessment of spread in pest risk analysis. *PLoS ONE* 7(10): e43366.
- Robinet C, Roques A, Pan H, Fang G, Ye J et al., 2009. Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS One* 4(2): e4646.
- Robinet C, Rousselet J, Roques A, 2014. Potential spread of the pine processionary moth in France: preliminary results from a simulation model and future challenges. *Annals of Forest Science* 71, 149–160.
- Robinet C, Van Opstal N, Baker R, Roques A, 2011. Applying a spread model to identify the entry points from which the pine wood nematode, the vector of pine wilt disease, would spread most rapidly across Europe. *Biological Invasions* 13, 2981–2995.
- Rodrigues JM, 2008. National eradication programme for the pinewood nematode. 3: 5-14. In: Mota M, Vieira P.R. (eds). *Pine wilt disease: A worldwide threat to forest ecosystems*, 405 pp.
- Roques A, Kenis M, Lees D, Lopez-Vaamonde C, Rabitsch W, Rasplus JY, Roy D (eds.), 2010. Alien terrestrial arthropods of Europe. *BioRisk*, special issue, 4. Sofia, BGR : Pensoft Publishers, pp. 1028.
- Roques A, Zhao LL, Sun J, Robinet C, 2015. Pine wood nematode, pine wilt disease, vector beetle and pine tree: how a multiplayer system could reply to climate change. 12:pp. In: Björkman C. & Niemelä P. (Eds), *Climate Change and Insect Pests*, CABI Climate Change Series, Oxfordshire, UK: CABI Editions, pp. 292.
- Roques L, Roques A, Berestycki H, Kretschmar A, 2008. A population facing climate change: joint influences of Allee effects and environmental boundary geometry. *Population Ecology* 50, 215–225.
- Rossi JP, Garcia J, Roques A, Rousselet J, 2015. Trees outside forests in agricultural landscapes: spatial distribution and impact on habitat connectivity for forest organisms. *Landscape Ecology*. pp. 12. doi: 10.1007/s10980-015-0239-8.
- Rousselet J, Zhao RX, Argal D, Simonato M, Battisti A et al., 2010. The role of topography in structuring the demographic history of the pine processionary moth, *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae). *Journal of Biogeography* 37, 1478–1490.
- Ruiz-Benito P, Gomez-Aparicio L, Zavala MA, 2012. Large-scale assessment of regeneration and diversity in Mediterranean planted pine forests along ecological gradients. *Diversity and Distributions* 18, 1092–1106.
- Rutherford TA, Mamiya Y, Webster JM, 1990. Nematode-induced pine wilt disease: factors influencing its occurrence and distribution. *Forest Science* 36, 145–155.
- Rutherford TA, Webster JM, 1987. Distribution of pine wilt disease with respect to temperature in North America, Japan and Europe. *Canadian Journal of Forest Research* 17, 1050–1059.
- Sakai AK, Allendorf FW, Holt JS, Lodge DM, Molofsky J et al., 2001. The population biology of invasive species. *Annual Review of Ecology and Systematics* 32, 305–332.

- Santini A, Ghelardini L, De Pace C, Desprez-Loustau ML, Capretti P et al., 2013. Biogeographical patterns and determinants of invasion by forest pathogens in Europe. *New Phytologist* 197, 238–250.
- Schmitt T, 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology* 4, pp. 13. Doi:10.1186/1742-9994-4-11.
- Schrey NM, Schrey AW, Heist EJ, Reeve JD, 2008. Fine-scale genetic population structure of southern pine beetle (Coleoptera : Curculionidae) in Mississippi forests. *Environmental Entomology* 37, 271–276.
- Senger SE, Roitberg BD, Thistlewood HMA, 2007. Relative flight responses of *Rhagoletis indifferens* as influenced by crowding, sex, and resources. *Entomologia Experimentalis et Applicata* 123, 91–100.
- Serra P, Vera A, Francesc Tulla A, Salvati L, 2014. Beyond urbanerural dichotomy: exploring socioeconomic and land-use processes of change in Spain (1991-2011). *Applied Geography* 55, 71–81.
- Sexton JP, Hangartner SB, Hoffmann AA, 2014. Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution* 68, 1–15.
- Shigesada N, Kawasaki K, 1997. *Biological invasions: theory and practice*, Oxford series in ecology and evolution. Oxford, UK: Oxford University Press, pp. 205.
- Shirk AJ, Wallin DO, Cushman SA, Rice CG, Warheit KI, 2010. Inferring landscape effects on gene flow: a new model selection framework. *Molecular Ecology* 19, 3603–3619.
- Shoda-Kagaya E, 2007. Genetic differentiation of the pine wilt disease vector *Monochamus alternatus* (Coleoptera: Cerambycidae) over a mountain range – revealed from microsatellite DNA markers. *Bulletin of Entomological Research* 97, 167–174.
- Shokralla S, Gibson JF, Nikbakht, Janzen DH, Hallwachs W et al., 2014. Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular Ecology Resources* 14, 892–901.
- Short Bull RA, Cushman SA, Mace R, Chilton T, Kendall KC, Landguth EL, et al., 2011. Why replication is important in landscape genetics: American black bear in the Rocky Mountains. *Molecular Ecology* 20, 1092–1107.
- Six DL, Bentz BJ, 2007. Temperature determines symbiont abundance in a multipartite bark beetle-fungus ectosymbiosis. *Microbial Ecology* 54, 112–118.
- Smith RM, Baker RHA, Malumphy CP, Hockland S, Hammon RP et al., 2007. Recent non-native invertebrate plant pest establishments in Great Britain: origins, pathways, and trends. *Agricultural and Forest Entomology* 9, 307–326.
- Soliman T, Mourits MCM, van der Werf W, Hengeveld GM, Robinet C et al., 2012. Framework for modelling economic impacts of invasive species, applied to pine wood nematode in Europe. *PLoS ONE* 7(9): e45505.
- Song, H., Buhay, J.E., Whiting, M.F., Crandall, K.A., 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the USA* 105, 13486–13491.
- Sousa E, Mota M, 2001. *Bursaphelenchus xylophilus* (Nematoda: Aphelenchoididae) associated with *Monochamus galloprovincialis* (Coleoptera: Cerambycidae) in Portugal. *Nematology* 3, 89–91.
- Sousa E, Naves P, Vieira M, 2013. Prevention of pine wilt disease induced by *Bursaphelenchus xylophilus* and *Monochamus galloprovincialis* by trunk injection of emamectin benzoate. *Phytoparasitica* 41, 143–148.
- Sousa E, Rodrigues J, Bonifacio L, Naves P, Rodrigues A, 2011. Management and control of the pine wood nematode, *Bursaphelenchus xylophilus*, in Portugal. 1: 21 pp. In: F. Boeri & J. Chung

- (Eds.), *Nematodes: Morphology, functions and management strategies*. New York, US: Nova Science publishers, pp. 327.
- Steinbach D, Gutrod O, Lummen P, Matthiesen S, Schom C et al., 2015. Geographic spread, genetics and functional characteristics of ryanodine receptor based target-site resistance to diamide insecticides in diamondback moth, *Plutella xylostella*. *Insect Biochemistry and Molecular Biology* 63, 14–22.
- Suckling DM, Tobin PC, McCullough DG, Herms DA, 2012. Combining tactics to exploit Allee effects for eradication of alien insect populations. *Journal of Economic Entomology* 105, 1–13.
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF, 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* 7, 453–464.
- Takai K, Suzuki T, Kawasu K, 2003. Development and preventative effect against pine wilt disease of a novel liquid formulation of emamectin benzoate. *Pest Management Science* 59, 365–370.
- Takasu F, Yamamoto N, Kawasaki K, Togashi K, Kishi Y, 2000. Modeling the expansion of an introduced tree disease. *Biological Invasions* 2, 141–150.
- Telfer S, Pierniey SB, Dallas JF, Stewart WA, Marshall F et al., 2003. Parentage assignment detects frequent and large-scale dispersal in water voles. *Molecular Ecology* 12, 1939–1949.
- Togashi K, 1990. A field experiment on dispersal of newly emerged adults of *Monochamus alternatus* (Coleoptera: Cerambycidae). *Research on Population Ecology* 32, 1–13.
- Togashi K, Shigesada N, 2006. Spread of the pinewood nematode vectored by the Japanese pine sawyer: modelling and analytical approaches. *Population Ecology* 48, 271–283.
- Tomminen J, 1993. Development of *Monochamus galloprovincialis* Olivier (Coleoptera, Cerambycidae) in cut trees of young pines (*Pinus sylvestris* L.) and log bolts in southern Finland. *Entomologica Fennica* 4, 137–142.
- Torres-Vila L, Zugasti C, De-Juan JM, Olivia MJ, Montero C, Mendiola FJ et al., 2015. Mark-recapture of *Monochamus galloprovincialis* with semiochemicalbaited traps: population density, attraction distance, flight behaviour and mass trapping efficiency. *Forestry* 88, 224–236.
- Vicente C, Espada M, Vieira P, Mota M, 2012. Pine Wilt Disease: a threat to European forestry. *European Journal of Plant Pathology* 133, 89–99.
- Vincent B, Koutroumpa F, Altemayer V, Roux-Morabito G, Gevar J et al., 2008. Occurrence of *Bursaphelenchus mucronatus* (Nematoda; Aphelenchoididae) in France and association with *Monochamus galloprovincialis* (Coleoptera : Cerambycidae). *Annals of Forest Science* 65, pp. 111.
- Vogt JT, Appel AG, S West MS, 2000. Flight energetics and dispersal capability of the fire ant, *Solenopsis invicta* Buren. *Journal of Insect Physiology* 46, 697–707.
- Walker BH, Steffen W, 1997. An overview of the implications of global change for natural and managed terrestrial ecosystems. *Conservation Ecology* 1, pp. 17.
- Wallin H, Schroeder M, Kvamme T, 2013. A review of the European species of *Monochamus* Dejean, 1821 (Coleoptera, Cerambycidae) – with a description of the genitalia characters. *Norwegian Journal of Entomology* 60, 11–38.
- Wang & Bradburd, 2014. Isolation by environment. *Molecular Ecology* 23, 5649–5662.
- Wang CY, Fang ZM, Wang Z, Zhang DL, Gu LJ et al, 2011. Biological control of the pinewood nematode *Bursaphelenchus xylophilus* by application of the endoparasitic fungus *Esteya vermicola*. *BioControl* 56, 91–100.
- Weir BS, Cockerham CC, 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370.
- Williamson M, Fitter A, 1996. The varying success of invaders. *Ecology* 77, 1661–1666.
- Wingfield MJ, Brockerhoff EG, Wingfield BD, Slippers B, 2015. Planted forest health: The need for a global strategy. *Science* 349, 832–836

- Yamamura K, Kishita M, Arakaki N, Kawamura F, Sadoyama Y, 2003. Estimation of dispersal distance by mark-recapture experiments using traps: correction of bias caused by the artificial removal by traps. *Population Ecology* 45, 149–155.
- Yamanaka T, Tatsuki S, Shimada M, 2001. Flight characteristics and dispersal patterns of fall webworm (Lepidoptera: Arctiidae) males. *Environmental Entomology* 30, 1150–1157.
- Yan ZL, Sun JH, Owen D, Zhang ZN, 2005. The red turpentine beetle, *Dendroctonus valens* LeConte (Scolytidae): an exotic invasive pest of pine in China. *Biodiversity and Conservation* 14, 1735–1760.
- Yoshimura A, Kawasaki K, Takasu F, Togashi K, Futai K, et al., 1999. Modeling the spread of pine wilt disease caused by nematodes with pine sawyers as vector. *Ecology* 80, 1691–1702.
- Yu M, Xu X, Ding P, 2011. Economic loss versus ecological gain: the outbreaks of invaded pinewood nematode in China. *Biological Invasions* 13, 1283–1290.
- Zeller KA, McGarigal K, Whiteley AR, 2012. Estimating landscape resistance to movement: a review. *Landscape Ecology* 27, 777–797.
- Zenni RD, Hoban SM, 2015. Loci under selection during multiple range expansions of an invasive plant are mostly population specific, but patterns are associated with climate. *Molecular Ecology* 24, 3360–3371.
- Zhao LL, Lu M, Niu HT, Fang GF, Zhang S et al., 2013. A native fungal symbiont facilitates the prevalence and development of an invasive pathogen–native vector symbiosis. *Ecology* 94, 2817–2826.
- Zhao LL, Zhang S, Wei W, Hao HJ, Zhang B, 2013. Chemical signals synchronize the life cycles of a plant-parasitic nematode and its vector beetle. *Current Biology* 23, 2038–2043.
- Zink MR, 2010. Drawbacks with the use of microsatellites in phylogeography: the song sparrow *Melospiza melodia* as a case study. *Journal of Avian Biology* 41, 1–7.

