



UNIVERSITE D'ORLEANS

COLLEGIUM SCIENCES ET TECHNIQUES

ITP SCIENCES BIOLOGIQUE ET CHIMIE DU VIVANT

Master sciences biologie spécialité BOPE Année 2012 2013

**Dynamique de l'expansion spatiale de la  
processionnaire du pin *Thaumetopoea pityocampa*  
dans le bassin Parisien. Apports de la génétique et de  
l'écologie du paysage**

Simon LAUBRAY

Sous la direction de Jérôme Rousselet (unité zoologie, INRA Centre Val de Loire)

et Jean-Pierre Rossi (centre de biologie pour la gestion des populations INRA de Montpellier)

INRA - Centre Val de Loire, Unité de Recherche de Zoologie Forestière (URZF)  
2163, avenue de la pomme de pin, Ardon CS 40001 F-45075 Orléans CEDEX 2

**Résumé :** Il est prouvé qu'avec les changements globaux de nombreuses espèces connaissent une expansion naturelle vers le Nord. Cela s'explique bien souvent par la hausse des températures, mais il est rarement mentionné le rôle de la connectivité de paysage. L'étude des flux de gènes au niveau du front de colonisation d'une espèce en expansion telle que la processionnaire du pin (*Thaumetopoea pityocampa* Schiffermüller, Lepidoptera, Notodontidae) a mis en évidence des zones du paysage connectées et utilisées comme voies de colonisation. Pour cela, la variabilité génétique observée en région Centre (microsatellite) est mise en relation avec la variabilité paysagère au travers de la densité des pins. Afin de représenter au mieux la connectivité dans la région, une carte des arbres hors forêts (AHF) est fusionnée avec la carte des peuplements forestiers de l'Inventaire Forestier Nationale (IFN). La présence de trois lignées différentes entraîne une forte variabilité génétique masquant les relations paysage-flux de gène à large échelle. Les relations entre distance génétique et structure du paysage s'observent cependant à l'interface entre les principales lignées génétiques en contact dans la région. La fragmentation du paysage joue un rôle décisif vis-à-vis du processus de dispersion de la processionnaire du pin. Nos résultats montrent en outre qu'il est important de prendre en compte les arbres hors forêt dans la caractérisation des paysages.

**Mots clés :** flux de gènes, structure génétique, perméabilité, corridors, barrières

## **Abstract**

It is proven that the global changes cause a natural expansion to the north of a lot of species. This is often due the increase of the temperatures, but it is rarely mentioned the role of landscape connectivity. The study of gene flow at the colonization front of a species growing as the pine processionary moth (*Thaumetopoea pityocampa* Schiffermüller.) can highlight areas of the landscape connected and used as a means of colonisation. For this, the genetic variability observed in the Centre region (microsatellite) is confronted with the landscape variability through the density of pines. In order to best represent the connectivity in the region, a map of Trees outside forest (TOF) is merged with the map of forest stands of the National Forest Inventory. The presence of three different lines increases genetic variability distorting the results. But when there is no reproductive isolation, it appears that the meeting and the admixture rate of these lines depend on the landscape. It is shown that the fragmentation of the landscape is crucial for species dispersal and it is important to take into account the TOF in the dispersal of pest such as the pine processionary moth.

Key words :gene flow, genetic structure, permeability, corridors, barrier



## I. Introduction

Avec les changements globaux, l'aire de répartition de nombreuses espèces évolue. Le réchauffement climatique provoque une tendance générale des espèces à disperser vers le Nord (Walther et al., 2002). De nombreuses espèces du bassin méditerranéen ont ainsi une aire de répartition qui s'élargit vers des latitudes plus septentrionales. En effet, la hausse des températures lève une barrière thermique qui libère de nouvelles zones favorables pour ces espèces. Ces zones deviennent alors potentiellement colonisables, et le mode de dispersion ainsi que la connectivité du paysage vont déterminer la vitesse de l'expansion de l'espèce (Bialozyt et al., 2005, Taylor et al., 1993).

Les modes de dispersion peuvent être classés en deux grands types selon la proportion d'événements de dispersion à longue distance : la dispersion diffusive et stratifiée (Ibrahim et al., 1996., Klein et al., 2006). La dispersion diffusive se fait uniquement de proche en proche et conduit en milieu homogène à une avancée progressive du front de colonisation, alors que la dispersion stratifiée combine également des événements rares de dispersion à longue distance et conduit à une avancée de plus en plus rapide du front. Ce type de dispersion peut être due par exemple à une combinaison de transport actif et passif ou à la variabilité des capacités de dispersion active des individus.

En conditions naturelles la dispersion se réalise rarement en milieu homogène et la question de la vitesse d'avancée d'un front ne peut pas être réduite simplement à celle du type de dispersion en faisant abstraction du paysage. En réalité, le support sur lequel se fait la dispersion est hétérogène, et l'expansion prend place dans une matrice paysagère composée d'une mosaïque d'éléments de nature différente, plus ou moins favorables à l'espèce. Pour une espèce en expansion, comme pour les autres, le milieu présente des zones à capacité biotique très différente et le paysage se compose d'une alternance de corridors, de filtres et de barrières à la dispersion (Bergès, 2010). Ces barrières à la dispersion occasionnent des ruptures dans les flux de migrants et donc de gènes, les rendant détectables par des marqueurs génétiques (Manel et al., 2003).

Quantifier la dispersion des individus à de grandes échelles spatiales est très difficile, en particulier chez les insectes, et on utilise donc très souvent les outils de la génétique des populations pour retracer le processus et décrire les patrons de dispersion. C'est l'objet de la génétique du paysage qui est une discipline récente (le terme est employé pour la première fois en 2003 par Manel et al.). La génétique du paysage a pour objectif de relier les structures génétiques observées avec la structure paysagère.

Notre modèle d'étude est la processionnaire du pin *Thaumetopoea pityocampa* (PP), un des

plus importants défoliateurs du pin en Europe. D'origine méditerranéenne, cette espèce est aujourd'hui en expansion et son aire de répartition s'étend en latitude (monte vers le nord de 0,27° dans le Bassin Parisien) et en altitude (un gain vertical de 110-230m dans les Alpes italiennes) (Battisti et al 2005). Au cours des vingt dernières années, la superficie de son aire en France a augmenté de près d'un tiers, avec près de 100 000 km<sup>2</sup> gagnés, dont une grande partie dans le sud du Bassin Parisien.

Les causes combinées de son expansion sont le réchauffement climatique, la disponibilité croissante des arbres hôtes et le transport involontaire par l'être humain.

– La hausse des températures hivernales observée depuis le début des années 1990, a entraîné une extension de l'aire de distribution de la processionnaire du pin vers le nord et en altitude (Battisti et al., 2005 ; Robinet et al., 2007). La progression s'est faite à une vitesse de 5,6 km/an en région Centre (Battisti et al.,2005; Robinet et al., 2007) et le front d'expansion atteint maintenant le sud de l'Ile-de-France (Robinet et al.,2011).

– Les pins *Pinus spp* sont des arbres qui ont été fortement utilisés pour le reboisement de la France à la fin du XIXème et au XXème siècles. Aujourd'hui, ils sont également de plus en plus utilisés comme arbres d'ornement. Cela offre à la PP une grande disponibilité en arbres-hôtes dans des milieux non forestiers (urbains, champêtres, prairiaux, ...). L'étude du rôle de ces Arbres Hors Forêt (AHF) dans la propagation de cet insecte et sa circulation entre les écopaysages de la région Centre font l'objet du projet ADRIEN 2012-2014 (Arbres Dispersés et Rôle dans les Invasions d'Espèces Nuisibles) dans lequel mon stage s'inscrit.

– On trouve également des foyers d'invasions de l'espèce à de longues distances du front de colonisation, en région parisienne et en Alsace. Ce sont des foyers créés par introductions humaines involontaires via le commerce des plants de pin (Robinet *et al.*, 2012).