Crédits photographiques

Figure 4 (page 23) : Nématode du pin et symptômes de la maladie du flétrissement du pin.
clichés: Ian Riley / <http://dmgcc.blogspot.fr> / <http://www.rncan.gc.ca>

Figure 5 (page 25): *Monochamus galloprovincialis*, vecteur du nématode du pin en Europe.
Clichés : <http://www.uni-goettingen.de>; <http://www.cerambyx.uochb.cz>

Annexes

Annexe n°1

Haran J, Koutroumpa F, Magnoux E, Roques A & Roux G 2015. Ghost mtDNA haplotypes generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and phylogeographic pattern. *Journal of Zoological Systematics and Evolutionary Research* 53, 109-115.

¹INRA, UR633 Zoologie Forestière, F-45075, Orléans France; ²Université d'Orléans, Orléans France; ³INRA, UMR 1392, iEES-Paris, Versailles France

Ghost mtDNA haplotypes generated by fortuitous NUMTs can deeply disturb infra-specific genetic diversity and phylogeographic pattern

JULIEN HARAN^{1,2}, FOTINI KOUTROUMPA³, EMMANUELLE MAGNOUX¹, ALAIN ROQUES¹ and GÉRALDINE ROUX^{1,2}

Abstract

Nuclear copies of mitochondrial DNA (NUMTs or mitochondrial pseudogenes) are known to impede the detection of interspecific genetic diversity. But the effect of these artifacts on phylogeographic reconstruction remains under evaluated. In this study, we analysed a set of 115 sequences of a fragment of the cytochrome c oxidase subunit I gene (COI) of *Monochamus galloprovincialis* (Coleoptera, Cerambycidae) for which overlapping signals in sequencing electropherograms were observed. Comparison of full and corrected 'ambiguities-free' data sets reveals the prevalence of numerous supernumerary haplotypes that deeply affect genetic diversity indices and phylogeographic patterns of this species. Slightly divergent pseudogenes were recovered in 49 of the 115 sequences. These results highlight the potential misdetection of NUMTs using current control methods and the consequences on phylogeographic structure. To test the frequency of unintended amplification of NUMTs, a cloning was performed on 15 individuals. An average of 3.72 and a maximum of six paralogous sequences with different levels of divergence were identified among individual cloned. Within individual pairwise distance between paralogs raised 1.4%. This work calls for awareness to the presence of undetected NUMTs within mitochondrial data sets, especially at infra-specific level.

Key words: mtDNA – NUMTs – COI – *Monochamus galloprovincialis* – phylogeography – intra-individual polymorphism

Introduction

Over the last decades, due to a shorter coalescent time compared to nuclear loci (Neigel and Avise 1986; Hudson 1991), mitochondrial DNA has been extensively used for shallow scale phylogenetics and phylogeographies (Galtier et al. 2009; Hickerson et al. 2010). The success of this genome as genetic marker is due to numerous striking characteristics, such as a small size, a relatively high rate of polymorphism and a simple and conserved structure providing sequences for universal primer design (Lunt et al. 1996). Recent developments in high throughput sequencing technologies, including the full mitogenome analysis, have allowed a deeper characterization of the mitochondrial genome (Cameron 2014), providing new prospects on mtDNA evolutionary processes (Song et al. 2010; Haran et al. 2013). Moreover, DNA-based species identification still required thorough taxonomic reference libraries of standardized regions, including the standard mtDNA barcode for the identification of animals (Hebert et al. 2003; Shokralla et al. 2014). Nevertheless, in recent years, there was a growing awareness of pitfalls and limits encountered with this genome.

Among these, undetected intra-individual variation generated by mitochondrial heteroplasmy, gene duplication events (Bensasson et al. 2000, 2001) or integration of mtDNA into the nuclear genome (nuclear mitochondrial pseudogenes or NUMTs, Lopez et al. 1994; Zhang and Hewitt 1996) can profoundly confound phylogenetic studies. Even if nuclear copies of mitochondrial sequences received considerable attention, the effects remained poorly reported in the literature when closely related species are studied (but see Ruiz et al. 2013), or when infra-specific level was investigated (Sword et al. 2007). We believe that the occurrence of NUMTs was generally underestimated, because of fortuitous amplification and random occurrence of NUMTs confounded with the putative mtDNA fragment. NUMTs are an ubiquitous phenomenon (review in Zhang and Hewitt 1996; Aylliffe et al. 1998; Bensasson et al. 2001; Richly and Leiste 2004; Leister 2005) and has been recorded in a wide range of organisms including plants, fungi, vertebrates and invertebrates among which insects have provided numerous examples (Bensasson

et al. 2000; Rokas et al. 2003; Pons and Vogler 2005; Keller et al. 2007; Black IV and Bernhardt 2009; Hlaing et al. 2009; Koutroumpa et al. 2009; Ruiz et al. 2013). Whether NUMTs are preferentially amplified instead of the authentic mitochondrial copy (Gjerde 2013) or co-amplified but not identified, they generate confusion of identity among the putative mtDNA and its paralogs. Because the fundamental principle of orthology is no longer respected (Calvignac et al. 2011), they result in erroneous interpretation of phylogenetic patterns. Moreover, this phenomenon leads to an overestimation of genetic diversity (Song et al. 2008), spilling out ghost haplotypes disturbing the phylogeographic reconstruction at infra-specific level (Cristiano et al. 2012). Several methods have been proposed to detect or avoid the presence of NUMTs, in particular for barcoding approaches (see Bensasson et al. 2001; Song et al. 2008; Buhay 2009; Calvignac et al. 2011). But these methods, especially *a posteriori* analysis (i.e. quality control of the sequence alignment), face strong limitations to detect slightly divergent paralogous sequences encountered within species (Hlaing et al. 2009; Koutroumpa et al. 2009; Bertheau et al. 2011), with severe consequences on the evolutionary processes highlighted.

A previous work brought out the presence of NUMTs in the mitochondrial genes cytochrome c oxidase subunit I and II (COI and COII) in the longhorn beetle *Monochamus galloprovincialis* Olivier (Coleoptera, Cerambycidae) (Koutroumpa et al. 2009). The nuclear origin of co-amplified sequences was confirmed with respect to intra-individual mitochondrial heteroplasmy using alkaline lysis. The occurrence of NUMTs seemed species specific and appeared not overlapping with mtDNA sequences of its sister species *M. sutor*, despite strong effects on the delineation of intra-specific lineages reported in within *M. galloprovincialis* (Koutroumpa et al. 2009, 2013). *Monochamus galloprovincialis* is the European vector of the invasive species *Bursaphelenchus xylophilus* (Steiner & Burher, 1934) Nickle, 1970 (Nematoda: Aphelenchoididae) recently introduced in Portugal (Mota et al. 1999) and Spain (Robertson et al. 2011). This nematode is the agent of the pine wilt disease, causing considerable damages to native pine forests. Evaluation of the genetic structure of *M. galloprovincialis* at large and local scale is a prerequisite to assess gene flow of this species and can provide elements to predict invasion route of the Pine Wood Nematode (PWN) in Europe.

Corresponding author: Géraldine Roux (geraldine.roux@orleans.inra.fr)

The presence of NUMTs in *M. galloprovincialis*, combined to the key role of this species in the natural dispersal of the PWN in Europe, make it of particular interest to evaluate the extent and the effect of these pitfalls.

Our goal was to highlight the bias encountered in reconstructing phylogeography using mitochondrial sequences affected by NUMTs. We show in this study (1) the effect of amplification of fortuitous NUMTs on genetic diversity and phylogeographic pattern and (2) the high intra-individual variation of co-amplified paralogous sequences.

Materials and Methods

Mitochondrial sequences and cloning

Mitochondrial sequences corresponding to a 792-bp fragment of the COI gene were considered for 115 specimens of *M. galloprovincialis* sampled in 32 localities in France (for details see Table S1). Sampling details, DNA extraction, PCR conditions and sequencing procedures are reported in Koutroumpa et al. (2009). Two data sets were examined for the analysis. We first considered the complete data set containing all 115 sequences resulting from direct sequencing. Some of these sequences have revealed intra-individual variability characterized by ambiguous sites or 'double peaks' (Fig. 1). These sites were positioned on diagnostic polymorphic sites of this species (up to nine ambiguous sites of the 13 polymorphic sites along the sequence). In most of these cases, competing signal was weak and base assignment was scored following the automatic acquisition with Sequencing Analysis v5.4., or corrected manually with CodonCode Aligner V.3.7.1., (CodonCode Corporation, Centerville, MA, USA) assigned as a 'N' when the two peaks overlapped. In a second data set, each individual showing ambiguous sites (i.e. 69 sequences) was removed and only sequences displaying clear sequencing signal patterns were considered.