Dans le cas de la PP, des observations de terrain ont suggéré que l'espèce progresse plus rapidement dans des paysages composés de milieux ouverts de type Beauce (donc *a priori* défavorables à cette espèce forestière en matière de capacité biotique) où la ressource est rare et l'habitat fragmenté. Inversement, toujours selon ces observations, les forêts présentant de grandes surfaces de peuplements denses de pins, comme la forêt d'Orléans, auraient freiné l'expansion de l'insecte. Ceci pourrait s'expliquer par le comportement généralement philopatric de femelles qui ne disperseraient loin de la zone natale que dans des milieux à capacité biotique faible. Nous avons ainsi formulé l'hypothèse suivante :

- En forêt de pin, la dispersion se ferait de proche en proche. La forte densité de pins disponibles entraînerait alors un ralentissement de la progression géographique de l'espèce. Une fois l'obstacle contourné et les territoires en aval colonisés, les flux de gènes de par et

d'autre resteraient limités, et la structure génétique des populations pourrait en garder une empreinte détectable avec des marqueurs hyperpolymorphes (si nombre et variabilité suffisante).

- En milieu ouvert, les principaux éléments boisés sont sous forme d'arbres hors forêt (AHF) et distribués sur de très grandes surfaces. La densité en arbres est donc beaucoup plus faible qu'en peuplement forestier. Ceci se traduirait par un habitat à faible capacité biotique mais favorable à la dispersion. Les AHF joueraient probablement un rôle important dans la dispersion des bioagresseurs mais ceci a rarement été étudié en détail. Dans des paysages fragmentés, la distance entre *patches* d'habitat favorable nécessite une dispersion à plus longue distance qu'en milieu continu avec un avantage important pour les individus bons disperseurs. Ce mode de dispersion stratifiée engendrerait une augmentation de la vitesse de colonisation (Ibrahim et al., 1996., Klein et al., 2006), avec comme tremplin les AHF .

La génétique du paysage constitue un outil intéressant liant la puissance de la génétique des populations pour caractériser les flux de gènes avec la résolution de l'écologie du paysage pour identifier les sources d'hétérogénéité environnementale affectant la dispersion. Le travail présenté dans ce rapport repose sur cette approche appliquée à l'étude de l'expansion de la PP dans la région Centre où existe des paysages variés, notamment en matière d'abondance en plantes-hôtes de la processionnaire. Mon stage a consisté à analyser les génotypes à 19 loci microsatellite de plus de 1000 individus qui y ont été collectés selon une maille régulière et à déterminer si le paysage peut rendre compte d'une partie de la structure génétique observée.

## **II. Matériels et méthodes**

### **2.1 Plan d'échantillonnage**

Durant l'hiver 2007/2008, 1287 individus ont été prélevés directement sur l'arbre hôte (*Pinus* spp). A cette période, la PP est au stade chenille et se développe dans un nid blanc et cotonneux que l'on trouve au sommet des branches. Le plan d'échantillonnages (annexe 1) est régulier et suit une grille couvrant 65536 km<sup>2</sup> dans le bassin parisien soit 20% de son aire de répartition. La grille est composée de cellules de 16 km sur 16 km. Cinq à quinze individus ont été prélevés par cellule. Les individus ont été géo référencés et l'espèce de l'arbre hôte répertoriée.

### **2.2 Acquisition des données génétiques**

Les étapes d'extraction et d'amplification des microsatellites ont été faite de novembre 2011 à février 2012. La lecture des génotypes, l'attribution des allèles et l'harmonisation des données ont été effectuées durant ce stage.

### **2.2.1 Extraction de l'ADN**

L'ADN a été préalablement extrait à partir de tubes de Malpighi et de fragments d'intestins de X individus. L'ADN a été isolé et purifié en suivant le protocole du Kit « NucleoSpin® Tissu » de Macherey-Nagel.