As mentioned above, the presence of NUMTs could be detected on chromatograms as double peaks and still stay largely readable. Sometimes, such features are fairly doubtful, often recorded as background noise instead of NUMTs. Nevertheless, when several peaks occur in sequence chromatograms, it is difficult to identify the focal mtDNA sequence from the non-target one, sometimes selecting the preferentially

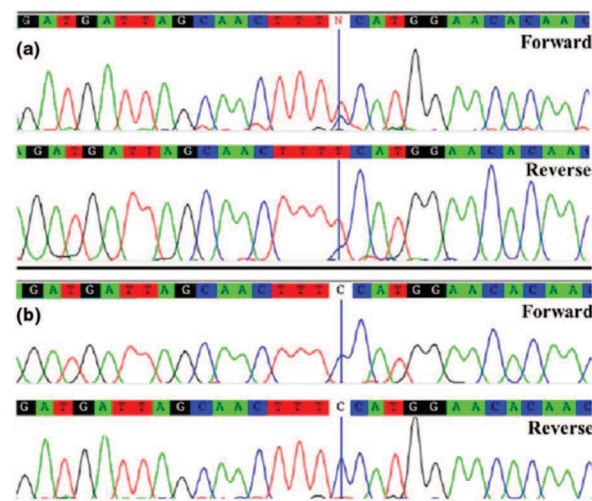


Fig. 1. Evidence for NUMTs in *M. galloprovincialis*. (A) Direct sequencing chromatogram after amplification of COI gene following total DNA extraction: co-amplification of mtDNA target molecule and a non-target molecule resulting in a double peak at the site where the two sequences differ from each other. (B) Direct sequencing chromatogram of the same mtDNA fragment gene after mitochondrial DNA isolation: putative mtDNA (corresponding to the sequence with a lower frequency in Fig. 1A).

amplified sequence instead of the putative mtDNA (as it shown in Fig. 1). Cloning of PCR-amplified fragments is known to be the most reliable way to explore intra-individual variation (Berthier et al. 2011). To explore in further details the diversity of paralogous mtDNA sequences, a cloning was performed on a selection of 15 individuals. These specimens displayed low (GLG1 and GCau2) to high ambiguity signals on the direct sequencing chromatograms (e.g. GCor2 or GLor1, see Table 1 for further details). Five of the specimens displayed apparent clean reads (GMey1, GEspM1, GGr7, GGrP1 and GAdr1). For control purposes, we also included two specimens of the sister species *M. sutor*, for which NUMTs have not been detected (Koutroumpa et al. 2009). The cloning was conducted using the StrataClone PCR Cloning kit (Agilent Technologies, Santa Clara, CA, USA). For each ligation, depending on the success of the cloning, 5–20 positive clones were identified by PCR analysis of the plasmid DNA. Successful amplification was confirmed by agarose gel electrophoresis, and PCR products were subsequently cleaned using a Nucleospin® Gel and PCR Clean-up Kit (Macherey-Nagel, Düren, Germany). For most specimens, 5–11 clones were sequenced from both strands with the BigDye® Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA), with the exception of two specimens displaying apparent clean reads (GMey6 and GAdr2) and for which 16 and 18 clones were, respectively, sequenced. The sequencing was carried out with an ABI 3100 automatic sequencer (Applied Biosystems). Only nucleotide heterogeneities shared among two or more sequences, i.e. those matching with diagnostic mutations, were considered to define authentic haplotypes within amplicon. All singletons were considered as Taq errors and were thus neglected. Putative mtDNA was confirmed for some specimens previously analysed in Koutroumpa et al. (2009). The identification of authentic mtDNA sequence in a set of cloned sequences is often challenging. For older specimens for which alkaline lysis was not successful, we followed Buhay (2009) and considered the most frequently encountered sequence obtained from a single individual to represent mitochondrial sequence. We thus referred it to a COI haplotype, aware of the bias that could be introduced following this method.

mtDNA sequences analysis

The same calculations were conducted on the two data sets (complete and corrected 'ambiguities-free') to assess the effect of NUMTs on genetic diversity and on the reconstruction of the genetic structure. We used the software GENALEX 6.5 (Peakall and Smouse 2012) to estimate mean number of haplotypes and mean haplotype diversity over populations before and after correction. The presence of stop codons, codon bias and variation in C-G content was tested using DNASP 5.0 (Rozas 2009). Haplotype diversity and relative abundance of each haplotype of the two data sets were projected on a map using ARCGIS 9.3 (ESRI, Redlands, CA, USA). Haplotype networks were reconstructed using the software TCS 1.21 (Clement et al. 2000). Pairwise distances within and between clones were computed using the p-distances method implemented in Mega 6 (Tamura et al. 2013).

Results

Mitochondrial DNA

Total number of haplotypes within *M. galloprovincialis* differed considerably when considering the complete data set versus the ambiguity-free one. A total of 44 haplotypes were recorded in the full data set, while haplotype number dropped to 6 in the cleaned one. A comparison of the haplotype distribution obtained is summarized in Fig. 2A & B. The mean number of haplotypes over populations decreased from 2.719 (SD: 0.402) to 1.389 (SD: 0.183) and the mean haplotype diversity over populations dropped from 0.359 (SD: 0.063) to 0.132 (SD: 0.054) after elimination of ambiguous sequences. Finally, the pairwise comparison of nucleotide diversity decreased from 0.00334 to 0.00245. No stop codon or frame shifts were identified in the complete data set. Most mutations occurred at the third position of the codon and did not change amino acid translation. A single non-

Table 1. Position and substitution type of polymorphic sites observed within *M. galloprovincialis* sequences resulting from direct sequencing (grey rows) and corresponding nucleotide after cloning. Base assignment observed for each ambiguous substitution site is indicated in brackets. Ambiguities that could not be assigned (N) were highlighted. The level of heterogeneity observed on the direct sequencing chromatogram was specified as high (***), low (**), and null (0) for each amplicon. Substitution position refers to the position in the reference genome (*D. yacuba*).

Substitution	632	635	653	728	769	829	870	938	962	1050	1052	1122	1124	Observed heterogeneity level
GLor1	A/G (N)	A/T (N)	A	T _C (T/N)	C	C/A (N)	G	C/T (N)	C	G	A/T (N)	C _T (C)	C/T (N)	***
7 clones	1A/6G	1A/6T	1A/6T	1C/6T	7C	7C/0A	7G	4C/3T	7C	7G	1A/6T	5C/1T	5C/2T	***
GLor3	A/G (N)	A/T (N)	A	T _C (N)	C	C _A (C)	G	C/T (N)	C	G	A/T (N)	C _T (C)	T _C (N)	***
7 clones	3A/4G	2A/5T	3A/4T	2C/5T	7C	6C/1A	7G	6C/1T	7C	7G	2A/5T	5C/2T	4C/3T	***
GLor6	A/G (N)	A/T (N)	A	T	C	C/A (N)	G	C/T (N)	C	G	A/T (N)	C _T (C)	C/T (N)	***
8 clones	4A/4G	2A/6T	4A/4T	2C/6T	8C	7C/1A	8G	6C/2T	8C	8G	2A/6T	6C/2T	4C/4T	***
GAdr3	A _G (A)	A/T (N)	A	T _C (N)	C	C _A (C)	G	C/T (N)	C	G	A/T (N)	C/T (N)	C/T (N)	***
11 clones	10A/1G	3A/8T	8A/3T	3C/8T	8C/3G	7C/4A	8C/3C	6C/5T	11C	8G/3A	3A/8T	3C/8T	4C/7T	***
GAud4	A _G (A)	A/T (N)	A	T _C (T)	C	C/A (N)	G	C/T (N)	C	G	T _A (N)	T _C (T)	C/T (N)	***
5 clones	2A/3G	2A/3T	2A/3T	2C/3T	5C	5C/0A	5G	5C/0T	5C	5G	3A/2T	2C/3T	2C/3T	***
GCou2	A/G (N)	A/T (N)	A	C/T (N)	C	C _A (C/N)	G	C/T (N)	C _A (C)	G	A/T (N)	C _T (C)	C _T (C)	***
10 clones	8A/2G	6A/4T	8A/2T	6C/4T	10C	8C/2A	10G	8C/2T	4C/6A	10G	6A/3T+1C	4C/6T	2C/8T	***
GCou2	A _G (A)	A _T (A)	A	C/T (N)	C	C _A (C)	G	C _T (C)	C _A (C)	G	A/T (N)	C/T (N)	C/T (N)	**
9 clones	9A/0G	5A/4T	9A	5C/4T	7C/2G	7C/2A	7G/2C	5C/4T	4C/5A	6C/3A	0A/9T	8C/1T	5C/4T	*
GLG1	A	A _T (A)	A	C _T (C)	C	C _A (C)	G	C _T (C)	C/A (N)	G	A _T (A)	T _C (N)	T	*
9 clones	8A/1G	7A/2T	8A/1T	7C/2T	9C	9C/0A	8G/1C	8C/1T	2C/7A	7G/2A	6A/3T	2C/7T	1C/8T	0
GMey6	A	A	A	C	C	C	G	C	A	G	A	T	C	0
16 clones	15A/1G	14A/2T	15A/1T	14C/2T	16C	15C/1A	16G	15C/1T	14A/2C	16G	14A/1T	14T/1C	10C	0
GEspM1	A	T	A	T	C	A	G	T	C	G	T	T	T	0
10 clones	10A	10T	10A	10T	9C/1G	9A/1C	10G	10T	10C	10G	10T	10T	10T	0
GGra7	A	T	A	T	C	C	G	C	C	G	A	T	T	0
7 clones	7A	7T	7A	7T	7C	6C/1A	7G	6C/1T	7C	7G	6A/1T	7T	7T	0
GGP1	A	T	A	T	C	C	G	C	C	G	-	-	-	0
11 clones	11A	11T	11A	11T	11C	11C	11G	11C	11C	11G	-	-	-	0
GPAdr1	A	A	A	C	C	C	G	C	A	G	-	-	T	0
18 clones	18A	18A	18A	18C	18C	18C	18G	18C	18A	18G	18A	18T	18T	0

Information relative to clones are displayed in *italic*. Bold values refer to double peaks noticeable on electropherograms.

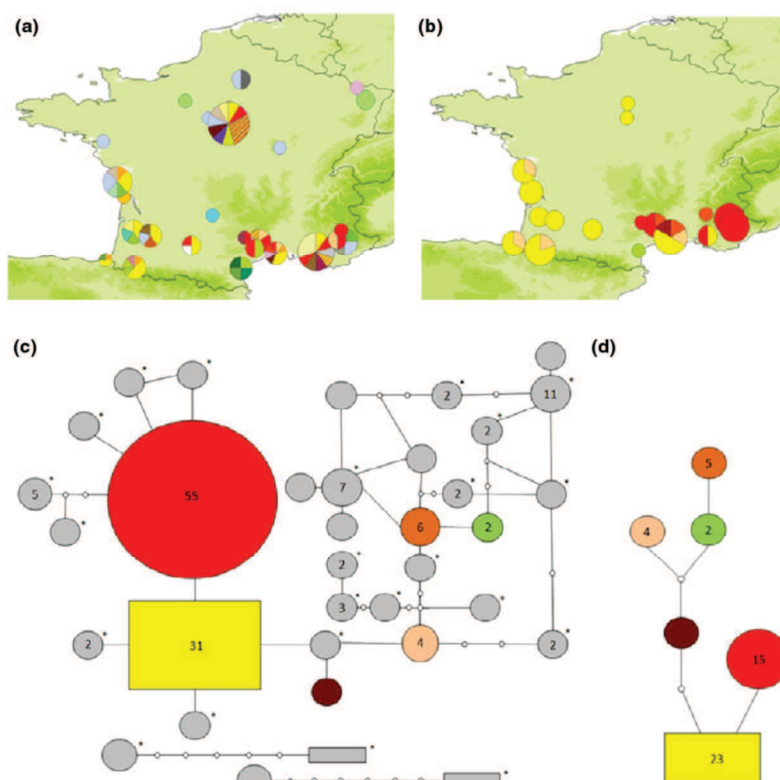


Fig. 2. (A,B) Reconstruction of the genetic structure of *M. galloprovincialis* in France based on COI sequences. (A) full data set (B) corrected 'ambiguities-free' data set. (C,D) Haplotype network of COI of *M. galloprovincialis*. (C) full data set, including cloned specimens. (*) represents the haplotypes containing at least one NUMT obtained by cloning. (D) after removal of ambiguous sequences

synonymous mutation was observed, in position 1124, where the transition T → C involves a change from Leucin (L) to Prolin (P). This mutation has been found in 20 of the 69 ambiguous sequences, but was not encountered in any of the non-ambiguous sequences. As a result, due to a lack of obvious diagnostic characters to recognize pseudogenes, 49 of the sequences with ambiguous sites failed to be recognized as NUMTs in the complete data set. We found a slight but significant decrease of C-G content with the correction (28.67%, $n = 115$; after correction: 28.60%, $n = 46$. Student test: $p\text{-val} < 0.001$), and we detected a significant decrease in codon bias index (uncorrected: 0.778; after correction: 0.774. Student test: $p\text{-val} < 0.001$).