### **2.2.2 Amplification des microsatellites.**

A partir de l'ADN génomique extrait, 19 microsatellites ont été amplifiés à l'aide d'amorces spécifiques (Rousselet et al., 2004. , Santos et al., 2007. , Burban et al., 2012.). Pour chaque locus, les amorces ont été marquées à l'aide de fluorochromes de différentes couleurs afin de permettre au séquenceur de les lire. Ainsi, un locus a une couleur bien défini, ce qui permet d'amplifier, par *Polymerase Chain Reaction* (PCR), plusieurs *locus* dans un même puits (PCR en multiplex). 1 µl de produits PCR a alors été déposé dans un des puits de la plaque avec un mélange de formamide et d'un marqueur de taille 600liz (respectivement 10 µl et 0.4 µl) pour obtenir un volume total de 11,4 µl.

La taille des fragments fluorescents a ensuite été déterminée par le séquenceur automatique *Abi Prism® 3100 Genetic Analyser* (Applied Biosystems). On obtient un graphique avec le niveau de fluorescence en fonction de la taille du fragment. Le passage d'un fragment devant le capteur du séquenceur crée un pic de fluorescence

### **2.2.3 Lecture des fragments**

Les microsatellites sont des marqueurs utilisés en génétique pour leur polymorphisme de taille. Ce sont des motifs répétés en tandem qui acquièrent leur variabilité par des mutations qui vont affecter le nombre de répétitions, un nombre de répétition donne un allèle. Le nombre variable de répétitions se traduit par des variations de tailles de fragment détectables par le séquenceur à l'aide des fluorochromes fixés sur l'amorce. Le séquenceur détecte la fluorescence et génère un graphe avec un pic de fluorescence lors du passage d'un fragment devant la lentille.

Les tailles de fragments obtenues sont lues à l'aide du logiciel *Genemapper®*. Lors de cette lecture, il y existe un écart entre les tailles attendues et les tailles observées du aux fluorochromes ou aux capillaires qui vont modifier l'activité électrophorétique. Cet écart est

variable d'un fragment à un autre, et chaque allèle a une gamme de tailles dans laquelle on trouve le pic de fluorescence. De plus, les données ont été acquises avec deux séquenceurs, deux marqueurs de taille et trois opérateurs différents générant trois jeux différents. Il a fallu homogénéiser les données afin de les fusionner en un seul. Pour cela il a fallu convertir la longueur des fragments en nombre de répétitions des motifs. La procédure a été de séquencer et génotyper des individus témoins et définir des bornes fixant la gamme de taille de chaque allèle.

Séquençage : amplification des fragments d'intérêts. Cette fois-ci, il y a eu une réaction PCR par locus et par puits selon le même protocole (Rousselet et al., 2004. , Santos et al., 2007. , Burban et al., 2012.). Après vérification sur gel d'agarose 3%, les produits PCR ont été purifiés sur colonne en suivant le protocole « NucleoSpin Extract II » de Macherey-Nagel. Ensuite, la réaction de séquence s'est fait suivant le protocole Abi Prism® Big Dye® Terminator V.3.1 Cycle Sequencing Kits (Applied Biosystems) en utilisant les amorces propres à chaque locus et des didésoxyribonucléotides marqués. Les fragments d'ADN obtenus ont été précipités à l'éthanol pour éliminer tous les débris résiduels de la réaction de séquences pour ensuite être dilués avec 30µl d'eau ultrapure. L'ADN en solution est passé au séquenceur automatique.

Calibrage : génotypage d'individus hétérozygotes, lus avec le marqueur de taille 500 liz ainsi que d'individus hétérozygotes lus avec le marqueur de taille 600liz. Les hétérozygotes choisis étaient de tailles proches. Ces individus ont été lus sur le même séquenceur avec le marqueur de taille 600liz. Ce qui a permis de déterminer le «degré » de décalage des pics.