The cloning of individuals exhibiting ambiguities on direct sequencing chromatogram revealed the presence of multiple sequences. Polymorphic sites and frequency of nucleotides found in each individual cloned are summarized in Table 1. The relative proportion of nucleotides observed in each polymorphic site was fortuitous and not always consistent with the proportion of the different peaks observed on the direct chromatogram. In some cases of overlapping signals, only one nucleotide type has been scored for all clones. Conversely, site ascribed as the major peak on the chromatogram, appeared mainly predominant among the clones sequenced. Furthermore, specimens with low or no polymorphism detected in the direct chromatogram (GLG1 or GMey6, GEsPm1, GGrA7) showed a substantial proportion of polymorphism at some sites (e.g. position 235 and 557 for GLG1), which presumably represent putative pseudogenes. No polymorphic sites could be recovered within the pseudogene-free species *M. sutor* and for two individuals displaying no polymorphism in direct chromatogram (GGrP1 and GPAdr1).

Alignment of the 5–18 cloned sequences resulted in 1–6 haplotypes per amplicon, and the mean number of haplotype per specimen was 3.72 (Table 2). Each individual gave at least one sequence that matched one haplotype of the ambiguities-free (mtDNA) data set. Sequence divergence between paralogs varied, ranking from 1 to 11 mutation steps from the putative mtDNA. Within individual maximum divergence between haplotypes ranged from 0.8 to 1.4%. We observed a maximum pairwise distance between all sequences of individuals cloned of 2.3%. No stop codons or frame shifts were detected in the set of cloned sequences.

Haplotype network of the 44 mtDNA-like and cloned sequences exhibited a complex starlike arrangement (Fig. 2C). The main haplotypes were surrounded by multiple unique haplotypes some of which being strongly interconnected. Some highly divergent haplotypes were not connected to the main network. These multiple unique haplotypes and highly divergent ones were not recovered in the haplotype network reconstructed from purged data sets (Fig. 2D).

Discussion

This study aimed to compare phylogeographic patterns obtained before and after elimination of mtDNA paralogous sequences. We brought evidence of the misleading interpretations that might be made from a data set by the inadvertent co-amplification or preferred amplification of NUMTs. Identifying these nuclear paralogs as putative mtDNA or scoring blindly the dominant peak in ambiguous sites both resulted in an overestimation of genetic diversity indices and an inconsistent phylogeographic structure. It

Table 2. Number of clones and haplotypes network of paralogous sequences obtained among the 15 specimens cloned. The basal square represents the assumed mitochondrial sequence of COI. Number of copies obtained for each haplotype is shown in the haplotype network.

Specimen cloned	Locality Country	No. of clones obtained	number of haplotypes	maximum divergence	Mean divergence	Haplotype network
GAdr3	St André les A. France	11	6	1.3%	0.009	
GAud4	Peyraiac de Mer France	5	3	1.3%	0.008	
GCau2	Grandes Cau. France	9	4	1.4%	0.008	
GCor2	St Jean de cornies France	10	4	1.4%	0.007	
GLg1	Puechagut le V. France	9	4	1.4%	0.006	
GLor1	Lorris France	7	5	1.4%	0.005	
GLor3	Lorris France	7	3	1.4%	0.008	
GLor6	Lorris France	7	4	1.4%	0.008	
GMey6	Meyrargues France	14	5	1.3%	0.003	
GEspM1	Murcia Spain	10	2	0.3%	0.001	
GGra7	Afete Greece	7	2	0.8%	0.002	
GGrP1	Patitiri Greece	11	1	–	–	
GPAdr1	St André les A. France	18	1	–	–	
<i>M. sutor</i>	St André les A. France	10	1	–	–	
<i>M. sutor</i>	Seyne France	10	1	–	–	

is worse at the infra-specific level when NUMTs are recently transposed and unnoticeable. For example, the phylogeography of the bark beetle *Ips typographus* based on mtDNA sequences showed that 1/3 of the total haplotype richness of this species was due to the occurrence of NUMTs (Bertheau et al. 2011). Likewise, in *Dendroctonus valens*, two putative COI haplotypes could be recovered after mtDNA purification against 14 from nuclear DNA of the same individuals (Cai et al. 2010). In our case, the elimination of ambiguous sequences of the data set reduced considerably the number of haplotypes, disentangling a more accurate geographic pattern of haplotype distribution. Because PCR is a stochastic process that might preferentially amplify NUMTs without ambiguities on direct chromatograms (Zhang and Hewitt 1996; Sorenson and Quinn 1998; Song et al. 2008; Buhay 2009), we cannot exclude the risk of persistence of NUMTs in the corrected data set and consequently the persistence of additional haplotypes on the phylogeography. Similar pattern of the genetic structure within *M. galloprovincialis* was observed in the course of microsatellites development for this species (Haran and Roux-Morabito 2014).

Various strategies were proposed to avoid NUMT occurrence within mitochondrial data sets (Bensasson et al. 2001; Song et al. 2008; Buhay 2009). Calvignac et al. (2011) have recently investigated several *a priori* approaches to prevent NUMTs co-amplification or preferred amplification. Among the strategies tested by the authors (mtDNA enrichment, cDNA amplification,

long-range amplification and pre-PCR dilution), they retained the last one as the most effective method to prevent the pollution of putative mtDNA data sets with NUMTs. Primer optimization (Sorenson and Quinn 1998; Podnar et al. 2007) could be useful to prevent persistence of NUMTs, especially at the infra-specific level for which universal primers are not a prerequisite, but this method still shows limited extent as described in Moulton et al. (2010). For conventional sequencing approaches, only a separation between mitochondrial and nuclear DNA by alkaline digestion (Tamura and Aotsuka 1988), a specific amplification of mitochondrial genome (Wolff et al. 2012) or a reverse-transcription PCR of mtRNA (Sun et al. 2012) provide the guarantee of the full absence of NUMTs. All these methods remain costly or involve heavy protocols requiring living or very well conserved organisms. It is therefore difficult to apply them at a large scale, for population genetic and phylogeographic studies.

A crucial point in the prevention of NUMTs remains their detection. *A posteriori* methods currently used for detecting NUMTs (i.e. quality control of the sequence alignment) face strong limitations, especially at infra-specific level. Among the tools available for that purpose, the most used and cost-effective method is a combination of several 'postsequencing' investigations aiming at targeting the presence of stop codons, frame shifts or non-synonymous mutations on the sequence (Bensasson et al. 2001; Song et al. 2008; Baldo et al. 2011). The cloning performed here has allowed isolation of each paralogous

sequences encountered in a single specimen. Thus, it was possible to assess their detection if preferentially amplified during PCR. Our observations show that none of the paralogous sequences obtained exhibited frames shifts or stop codons. Non-synonymous mutations permitted to identify several NUMTs within the paralogous sequences, but some differing by 1–3 bp only remained undetected. We found a majority of mutation occurring at the third codon position in NUMTs sequences that did not change amino acid translation. This contradicts the hypothesis of random mutations of these sequences when included in the nuclear genome, where they lose their functionality (Bensasson et al. 2001), indicating a rather recent transposition event as suggested in Koutroumpa et al. (2009).

An intra-individual COI variation study in two grasshopper species revealed that 70–75% of cloned sequences diverged by less than 1% from the putative mtDNA (Berthier et al. 2011). The authors ascribed these COI-like sequences to heteroplasmy, in addition to more distant cloned sequences that diverged by more than 9% and assigned as NUMTs. Bertheau et al. (2011) have applied the term of cryptic NUMTs to describe slightly divergent NUMTs that remain undetected by the control methods mentioned above. Cryptic NUMTs were also described for *Aedes aegypti* (Hlaing et al. 2009). It should be noticed that the occurrence of these closely related mtDNA sequences is more prevalent than recorded but not always detected, due mainly to the haphazard of the PCR-amplification and primers used. Heteroplasmic molecules and fortuitous NUMTs are both a major problem for phylogeography as they lead to an overestimation of genetic diversity and disturb the haplotype networks while remaining undetected. Haplotype networks reconstructed with NUMTs exhibiting single mutational events exhibit a starlike network pattern, which is usually considered as the signature of a recent demographic expansion of populations (Avise 2000). This phenomenon is thought to explain a significant part of the excess of haplotypes encountered in our study. Indeed, 49 of the 69 *M. galloprovincialis* sequences with ambiguities were not detected and can be assigned to cryptic NUMTs, as no abnormality could be detected on their open reading frames.

The case of *M. galloprovincialis* presented here reminds that despite awareness against NUMTs over the last decades, this phenomenon remains poorly known and still escapes control methods. It calls for special attention in future studies to thoroughly examine mitochondrial sequence data sets displaying NUMT characteristics, as it has been performed for *I. typographus* (Mayer et al. 2014) or for two termite genera (Hausberger et al. 2011). It might also lead to the reconsideration of previous phylogeographic studies, as it has been performed for few examples (Arthofer et al. 2010; Bertheau et al. 2011). Finally, checking out the congruence between several independent markers as well as considering a sufficient number of individuals would contribute to detect and prevent persistence of NUMTs in mitochondrial data sets.

More generally, fortuitous NUMTs also address the question of the reliability of public COI databases and their current use in metabarcoding approaches, especially given the combination of next-generation sequencing and 'mini-barcodes' for species identification (Hajibabaei et al. 2011; Shokralla et al. 2014). In a recent study, Shokralla et al. (2014) identified 10% of heteroplasmic-like COI sequences using 454-pyrosequencing performed on 190 specimens of Lepidoptera, although the heteroplasmic origin of paralogous sequences was not clearly established and might be related to NUMTs. Furthermore, this study highlighted the absence of evident phylogenetic pattern for the presence of this bias in this group. This species specificity of NUMTs evoked also for the genus *Monochamus* (Koutroumpa et al. 2009), even

if not generalized, raises another dimension in their consideration. Indeed, this observation suggests no link between the taxonomic placement of a species and the frequency of NUMTs. Subsequently, to ensure the highest mtDNA quality, each taxonomic group and species should be carefully considered in the light of the NUMT bias. For this purpose, next-generation sequencing would help in recovering heteroplasmic or NUMT features (Shokralla et al. 2014).

Acknowledgements

We acknowledge Sophie Delamaire for her assistance with cloning and DNA sequencing. We are grateful to Carole Kerdelhué and two anonymous reviewers for their comments and valuable remarks on the previous manuscript. This work was supported by INRA-EFPA and the European Project REPHRAME. PhD grant of the first author was provided by the French Ministry of Research and Education.

References

- Arthofer W, Avtzi DN, Riegler M, Stauffer C (2010) Mitochondrial phylogenies in the light of pseudogenes and *Wolbachia*: re-assessment of a bark beetle dataset. In: Cognato AI, Knížek M (Eds) Sixty years of discovering scolytine and platypodine diversity: A tribute to Stephen L. Wood. *ZooKeys* **56**:269–280.
- Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge.
- Ayliffe MA, Scott NS, Timmis JN (1998) Analysis of plastid DNA-like sequences within the nuclear genomes of higher plants. *Mol Biol Evol* **15**:738–745.
- Baldo L, de Queiroz A, Hedin M, Hayashi CY, Gates J (2011) Nuclear-mitochondrial sequences as witnesses of past interbreeding and population diversity in the jumping bristletail *Mesomachilis*. *Mol Biol Evol* **28**:195–210.
- Bensasson D, Zhang DX, Hewitt GM (2000) Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Mol Biol Evol* **17**:406–415.
- Bensasson D, Zhang D, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol Evol* **16**:314–321.
- Bertheau C, Schuler H, Krumböck S, Arthofer W, Stauffer C (2011) Hit or miss in phylogeographic analyses: the case of the cryptic NUMTs. *Mol Ecol Resour* **11**:1056–1059.
- Berthier K, Chapuis MP, Moosavi SM, Tohidi-Esfahani D, Sword GA (2011) Nuclear insertions and heteroplasmy of mitochondrial DNA as two sources of intra-individual genomic variation in grasshoppers. *Syst Entomol* **36**:285–299.
- Black IV WC, Bernhardt SA (2009) Abundant nuclear copies of mitochondrial origin (NUMTs) in the *Aedes aegypti* genome. *Insect Mol Biol* **18**:705–713.
- Buhay JE (2009) COI-like sequences are becoming problematic in molecular systematic and DNA barcoding studies. *J Crustacean Biol* **29**:96–110.
- Cai Y, Cheng X-Y, Duan D, Xu R (2010) Mitochondrial COI gene transfers to the nuclear genome of *Dendroctonus valens* and its implications. *J Appl Entomol* **135**:302–310.
- Calvignac S, Konecny L, Malard F, Douady CJ (2011) Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion* **11**:246–254.
- Cameron SL (2014) Insect mitochondrial genomics: implications for evolution and phylogeny. *Annu Rev Entomol* **59**:95–117.
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**:1657–1659.
- Cristiano MP, Fernandes-Salomao TM, Yotoko KSC (2012) Nuclear mitochondrial DNA: an Achilles' heel of molecular systematics, phylogenetics, and phylogeographic studies of stingless bees. *Apidologie* **43**:527–538.
- Galtier N, Nabholz B, Glémin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol* **18**:4541–4550.

- Gjerde B (2013) Characterisation of full-length mitochondrial copies and partial nuclear copies (numts) of the cytochrome b and cytochrome c oxidase subunit I genes of *Toxoplasma gondii*, *Neospora caninum*, *Hammondia heydorni* and *Hammondia triffittae* (Apicomplexa: Sarcocystidae). *Parasitol Res* **112**:1493–1511.
- Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One* **6**:e17497.
- Haran J, Roux-Morabito G (2014) Development of 12 microsatellite loci for the longhorn beetle *Monochamus galloprovincialis* (Coleoptera Cerambycidae), vector of the Pine Wood Nematode in Europe. *Conserv Genet Resour* **6**:975–977.
- Haran J, Timmermans MJTN, Vogler AP (2013) Mitogenome sequences stabilize the phylogenetics of weevils (Curculionidae) and establish the monophyly of larval ectophagy. *Mol Phylogenet Evol* **67**:156–166.
- Hausberger B, Kimpel D, van Neer A, Korb J (2011) Uncovering cryptic species diversity of a termite community in a West African savanna. *Mol Phylogenet Evol* **61**:964–969.
- Hebert PDN, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome c oxidase subunit I divergences among closely related species. *Proc R Soc Lond B Biol Sci* **270**:S96–S99.
- Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victorian PF, Yoder AD (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. *Mol Phylogenet Evol* **54**:291–301.
- Hlaing T, Tun-Lin W, Somboon P, Socheat D, Setha T, Min S, Chang MS, Walton C (2009) Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. *BMC Genet* **10**:11.
- Hudson RR (1991) Gene genealogies and the coalescent process. *Oxf Surv Evol Biol* **7**:1–44.
- Keller I, Bensasson D, Nichols RA (2007) Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. *PLoS Genet* **3**:e22.
- Koutroumpa FA, Lieutier F, Roux-Morabito G (2009) Incorporation of mitochondrial fragments in the nuclear genome (Numts) of the longhorn beetle *Monochamus galloprovincialis* (Coleoptera, Cerambycidae). *J Zool Syst Evol Res* **47**:141–148.
- Koutroumpa FA, Rougon D, Bertheau C, Lieutier F, Roux-Morabito G (2013) Evolutionary relationships within European *Monochamus* (Coleoptera: Cerambycidae) highlight the role of altitude in species delineation. *Biol J Linn Soc Lond* **109**:354–376.
- Leister D (2005) Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet* **21**:655–663.
- Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJO (1994) *Nmt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J Mol Evol* **39**:174–190.
- Lunt DH, Zhang DX, Szymura JM, Hewitt GM (1996) The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Mol Biol* **5**:153–165.
- Mayer F, Björklund N, Wallén J, Långström B, Cassel-Lundhagen A (2014) Mitochondrial DNA haplotypes indicate two postglacial recolonization routes of the spruce bark beetle *Ips typographus* through northern Europe to Scandinavia. *J Zool Syst Evol Res* **52**:285–292.
- Mota M, Braasch H, Bravo MA, Penas AC, Burgermeister W, Metge K, Sousa E (1999) First report of *Bursaphelenchus xylophilus* in Portugal and in Europe. *Nematology* **1**:727–734.
- Moulton MJ, Song H, Whiting MF (2010) Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Mol Ecol Resour* **10**:615–627.
- Neigel JE, Avise JC (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: Karlin S, Nevo E (eds), *Evolutionary Processes and Theory*, Academic Press, New York, pp 515–534.
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* **28**:2537–2539.
- Podnar M, Haring E, Pinsker W, Mayer W (2007) Unusual origin of a nuclear pseudogene in the Italian wall lizard: intergenomic and interspecific transfer of a large section of the mitochondrial genome in the genus *Podarcis* (Lacertidae). *J Mol Evol* **64**:308–320.
- Pons J, Vogler AP (2005) Complex pattern of coalescence and fast evolution of a mitochondrial rRNA pseudogene in a recent radiation of tiger beetles. *Mol Biol Evol* **22**:991–1000.
- Richly E, Leiste D (2004) NUMTs in sequenced eukaryotic genomes. *Mol Biol Evol* **21**:1081–1084.
- Robertson L, Cobacho AS, Escuer M, Santiago Merino R, Esparrago G, Abelleira A, Navas A (2011) Incidence of the pinewood nematode *Bursaphelenchus xylophilus* Steiner & Bührer, 1934 (Nickle, 1970) in Spain. *Nematology* **13**:755–757.
- Rokas A, Ladoukakis E, Zouros E (2003) Animal mitochondrial DNA recombination revisited. *Trends Ecol Evol* **18**:411–417.
- Rozas J (2009) DNA sequence polymorphism analysis using DnaSP. *Methods Mol Biol* **537**:337–350.
- Ruiz C, de J May-Itza W, Quezada-Euan JGG, De la Rua P (2013) Presence of nuclear copies of mitochondrial origin (NUMTs) in two related species of stingless bee genus *Melipona* (Hymenoptera: Meliponini). *J Zool Syst Evol Res* **51**:107–113.
- Shokralla S, Gibson JF, Nikbakht H, Janzen DH, Hallwachs W, Hajibabaei M (2014) Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Mol Ecol Resour* **14**:892–901.
- Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proc Natl Acad Sci USA* **105**:13486–13491.
- Song HJ, Sheffield NC, Cameron SL, Miller KB, Whiting MF (2010) When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. *Syst Entomol* **35**:429–448.
- Sorenson MD, Quinn TW (1998) Numts: a challenge for avian systematics and population biology. *Auk* **115**:214–221.
- Sun Y, Ji Y, Luo D, Liao DJ (2012) Pseudogenes as weaknesses of ACTB (Actb) and GAPDH (Gapdh) used as reference genes in reverse transcription and polymerase chain reactions. *PLoS One* **7**:e41659.
- Sword GA, Senior LB, Gaskin JF, Joern A (2007) Double trouble for grasshopper molecular systematics: intra-individual heterogeneity of both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer ribosomal DNA sequences in *Hesperotettix viridis* (Orthoptera: Acrididae). *Syst Entomol* **32**:420–428.
- Tamura K, Aotsuka T (1988) Rapid isolation method of animal mitochondrial DNA by the alkaline lysis procedure. *Biochem Genet* **26**:815–819.
- Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**:2772–2779.
- Wolff JN, Shearman DCA, Brooks RC, Ballard JWO (2012) Selective enrichment and sequencing of whole mitochondrial genomes in the presence of nuclear encoded mitochondrial pseudogenes (numts). *PLoS One* **7**:e37142.
- Zhang DX, Hewitt GM (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol Evol* **11**:247–251.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Sampling details of the 115 french and 13 spanish and greek specimens considered. *N* refers to the number of individuals collected in each locality.

Table S1: Sampling details of the 115 french and 13 spanish and greek specimens considered. N refers to the number of individuals collected in each locality.

Code	N	Country / Location	Collector	Host species	Date	Latitude	Longitude
GCom	2	France / Compiègne	D. Mure	<i>P. sylvestris</i>	2004	49°22' N	02°54' E
GBos	1	France / Prénouvellon	J. Mouthiez	No forest	2004	47°58' N	01°32' E
Gpig	2	France / FD 3 Pignons (Villeneuve St Denis)	B. Germain	<i>P. pinaster</i>	2003	48°24' N	02°33' E
GRam	1	France / Rambouillet (St Leger en Yvelines)	A. Jarsaillon	<i>P. sylvestris</i>	2003,2004	48°38' N	01°49' E
GLor ^{1,2}	11	France / Lorris	Authors	<i>P. sylvestris</i>	2003,2004	47°53' N	02°30' E
GSag	1	France / Savigny	B. Hutteringer	<i>P. nigra (laricio)</i>	2004	47°03' N	04°49' E
GStu	1	France / Sturzelbronn	H. Schmuck	<i>P. sylvestris</i>	2004	49°03' N	07°35' E
GHag	2	France / Haguenau	F. Dorffer	<i>P. sylvestris</i>	2003	48°49' N	07°47' E
GEyg	1	France / FD Eygues (com. Rosans)	P. Fee	<i>P. nigra</i>	2004	44°24' N	05°28' E
GAdr	5	France / St André les Alpes	L. Micas	<i>P. sylvestris</i>	2004	43°59' N	06°30' E
GDig	2	France / Digne (Les Bains)	L. Micas	<i>P. sylvestris</i>	2005	44°05' N	06°13' E
GSeY	1	France / Seyne (Maison Forestière 1600m)	D. Rougon	<i>P. sylvestris</i>	2003	44°21' N	06°23' E
GpFau	2	France / Seyne (Le Faut 1350m)	D. Rougon	<i>P. sylvestris</i>	2004	44°19' N	06°24' E
GpVer	5	France / Verdaches 1100m	D. Rougon	<i>P. sylvestris</i>	2005	44°15' N	06°20' E
GpAdr	4	France / L'Ergatiere (St André les Alpes 1100m)	D. Rougon	<i>P. sylvestris</i>	2005	43°59' N	06°30' E
GMey ¹	10	France / Meyrargues	J. Hintzy	<i>P. halepensis</i>	2003,2004	43°28' N	05°38' E
GCor ¹	9	France / St Jean de Cornies	P. Girard	<i>P. halepensis</i>	2003,2004	43°44' N	04°00' E
GLG	6	France / Puechagut le Vigan	C. Rulliere	<i>P. nigra</i>	2003,2004	44°01' N	03°34' E
GCau	2	France / FD Grands Causses	J-P. Ansonnaud	<i>P. nigra (laricio)</i>	2004	43°58' N	03°15' E
GLar	2	France / Larzac	J-P. Ansonnaud	<i>P. nigra (laricio)</i>	2005	43°58' N	03°11' E
GAud ²	4	France / FD de Foufroide (com. Peyriac de Mer)		<i>P. halepensis</i>	2004	43°17' N	01°48' E
GTou	1	France / Toulouse		<i>P. nigra</i>	2003	43°36' N	01°26' E
GArg	1	France / Argentat		<i>P. nigra</i>	2004	45°05' N	01°56' E
GHer	4	France / Herm (Léon)	J-L. Hautclocq	<i>P. pinaster</i>	2004	43°51' N	01°14' W
Gpis	5	France / Pissos	P. Ninosque	<i>P. pinaster</i>	2004	44°18' N	00°46' W
GFar	5	France / Fargues	R. Delpont	<i>P. pinaster</i>	2004	44°11' N	00°11' E
GSarI	8	France / Sare (St Ignace)	C. Van-Meer	<i>P. nigra (laricio)</i>	2004	43°20' N	01°36' W
GSarL	4	France / Sare (Lizarrieta)	C. Van-Meer	<i>P. nigra (laricio)</i>	2004	43°16' N	01°37' W
GHou	3	France / Hourtin	J-P. Coste	<i>P. pinaster</i>	2004,2005	45°12' N	01°04' W
GHnor	1	France / Hourtin (NORD)	Authors	<i>P. pinaster</i>	2005		
Gole	8	France / St Pierre d' Oléron	R. Pelloquint	<i>P. pinaster</i>	2003,2004	45°58' N	01°19' W
GGav	1	France / Gavre	J. Belloir	<i>P. pinaster</i>	2004	46°39' N	00°44' W
GEspM	5	Spain / Murcia	D. Gallego	<i>P. pinaster</i>	2004	37°59' N	01°07' W
GGrA	7	Greece / Afete(Magnesia) 300m	A. Koutroumpas	<i>P. brutia</i>	2004,2005	39°22' N	22°56' E
GGrP	1	Greece / Patitiri, Alonissos island	S. & N. Eustathiou	<i>P. brutia</i>	2005	39°09' N	23°51' E