#### **2.2.4 Analyse des données génétiques**

L'analyse des données génétiques a permis de visualiser les flux de gènes au travers d'indices de structurations. Lorsque les flux de gènes sont importants, la structuration des populations est plus faible. C'est-à-dire qu'à un certain taux de migrant la différenciation entre les populations diminue. L'isolation par la distance est un principe de caractérisation des flux de gènes. Ce principe montre que le flux de gènes est important entre deux populations proches créant peu de différenciation entre ces populations, mais au fur et à mesure que la distance augmente le flux de gènes diminue provoquant petit à petit de la différenciation. La mesure de cette différenciation, ce fait deux à deux. Pour estimer la différenciation entre deux populations, le logiciel *Genetix* (v 4.05.2) ( Belkhir *et al.*,) permet de calculer un  $F_{st}$  apparié (Weir et Cockerham 1984). Le  $F_{st}$  est une comparaison des fréquences d'hétérozygotes en considérant que les deux populations en forment une seule (flux de gènes maximale) avec les



fréquences observées au sein des populations. Plus le  $F_{st}$  est différent de 0, plus les fréquences des deux populations sont différentes de la fréquence de l'ensemble. Ce qui signifie qu'il y a une structuration entre les populations (flux de gènes limités). *Genetix* est capable de calculer ces  $F_{st}$  avec N populations. Les résultats sont affichés sous forme d'une matrice donnant les  $F_{st}$  entre les différentes populations. Pour valider la significativité des  $F_{st}$ , le logiciel permute les individus (et leur génotype) un grand nombre de fois (10 000). Cela crée des populations au hasard avec de nouvelles associations d'individus dont le  $F_{st}$  est calculé à chaque fois. Le but de ces permutations est de casser les relations entre individus du jeu de données en estimant que le hasard détruit la structure. Si 5 % des  $F_{st}$  créés par le hasard sont supérieurs au  $F_{st}$  réel, les populations ne sont pas structurées.

*Genetix* permet également de calculer des indices de dissimilarité entre individus tel que la distance de Nei 1978 (Nei 1978). Cet indice estime la distance entre deux individus à partir de différences génétiques. Plus les individus ont de différences dans leur génotype (moins ils partagent d'allèle commun) plus ils sont distants. Le résultat est une matrice donnant la distance génétique entre les individus.

Afin de visualiser graphiquement la structuration spatiale, les individus ont été répartis en *cluster* à l'équilibre de Hardy-Weinberg (HW) grâce au logiciel *Structure*® (v 2.3.4) (Pritchard et al. (2000a)). En utilisant comme hypothèse que l'ensemble des individus est réparti en K *clusters* à l'équilibre. Le logiciel va assigner les individus de façon à ce que les *clusters* formés soient à l'équilibre. Le logiciel fonctionne de façon probabiliste et sur permutation. Un grand nombre de permutations permet de faire un grand nombre d'associations d'individus afin de trouver celle qui est la plus proche de l'équilibre. Le nombre d'itérations permet d'effectuer le calcul avec plusieurs associations d'individus de départ. A partir de là, les choix des paramètres de calculs ont été : fréquence allélique indépendante, K compris a priori entre 1 et 10, 100 000 permutations et 10 itérations.

Pour chaque individu le logiciel attribue une probabilité d'appartenir à chaque *cluster*. Grâce au géo référencement des individus, les probabilités d'appartenance ont ensuite été projetées sur la carte de la région Centre à l'aide du logiciel Mapinfo®.

## **2.3 Cartographie de la densité en arbres hôtes dans la région centre**

### **2.3.1 Carte IFN (Inventaire Forestier National)**

L'IFN a réalisé et met à jour une carte de la couverture du sol selon le type de formation végétale, les essences et la densité du couvert forestier. Cette carte est établie sur la base d'unités de mesure de 2,25 ha (carré de 150 m de côté).

Nous avons utilisé ces données fournies par l'IFN sous la forme d'un *raster* c'est-à-dire une grille géo référencée dans laquelle chaque cellule ou pixel est une surface carrée de 250 m de côté.

Les cartes retenues sont celles représentant la répartition des pins noirs (Laricio et Noir), du pin sylvestre, du pin à crochets du pin cembro et du pin mugho. Ces arbres sont les hôtes préférentiels de la processionnaire.

### **2.3.2 Carte préliminaire des arbres hôtes isolés**

Les difficultés inhérentes à l'inventaire des AHF (faible densité, distribution spatiale complexe et mal connue) ont été contournées en élaborant un modèle statistique reliant la probabilité d'observer un AHF avec différentes variables décrivant paysage local. Ce travail est actuellement en cours dans le cadre du projet régional ADRIEN.

Nous avons utilisé les sorties d'un modèle préliminaire reliant l'occurrence d'AHF et la distance au plus proche élément bâti. Les données décrivant la distribution du bâti (bâtiments agricoles, administratifs, industriels, etc.) nous ont été fournies par l'IGN. Le modèle a été construit à partir de données de terrain consistant en un recensement exhaustif de tous les éléments boisés comportant un arbre ou plus (arbre, linéaire, bosquet ou une petite plantation) dans une surface de 22km × 22km dans la Beauce.

Le modèle liant la présence des AHF à la distribution du bâti a été construit à l'aide du logiciel R. Le modèle ajusté a ensuite été extrapolé à l'ensemble de la région centre. Il s'agit d'une approche préliminaire qui sera améliorée lorsque le modèle sera ajusté dans différents types de paysage de la région (travail en cours dans le cadre du projet régional ADRIEN et du méta-programme INRA SMaCH).

### **2.3.3 Modèle de conductance (ou de perméabilité)**

Avec l'aide du logiciel R et du *package gdistance* les cartes représentant la densité des pins (IFN et AHF) sont croisées pour former une carte complète de la région prenant en compte à la fois les éléments boisés et les arbres isolés. A partir de cette carte on peut créer une modélisation de la perméabilité du paysage à la progression de la processionnaire. La modélisation de la perméabilité est construite en suivant les principes des circuits électriques. De la même façon qu'un corps conduit plus ou moins bien le courant électrique en fonction de sa résistance, le paysage régule les flux d'organisme (et donc de gène) en fonction de sa composition (MacRae et al., 2008). Cela se traduit par la perméabilité (facilité de l'organisme à traverser éléments du paysage). La perméabilité peut être définie comme l'inverse du coût

fourni par l'organisme pour traverser le paysage. Ainsi plus le coût est faible, plus la perméabilité est forte. Il a fallu alors définir comment le coût de déplacement évolue dans le paysage en fonction de la densité en pin.

Des observations de terrain montrent que les adultes seraient plutôt phylopatriques (surtout les femelles). Il y aurait donc un phénomène de rétention (sédentarisation) dans les zones de ressources abondantes comme les lisières des peuplements de pin (densité en pin élevée). De plus, il existe un fort effet lisière chez la PP (Samalens & Rossi, 2011), le nombre de nids est beaucoup plus important en bordure de peuplement, ce qui traduit une faible dispersion vers le centre du peuplement. On peut en déduire que la perméabilité est faible dans un milieu avec une forte densité en pin. La traversée d'un élément du paysage dense en pin est donc coûteuse.

Au contraire, dans les zones où la ressource est présente ponctuellement (arbres isolés), la dispersion devient rapidement obligatoire. La faible densité en pin engendre une dispersion plus importante. On peut en déduire qu'un milieu avec une faible densité en pin à une perméabilité plus forte. En termes de coût, la traversée d'un élément du paysage avec peu de pin est moins dépensière.

Nous traduisons ces hypothèses dans un modèle définissant le coût du passage entre deux cellules (pixel) adjacentes par la fonction :  $F(x) = 1/(v_{\max}(x) - X_1 + X_2)$

Avec  $F(x)$  : taux de perméabilité de la cellule 1 vers la cellule 2

$X_1$  : densité en pin de la cellule d'origine

$X_2$  : densité en pin de la cellule voisine

$v_{\max}$  : valeur maximum entre  $X_1$  et  $X_2$

Ainsi, le coût est faible dans le sens :

- Faible densité en pin vers faible densité en pin (dispersion forcée)
- Faible densité en pin vers forte densité en pin (ressources plus abondante)

Et le coût est fort pour le passage dans le sens :