Annexe n°2

Haran J & Roux G 2014. Development of 12 microsatellites loci for the longhorn beetle *Monochamus galloprovincialis* (Coleoptera Cerambycidae), vector of the pinewood nematode in Europe. *Conservation Genetics Resources* 6, 975-977.

Development of 12 microsatellites loci for the longhorn beetle *Monochamus galloprovincialis* (Coleoptera Cerambycidae), vector of the Pine Wood Nematode in Europe

Julien Haran · Géraldine Roux-Morabito

Received: 4 June 2014 / Accepted: 7 July 2014
© Springer Science+Business Media Dordrecht 2014

Abstract *Monochamus galloprovincialis* (Coleoptera, Cerambycidae) is the vector of the Pine Wood Nematode (*Bursaphelenchus xylophilus*) in Europe. This nematode is the agent of the Pine Wilt Disease (PWD), a syndrome causing important damages in Eurasian pine forests. We report here the development of 12 polymorphic microsatellites loci specific to this species. We observed a moderate allelic richness over all loci. The number of allele per locus ranked from 2 to 8, with an average of 3.75. These markers will help to characterize the effect of the spread of an invader on the genetic structure of a native vector species.

Keywords Microsatellites · *Monochamus galloprovincialis* · Coleoptera · Cerambycidae · Pine Wood Nematode · *Bursaphelenchus xylophilus* · Europe

The longhorn beetle *Monochamus galloprovincialis* (Coleoptera, Cerambycidae) is a widely distributed and native species in Europe. It performs its life cycle in the wood of stressed or freshly dead trees of the genus *Pinus*. This species has been recognized as the main vector involved in the spread of Pine Wood Nematode (PWN) *Bursaphelenchus xylophilus*, in Europe. The PWN is native to North America and has been introduced in several Eurasian countries. It is the agent of the Pine Wilt Disease, a syndrome causing important tree mortality in both plantations and natural pine ecosystems. The amount of dead

trees generated by the invasion of the PWN is strongly affecting the population dynamics of *M. galloprovincialis*. This situation is a unique opportunity to study the effect of an alien spread on genetic structure of an associated native species. We developed a set of 12 microsatellites loci to identify changes in genetic structure of *M. galloprovincialis* involved by the invasion of the PWN.

Total genomic DNA was extracted from a leg using a Nucleospin® Kit (Macherey–Nagel, Düren, Germany) and sent to GenoScreen for development and optimization of microsatellites markers (www.genoscreen.fr). One mg of DNA was used for the development of microsatellite libraries through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries as described in Malausa et al. (2011). Total DNA was mechanically fragmented and enriched for AG, AC, AAC, AAG, AGG, ACG, ACAT and ATCT repeat motifs. PCR products were purified, quantified and GsFLX libraries were then carried out following manufacturer's protocols and sequenced on a GsFLX PTP. Sequences analysis and primer design were made using the software QDD (Megléczy and Martin 2009). A number of 386 loci were identified and 48 were selected for further assessment. Specificity and polymorphism assessment were tested using 14 individuals of *M. galloprovincialis* from various points of its whole distribution ranges. Among the 48 loci selected, 22 were validated as specific at one locus. Polymorphism was assessed on 18 loci and 12 were found to be polymorphic for the 14 individuals tested. A fluorescent dye (6'FAM, NED, PET VIC) was attached to 25 % of the forward primer of each pairs for peaks discrimination. The 12 loci were grouped into three multiplexes.

Multiplexes conditions were validated in a 10 µL PCR reaction volume using 25 ng of genomic DNA, 0.4 U of DreamTaq DNA Polymerase (Thermo Scientific®), 0.75 µL Dream Taq Green Buffer (including 20 mM

Electronic supplementary material The online version of this article (doi:10.1007/s12686-014-0262-0) contains supplementary material, which is available to authorized users.

J. Haran (✉) · G. Roux-Morabito
INRA, UR633 Zoologie Forestière, 45075 Orléans Cedex 2,
France
e-mail: julien.haran@gmail.com

Published online: 17 July 2014

 Springer

Table 1 Details of loci characteristics and basic genetic parameters of the population tested (Orleans, France, n = 30), including the multiplexing of primers “Multiplex” and relative amount of each primer “Amount”

Locus	Primers sequence 5' → 3'	Multiplex	Amount (pmol)	Motif	Size range	Na	Ho	He	Fis
Mon 01	F: TTCACGCACATCATTTCTTTG R:TCAAGCAGGAAACGAAAAGC	1	2.4	(aac)5	122–146	2	0.455	0.351	−0.294
Mon 08	F: TGGTGTCTGTAGAACGCTTCA R:GCTTATTAGCTCTCATCAGTATGCTC	3	1	(tata)5	192–196	2	0.074	0.071	−0.038
Mon 17	F: TAGTTTTACTGGGGCCAATG R:GAAGTCATGAACGGATATAAATGAA	3	1.6	(gt)6	149–153	3	0.286	0.487	0.413
Mon 23	F: ATTTATTCCAAATTGCCAATACTACA R:GTGTAAGGTGGAAGTGTCAAAAGC	1	2	(ca)7	142–144	2	0.379	0.307	−0.234
Mon 27	F: ACAATCTCTTTTCGATACCGTTGA R:TTTGCTACAAAGATGTTCTTAAAGT	3	2	(tg)7	118–124	4	0.167	0.517	0.677
Mon 30	F: TTCTAGTTTGCTGTATCCCG R:AAAGCGGGTGTGAAGTACCA	1	1.6	(ag)8	236–248	4	0.133	0.340	0.608
Mon 31	F: GTAAGAGAACCCAACCAACCG R:TTATCCTCACCGGACCGTTA	2	0.64	(ag)8	146–150	3	0.586	0.490	−0.197
Mon 35	F: TTTCATCTGACTTAATTTCTTTCTCA R:AGGGACGTGCAGATTAGGAA	2	7.2	(tc)8	110–126	8	0.519	0.784	0.339
Mon 36	F: ACGGTGCAGAACTAAAGTTAGCC R:GACTCGGACGGAGCTTCT	2	0.8	(tg)8	196–200	3	0.448	0.437	−0.026
Mon 41	F: ACGGTAGCGCAATCTTGAGT R:ACGTGAGCAGTCCTGTTGC	3	0.8	(ct)9	116–120	3	0.200	0.209	0.045
Mon 42	F: CCTTATTTAGGTCAAGAATTCGC R:GCGTGTCAATTATCCAAGGAC	1	1.6	(gt)9	138–146	5	0.586	0.627	0.065
Mon 44	F: AACCTGGACCTAGCTCGGAA R: TAGGAGAAGGTGGAGCAGGA	2	0.64	(act)10	98–113	6	0.552	0.667	0.173

Na number of alleles, Ho and He observed and expected heterozygotes, estimates of Fis values

MgCl₂, Thermo Scientific®), 1 μM Betaine, 0.24 μL dNTP (10 μM) and deionized H₂O. Detail of primers proportions of multiplexes is provided in Table 2. PCR amplifications were run on a Veriti® 96 well fast Thermal cycler (Applied biosystems®) using the following settings: a first denaturation step at 95 °C during 10 min; 40 cycles of denaturation (30 s at 95 °C), hybridization (30 s at 55 °C) and elongation (1 min at 72 °C); and a final elongation step at 72 °C during 10 min. One μL of PCR products were denatured within a mix of 10 μL of formamide and 0.3 μL of 600 Liz marker before being run on an ABI PRISM 3500 sequencer (Life Technologies®). Genotypes were read using the software GeneMapper V 4.1 (Applied Biosystems®).

Genetic parameters of the 12 microsatellites loci were calculated for one population from France (Orleans, n = 30). Deviation from Hardy–Weinberg equilibrium at each locus (Ho, He) and Fis were calculated using the software Genepop 4.2 (Raymond and Rousset 1995). We used the software Genepop 4.2 (Raymond and Rousset 1995) and a False Discovery Rate (FDR) correction (QVALUE, Storey 2002) to test the presence of linkage

disequilibrium in the set of loci. Null allele frequencies were estimated using the software Micro-checker (Van Oosterhout et al. 2004). The transferability of the 12 loci was tested on 3 congeneric European species: *Monochamus sutor* (Linnaeus), *M. sartor* (Fabricius), *M. rosenmuelleri* (Cederjelm) and on the asian vector of the PWN (*M. alternatus*, Hope).

PCR of the three multiplexes gave positive results and the 12 loci could be amplified for all individuals tested. The genetic parameters are presented in Table 1. We observed a moderate allelic richness over all loci. The number of allele per locus ranked from 2 to 8, with an average of 3.75. No significant linkage disequilibrium was detected in this set of loci. An excess of homozygotes was found at locus Mon17, Mon27, Mon30 and Mon35, suggesting the presence of null alleles for those markers in the population studied. Cross priming assessment gave positive results for the four species tested (Table 2). At least 5 loci were fully or partially amplified for each species. *M. sutor* gave the best results, with 9 of the 12 loci amplified. The loci Mon 31, 36, 8 and 17 were fully or partially amplified for all the species.

Table 2 Transferability of the 12 loci developed for *M. galloprovincialis*

Species	N	Microsatellite loci											
		01	23	30	42	31	35	36	44	08	17	27	41
<i>M. sutor</i>	26	–	100	11	–	100	23	100	–	42	88	19	11
<i>M. sartor</i>	4	–	–	75	–	100	–	100	100	100	100	–	–
<i>M. rosenmuelleri</i>	2	–	–	100	100	100	–	100	100	100	100	100	–
<i>M. alternatus</i>	2	–	–	100	–	100	–	50	–	100	100	100	–

Values indicate the rate of positive amplification

N number of individuals tested, “–” no amplification

Acknowledgments This work was supported by the European project REPHRAME.