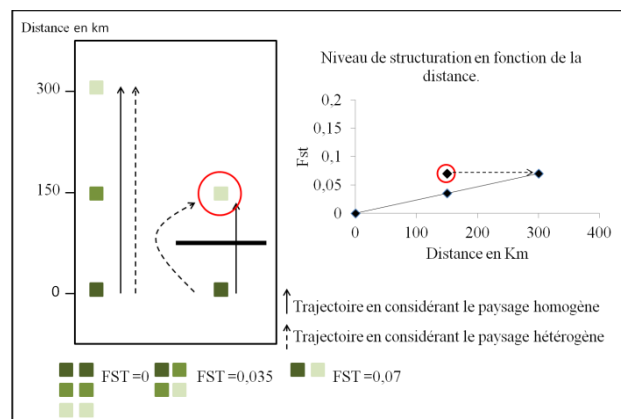
- Forte densité en pin vers forte densité en pin (effet lisière)
- Forte densité en pin vers faible densité en pin. (femelle à tendance phylopatrique)

Le modèle de perméabilité est alors l'inverse de cette fonction.

## 2.4 Génétique du paysage

Les analyses de génétique du paysage consisteront à vérifier la corrélation entre la structure paysagère et la structure génétique. Dans le cas où le milieu est homogène, la matrice de distance génétique est corrélée avec la matrice de distance euclidienne. Il y a un phénomène d'isolation par la distance entraînant une perte de variabilité avec l'avancée du front. Dans un milieu hétérogène, les différences de flux de gènes coïncident avec les éléments du paysages (corridors, filtres et barrières) ce qui est mis en évidence par une corrélation entre la matrice de distance génétique et la matrice de distance paysagère.

### 2.4.1 La différenciation génétique et le paysage



**Figure 1.** Schéma de la différenciation des populations en fonction de la distance. Comparaison entre un paysage homogène et un paysage hétérogène.

Le paysage est un ensemble hétérogène créant de la structure en régulant la perméabilité des flux de gènes :

- D'une part, les corridors permettent un fort flux de gènes entraînant une faible différenciation et donc la distance génétique correspond à celle que l'on trouverait en milieu homogène.

- D'autre part, les barrières et les filtres réduisent fortement ou partiellement les flux de gènes entraînant une différenciation importante de part et d'autre de la barrière. Dans ce cas la distance génétique est supérieure à celle qu'on attend en milieu homogène.

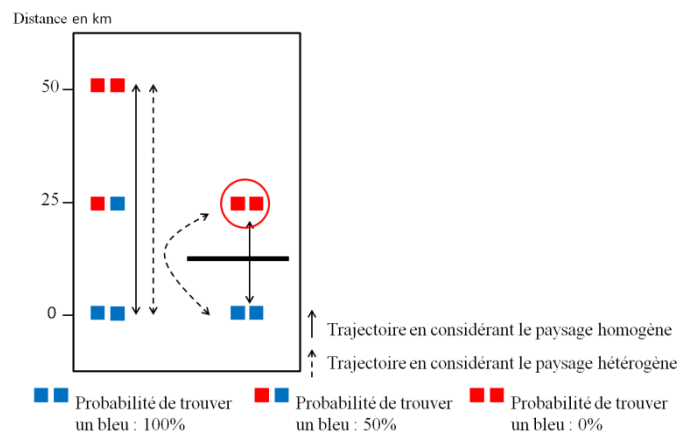
Les variations de flux de gènes au sein du paysage sont traduites en matrice de distance génétique. Cette matrice représente la distance de Nei 1978 entre les individus .

L'hétérogénéité du paysage est transformée en matrice de distance paysagère. A partir de la modélisation de la perméabilité du paysage, des trajectoires de moindre coût sont calculées entre les points d'échantillonnage ayant fourni les données génétiques (selon l'algorithme de

Dijkstra *package gdistance* (Etten 2012)). Les coûts obtenus entre deux points correspondent à la distance paysagère séparant ces points. Les distances de moindre coût quantifient la perméabilité du paysage à la dispersion de la PP.