References

- Malausa T, Gilles A, Meglécz E, Blanquart H, Duthoy S, Costedoat C, Martin JF (2011) High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Mol Ecol Res* 11:638–644
- Megléc E, Martin JF (2009) QDD version 1 a user-friendly program to select microsatellite markers and design primers from large sequencing projects (online software)
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Heredity* 86:248–249
- Storey JD (2002) A direct approach to false discovery rates. *J R Stat Soc B* 64:479–498
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4:535–538

Annexe n°3

Congrès nationaux et internationaux relatifs à cette thèse.

Gruffudd H, Evans H, Haran J, Roux-Morabito G, Roques A & Robinet C, 2013. How could climate change affect the potential spread of pine wilt disease in Europe? ClimTree, Zurich (Suisse), 2-4 Septembre.

Haran J, Roques A & Roux-Morabito G, 2013. Assessing potential expansion of the pinewood nematode (*Bursaphelenchus xylophilus*) from the spatial genetic structure of the vector (*Monochamus galloprovincialis*). IUFRO 7.02.10 "Pine Wilt Disease Conference 2013" Braunschweig, (Allemagne), 15–18 Octobre.

Haran J, Garcia J, Bernard A, Roques A, Robinet C, Roux-Morabito G, 2013. Etude des processus de dispersion de *Monochamus galloprovincialis*, vecteur du nématode du pin : rôle potentiel de la chaîne pyrénéenne. Groupe des Entomologistes Forestiers Francophones (GEFF), Bedoin (France), 2 Octoctbre.

Haran J, Robinet C, Roques A, Roux G, 2014. Etude des processus de dispersion de *Monochamus galloprovincialis*, vecteur du nématode du pin : le rôle potentiel de la chaîne pyrénéenne. 27^{ème} Colloque Biotechnocentre, Seillac (France), 9-10 Octobre.

Haran J, Robinet C, Bernard A, Roques A, Roux G, 2014. Etat des connaissances sur la dispersion de *Monochamus galloprovincialis*, vecteur du nématode du pin en Europe. Groupe des Entomologistes Forestiers Francophones (GEFF), Lucelle (France), 16-18 Septembre.

Haran J, Roques A, Robinet C, Roux-Morabito G, 2014. Assessing dispersal routes and ongoing gene flow of the vector of the pinewood nematode, *Monochamus galloprovincialis*, at different spatial scales. XXIV IUFRO World Congress, Salt Lake City, (USA), 5-11 Octobre.

Haran J, Roques A, Roux-Morabito G, 2014. Evolutionary history and ongoing gene flow of *Monochamus galloprovincialis* (Coleoptera, Cerambycidae), vector of the pinewood nematode. IUFRO 7.03.14, 7.03.06, 7.03.01 Joint Meeting, Antalya (Turquie), 9-14 Avril.

Haran J, Robinet C, Roques A, Roux G, 2014. Etude des processus de dispersion de *Monochamus galloprovincialis*, vecteur du nématode du pin: le rôle potentiel de la chaîne pyrénéenne. Colloque de lancement GDR Invasions Biologiques, Rennes (France), 20-22 Octobre.

Haran J, Robinet C, Roques A, Roux G, 2014. Landscape genetics of *Monochamus galloprovincialis*, Vector of the pinewood nematode in Europe. Neobiota, 8th International Conference on Biological Invasions, Antalya (Turquie), 03-08 Novembre.

Haran J, Roques A, Roux G, 2014. Landscape genetics of *Monochamus galloprovincialis*, Vector of the pinewood nematode in Europe. Le Studium Conference, Orléans (France), 17-19 Decembre.

Haran J, Pajares JA., Sousa E, Rossi JP, Roques A, Robinet C, Roux G, 2015. The importance to combine landscape-level replicates and multiple scales in landscape genetics analysis: implication for inference of gene flow in insects vectors of forest pathogens. 5th International Workshop on the Genetics of Tree-Parasite Interactions, Orléans (France), 23-28 Août.

Annexe n°4

Scripts R utilisés pour les articles n°2 (page 216) et n°3 (page 219).

##(Article n°2) Script loops for partial Mantel tests in sampling area of multiple scales and location (example for elevation (E) as resistant feature):

```
library("ecodist")
library("ade4")
library("raster")
library("gdistance")
library("vegan")

# Definition of the grid / Projection: LAEA metric system
x <- seq(2600+0.5,3800-0.5,by=1)
y <- seq(1500+0.5,2500-0.5,by=1)
vectxy <- expand.grid(x,y)
nx = length(x)
ny = length(y)
dx = 1
dy = 1

# Transition matrix for raster: Elevation
elev = raster("elevpiberiq10")
vectpin = extract(elev,cbind(vectxy[,1]*1000,vectxy[,2]*1000))
vectpin[is.na(vectpin)] = 0
matpin = matrix(vectpin,byrow=F,ncol=ny)
image(x*1000,y*1000,-matpin11,asp=1,xlab="",ylab="",main="elevation")
ft <- function(x) (10000-x[1]) # Transformation: high elevation=resistance
Telev <- transition(elev, transitionFunction=ft, directions=4)
Telev<-geoCorrection(Telev, type="c")
plot(raster(Telev),main="Transition Elevation ")

# Spatial coordiantes Spain (LAEA)
XLAEAj<- read.csv("Spain xcoordinates.txt", header = FALSE, sep = "")
YLAEAj<- read.csv("Spain ycoordinates.txt", header = FALSE, sep = "")
locABCD<-locbis<-data.frame(XLAEAj,YLAEAj)
points(XLAEAj,YLAEAj)

# Genetic data (1 row = 1 individual, each allele coded as : 0=absent,
1=hereozygot, 2=homozygot)

matG<-read.csv("Matrice Genet et Coord JitterPT.txt", header = FALSE, sep =
"")
matG2<-as.matrix(matG)

#-----#
# Loop for mantel test in sampling area at multiple scales and location

for(Rayon in seq(110000, 500000, by = 10000)) ## Loop for multiple scale
{
  resmanteltotal = NULL
  for(i in 1:length(XLAEAj)) ## Loop for multiple location
  {

#definition of sampling area (circle)
Xcircle <- ( XLAEAj [i] + Rayon*cos(seq(0,2*pi,length.out=100)))
Ycircle <- ( YLAEAj [i] + Rayon*sin(seq(0,2*pi,length.out=100)))
#extraction of coordinates in sampling area
expr <- point.in.polygon(XLAEAj, YLAEAj , Xcircle , Ycircle )
XLAEAj[expr==1]
YLAEAj[expr==1]
coord <- cbind (XLAEAj[expr==1], YLAEAj[expr==1])
CoordOrdre<- coord[order(coord[,1],decreasing=FALSE),]
```



```

locOrdre<-data.frame(CoordOrdre)

# Matrix of euclidian geographic distances
DisGeoEucl<-dist(locOrdre, method = "euclidean", diag = TRUE, upper = TRUE)
DisGeoEucl <- as.dist(DisGeoEucl)## head(DisGeoEucl)

# -A- Matrix of genetic disntances (=bray curtis dissimilarity)
listcoord = (1:992)[expr==1]
Genet = NULL
for(h in listcoord)
{
tmp <- matG2[(matG2[, 1]==XLAEAj[h])&(matG2[, 2]== YLAEAj[h]), ]## ajouter
3:118 après la dernière virgule pour sélectionner les colonnes voulues
Genet = rbind(Genet,tmp)
}
GenetOrdre<- Genet[order(Genet[,1],decreasing=FALSE),]
GenetOrdreSanscoord <- GenetOrdre[,-c(1,2)]
MatdistGenet<- vegdist(GenetOrdreSanscoord, method="bray", binary=FALSE,
diag=FALSE, upper=TRUE, na.rm = TRUE)
MatdistGenet <- as.dist(MatdistGenet)

# Matrix of distances of landscape resistance
spatiallocX <- locOrdre[,1]
spatiallocY <- locOrdre[,2]
SpaLoc <- SpatialPoints(cbind(spatiallocX, spatiallocY))
Comelev<- commuteDistance(Telev, SpaLoc)
Comelev<-as.dist(Comelev, diag = TRUE, upper=TRUE)

# Partial mantel tests
MantelpRes <- mantel.partial(MatdistGenet, Comelev, DisGeoEucl, method =
"pearson", permutations = 1000)
MantelpResinv <- mantel.partial(MatdistGenet, DisGeoEucl, Comelev, method =
"pearson", permutations = 1000)
resmanteltotal <- rbind (resmanteltotal, cbind (XLAEAj [i], YLAEAj
[i],length(XLAEAj[expr==1]), MantelpRes[3], MantelpRes[4], MantelpResinv
[3], MantelpResinv [4] ))

}
file <- paste("ResmantelIBRelevALLSPT", Rayon ,".txt", sep="")
write.table(resmanteltotal, file, col.names = FALSE )
}

# -B-

# For Commonality analysis, #
# replace the -A- to -B- part #
# by the following subscript #

# Matrix of genetic disntances (= bray curtis dissimilarity)
listcoord = (1:992)[expr==1]
Genet = NULL
for(h in listcoord)
{
tmp <- matG2[(matG2[, 1]==XLAEAj[h])&(matG2[, 2]== YLAEAj[h]), ]
Genet = rbind(Genet,tmp)
}
GenetOrdre<- Genet[order(Genet[,1],decreasing=FALSE),]
GenetOrdreSanscoord <- GenetOrdre[,-c(1,2)]
MatdistGenet<- vegdist(GenetOrdreSanscoord, method="bray", binary=FALSE,
diag=FALSE, upper=TRUE, na.rm = TRUE)
MatdistGenet <- as.dist(MatdistGenet)

```

```

MatdistGenetA <- as.matrix(MatdistGenet)
Matgenet3 <- melt(MatdistGenetA)[melt(upper.tri(MatdistGenetA))$value,]

# Matrix of distances of landscape resistance
spatiallocX <- locOrdre[,1]
spatiallocY <- locOrdre[,2]
SpaLoc <- SpatialPoints(cbind(spatiallocX, spatiallocY))
Comelev<- commuteDistance(Telev, SpaLoc)
Comelev3 <- as.matrix(Comelev)
DatafElevA <- melt(Comelev3)[melt(upper.tri(Comelev3))$value,]

# Commonality analysis
Data <- cbind (Matgenet3, DatafElevA)
names(Data) <- c( "X1", "X2", "DistGenet", "X1", "X2","E")
apsOut3=aps(Data,"DistGenet",list("E"))
G<-commonality(apsOut3)
sum(G[,1])

rescommonalitytotal <- rbind (rescommonalitytotal, cbind (XLAEAj [i],
YLAEAj [i],length(XLAEAj[expr==1]), sum(G[,1]) ))

}
file <- paste("RescommonalityALLSPT", Rayon ,".txt", sep="")
write.table(rescommonalitytotal, file, col.names = FALSE )
}

```

```

# (Article n°3) SPREAD MODEL PWN-infested populations across Pyrenees

# Definition of the grid / Projection: LAEA metric system
x <- seq(3410+0.5,3460-0.5,by=1)
y <- seq(2240+0.5,2290-0.5,by=1)
vectxy <- expand.grid(x,y)
nx = length(x)
ny = length(y)
dx = 1      # resolution of grid = 1 km
dy = 1

# Focus on Pyrenees
x <- seq(3250+0.5,3800-0.5,by=1)
y <- seq(2100+0.5,2390-0.5,by=1)
vectxy <- expand.grid(x,y)
nx = length(x)
ny = length(y)
dx = 1
dy = 1

# Load data
library(raster)
elev = raster("elev1kmlaea")
vectelev = extract(elev,cbind(vectxy[,1]*1000,vectxy[,2]*1000))
vectelev[is.na(vectelev)] = 0
matelev = matrix(vectelev,byrow=F,ncol=ny)
image(x,y,-
matelev,cex.axis=1.3,cex.lab=1.3,xlab="",ylab="",main="Elevation")

ppin = raster("pinuspin")
vectppin = extract(ppin,cbind(vectxy[,1]*1000,vectxy[,2]*1000))
vectppin[is.na(vectppin)] = 0
matppin = matrix(vectppin,byrow=F,ncol=ny)
image(x*1000,y*1000,-matppin,asp=1,xlab="",ylab="",main="Pinus pinaster")

psyl = raster("pinussyl")
vectpsyl = extract(psyl,cbind(vectxy[,1]*1000,vectxy[,2]*1000))
vectpsyl[is.na(vectpsyl)] = 0
matpsyl = matrix(vectpsyl,byrow=F,ncol=ny)
image(x*1000,y*1000,-matpsyl,asp=1,xlab="",ylab="",main="Pinus sylvestris")

poth = raster("pinemisc")
vectpoth = extract(poth,cbind(vectxy[,1]*1000,vectxy[,2]*1000))
vectpoth[is.na(vectpoth)] = 0
matpoth = matrix(vectpoth,byrow=F,ncol=ny)
image(x*1000,y*1000,-matpoth,asp=1,xlab="",ylab="",main="Other Pinus")

# addition of pine cover
matpine = matppin+matpsyl+matpoth
image(x*1000,y*1000,-matpine,asp=1,xlab="",ylab="",main="All pines")
image(x*1000,y*1000,1*(matpine>5),asp=1,xlab="",ylab="",main="Pines    >
5%",breaks=c(-1,0.5,1.5),col=c("white","green"))
image(x,y,-matpine,cex.axis=1.3,cex.lab=1.3,xlab="",ylab="",main="All
pines")

# EXPANSION MODEL, planimetric distance
i0=2 # Starting line
j0=2 # Starting column
P0 = 200 # Insect number at starting point

```

```

dif <- 6.480 # Diffusion coefficient
eps <- 2.17 # Growth rate
K=0.03*(matpine/100)*156000 # biotic capacity
min(K[K>0])
K[K==0] = 0.0001

POP = matrix(rep(0,nx*ny),ncol=ny)
POP[i0,j0]=P0
image(x,y,-POP,cex.axis=1.3,cex.lab=1.3,xlab="",ylab="",main="INITIAL
POPULATION") # Definition of starting point
nom <- heat.colors(11)

# EXPANSION MODEL, true surface distance (accounting for slope)

El1 = matelev[-1,]
El1 = rbind(El1,matelev[nx,])
El2 = matelev[-nx,]
El2 = rbind(matelev[1,],El2)
El3 = matelev[, -1]
El3 = cbind(El3,matelev[,ny])
El4 = matelev[, -ny]
El4 = cbind(matelev[,1],El4)

dXi = sqrt(((abs(El1-El2)/1000)^2) + (2*dx)^2)/2
dYi = sqrt(((abs(El3-El4)/1000)^2) + (2*dy)^2)/2
dX2 = 1/(dXi^2)
dY2 = 1/(dYi^2)

tp<-max(floor(2*dif*max(max(dX2)+max(dY2))))+1,1)

# Function (accounting for resident populations)

surf = function(N, valnem=4,valres=0.9,seuilpwn = 1)
{
  P = POP # density of infested vector beetle
  Knem=valnem*K # definition of ratio between K0 et Ki (=Knem)
  R = K*valres

  # Definition of variables (latency in recuitment of resident
populations)
# exprPWNn : cells infested at year n (after dispersal of year n) or
before
# exprPWNnn : cells newly infested at year n, but not before year n
# exprPWNn1 : cells infested at year n-1 (after dispersal of year n-1
) or before
# exprPWNnn1 : cells newly infested at year n-1
# exprPWNn2 : cells infested at year n-2 (after dispersal of year n-2
) or before

  exprPWNnn1= (P>seuilpwn ) # cells colonized at year n-1
  exprPWNn1 = (P>seuilpwn ) # cells alread colonized at year n-1 (or
before)

  for (n in 1:N)
  { print(n)

```

```

# SPREAD MODEL

for (ind in 1:tp)
{
  Mi1 <- P[-1,] # shift of values of the cell matrix for the loop
  Mi1 <- rbind(Mi1,rep(0,ny))
  Mi2 <- P[-nx,]
  Mi2 <- rbind(rep(0,ny),Mi2)
  Mi3 <- P[, -1]
  Mi3 <- cbind(Mi3,rep(0,nx))
  Mi4 <- P[, -ny]
  Mi4 <- cbind(rep(0,nx),Mi4)

  P= P + (dif*dX2/tp)*(Mi1+Mi2-2*P) + (dif*dY2/tp)*(Mi3+Mi4-2*P)

  P[1,]<-P[2,] # accounting for the edges of the cell matrix
  P[nx,]<-P[nx-1,]
  P[,1]<-P[,2]
  P[,ny]<-P[,ny-1]

  P[P<0]=0
  P[matelev>1580]=0 # empirical threshold for vector survival
  exprPWNNn = (P>seuilpwn )
  P[exprPWNNn & (P>Knem)]=Knem[exprPWNNn & (P>Knem)]
  P[(!exprPWNNn) & (P>K)]=K[(!exprPWNNn) & (P>K)]
}

# GROWTH MODEL and recruitment steps

# 1) Biotic capacity of non infested cells (K0) at year n

P[!exprPWNNn] = 0

# 2) Growth in newly infested cells at year n (biotic capacity =
Knem, but no recruitment of resident populations)

exprPWNNnn = (exprPWNNn) & (!exprPWNNn1)
P[exprPWNNnn] = P[exprPWNNnn] * (1+eps*(1-
((P[exprPWNNnn]+R[exprPWNNnn])/Knem[exprPWNNnn])))

# 3) Growth in cells infested the year before (n-1), (biotic capacity
= Knem, and recruitment of resident populations)

P[exprPWNNnn1] = (P[exprPWNNnn1] + R[exprPWNNnn1]) * (1+eps*(1-
((P[exprPWNNnn1]+R[exprPWNNnn1])/Knem[exprPWNNnn1])))

# 4) Growth in cells infested at year (n-2) or before

exprPWNNn2 = (exprPWNNn & (!exprPWNNnn) & (!exprPWNNnn1))
P[exprPWNNn2] = P[exprPWNNn2] * (1+eps*(1-
((P[exprPWNNn2])/Knem[exprPWNNn2]))) # resident population considered in P
(infested population)

exprPWNNnn1 = exprPWNNnn # print newly colonized cells as n-1 for
following iteration (year)
exprPWNNn1 = exprPWNNn

exprPWNNn = (P>seuilpwn )

```



```

# definition of maximum P: maximum density of infested cells = Knem
P[exprPWNn & (P>Knem)]=Knem[exprPWNn & (P>Knem)]

# definition of maximum P: maximum density of non infested cells = K
P[(!exprPWNn) & (P>K)]=K[(!exprPWNn) & (P>K)]
P[P<0] = 0

image(x,y,P,xlab="Longitude",ylab="Latitude",cex.axis=1.3,cex.lab=1.3,main=
paste("N = ",n),breaks=c(0,0.1,10,100,1000,10000,200000), col=c("white",
"green",nom[8],nom[6],nom[4],nom[1]), asp=1)
}
return(P)
}

Resultat<-surf(N=60, valnem=4,valres=2,seuilpwn = 1)

# Calculation of invaded area (km square)
sum(Resultat>1)

```


Dispersion des espèces impliquées dans une association phorétique vecteur - pathogène nouvellement formée: Le cas de *Monochamus galloprovincialis*, vecteur natif d'un nématode invasif en Europe (*Bursaphelenchus xylophilus*)

Résumé: Les invasions biologiques se sont intensifiées au cours des dernières décennies en raison d'une accélération des échanges commerciaux. Ces invasions représentent une menace pour les écosystèmes et de nombreuses activités anthropiques, il est donc crucial de comprendre les mécanismes qui les sous-tendent afin de mieux prévoir et limiter leurs impacts. Dans cette thèse, j'aborde la question du potentiel dispersif d'espèces natives et non natives impliquées dans une association phorétique nouvellement formée. En particulier, je me focalise sur le cas de l'association entre un nématode invasif ravageur des pinèdes, le nématode du pin (*Bursaphelenchus xylophilus*) et son insecte vecteur endémique en Europe (*Monochamus galloprovincialis*). J'ai tout d'abord étudié les flux de gènes de l'insecte vecteur seul afin d'identifier les barrières à sa dispersion. J'ai ensuite simulé l'expansion spatiale du couple nématode-vecteur à l'aide d'un modèle de dispersion, en intégrant l'effet synergique de cette nouvelle association. Les résultats obtenus au cours de cette thèse montrent qu'il existe un important potentiel de dispersion du nématode invasif en Europe par le biais de cette association phorétique. En revanche, certains paramètres de l'environnement tels que les reliefs et les températures basses qui leur sont associées, ainsi que les fortes densités en pins constituent des barrières à la dispersion du vecteur et donc des obstacles potentiels à l'expansion du nématode invasif. Au-delà des apports relatifs au modèle d'étude, cette thèse a conduit au développement de plusieurs méthodes pouvant être adaptées à d'autres cas d'associations phorétiques nouvelles et, par extension, contribuer à la compréhension de la dispersion des espèces au sein de ces systèmes complexes et peu étudiés.

Mots clés : invasion biologique, association phorétique, dispersion, phylogéographie, génétique spatialisée, modèle de dispersion.

Dispersal of species involved in a novel vector-pathogen phoretic association: The case of *Monochamus galloprovincialis*, native vector of an invasive nematode in Europe (*Bursaphelenchus xylophilus*)

Summary: Biological invasions dramatically increased over the last decades due to the intensification of international trade. These invasions constitute a threat for ecosystems and many anthropic activities, therefore it is crucial to understand underlying processes in order to better predict and manage their impacts. In this PhD thesis, I explore the potential of dispersion of native and non-native species involved in a novel phoretic association. I focus on the case of the association between a pest for pine forests, the pinewood nematode (*Bursaphelenchus xylophilus*) introduced in Europe, and its endemic insect vector (*Monochamus galloprovincialis*). I first estimated gene flows of the insect vector alone in order to identify the barriers and corridors to dispersal of this species. Then I have simulated the spatial spread of the nematode-vector couple using a spread model, and accounting for the synergistic effect of this novel association. The results obtained during this PhD showed that the invasive nematode has an important potential to spread through this phoretic association. However, some environmental features such as elevation, areas with low temperatures, and the high pine densities constitute barriers to dispersal of the vector and so, potential obstacles to the spread of the invasive nematode. Beyond these results focused on the model of study, this thesis has led to the development of several methods that may be adapted to other cases of novel phoretic association and, by extension, may contribute to a better understanding of dispersal of species involved in those complex and poorly known systems.

Keywords: biological invasion, phoretic association, dispersal, phylogeography, spatial genetics, spread model.