Dans le cas d'un paysage homogène, la trajectoire de moindre coût correspond à une trajectoire en ligne droite. Dans ce cas, le coût correspond à la distance euclidienne. Le *package R raster* permet de construire une matrice de distance euclidienne qui contiendra la distance à vol d'oiseau séparant 2 points.

## 2.4.2 L'admixture et le paysage



**Figure 2.** Schéma de l'admixture de deux lignées. Comparaison entre paysage homogène et paysage hétérogène.

La région Centre est la zone de contact de deux lignées maternelles provoquant des taux de différenciation plus élevés que ceux que l'on pourrait trouver avec une seule lignée.

Dans ce cas, les corridors ont potentiellement un taux d'admixture plus fort que des zones séparées par des barrières. C'est-à-dire que le mélange entre les lignées augmente lorsqu'elles progressent plus facilement l'une dans l'autre dans un paysage où la progression est facilitée. Nous avons testé cette hypothèse en confrontant le niveau de perméabilité au taux de diffusion d'une lignée dans l'autre, sous réserve qu'il n'y ait pas d'isolement reproducteur. Pour cela, de nouvelles matrices de distance génétique ont été calculées à partir des probabilités d'appartenir à un *cluster*, déterminées par la méthode implémentée dans le logiciel *Structure*. En cas de corridors les probabilités d'appartenance à un *cluster* diminuent progressivement avec la distance, en cas de barrières ou de filtres elles diminuent brutalement. Ce qui se vérifie par une corrélation entre matrice de génétique et matrice de distance paysagère et une absence de corrélation significative avec la distance euclidienne.

Dans les 2 cas, la comparaison des matrices de données génétiques et celle des coûts est une façon de confronter les données génétiques aux données paysagères. Ceci a été réalisé en appliquant les tests de Mantel simples et partiels (Mantel, 1967) sur ces matrices de distance.

### **III. Résultats**

#### **3.1 Diversité génétique**

Données manquantes : lors du génotypage, il est apparu que plusieurs microsatellites posaient problème. Parmi les 19 locis génotypés 11 ont été maintenus.

Pour certains individus les données sont incomplètes. Parmi 1287 individus, 1026 ont été retenus pour les analyses (soit 80%). Les 20% d'individus écartés sont les individus ayant plus de 25% de données manquantes (annexe 2). 18 mailles ont été éliminées car elles ne possèdent aucun individu (annexe3). Les analyses génétiques ont porté sur tous les individus, pour les analyses de génétiques du paysage, seuls les individus des mailles comprises dans la région Centre ont été retenus.

L'harmonisation des trois jeux de données a été vérifiée en calculant des FST et des FIS dans des zones contenant des individus issus de jeux de données différents (résultats non communiqués). Ces indices ne montrent pas de déséquilibres dus à l'opérateur. De plus, la structure révélée n'apparaît pas souffrir de biais.

Pour avoir une idée du jeu de donnée globale les calculs de génétiques de populations classiques ont été faits sur les individus de la grille 64 x 64km. L'avantage de cette échelle est qu'il y a 15 individus par maille, aux autres échelles il n'y en a que 5. Les mailles 002, 014, 202 et 251 ont servi de population représentative pour les comparaisons des niveaux de structuration.

Les FST ont été calculés avec 15 individus ce qui est un effectif faible, mais le but de cette première analyse est une description du jeu de données selon des indices de génétique des populations classiques. Le FIS de 9 mailles sur 13 est non significativement différent de 0. Ce qui signifie que les populations sont à l'équilibre de HW. Les 4 autres populations ont un FIS > 0, ce qui signifie que ces populations ont un déficit d'hétérozygote traduisant une variabilité intrapopulation plus faible.

La valeur minimum de FST est 0 et la valeur maximum est 0,25. Cette variation de FST montre qu'il existe plusieurs niveaux de structuration. Un FST égal à 0 correspond au niveau de différenciation entre 2 populations proches géographiquement, et montre qu'il n'y a pas de

structuration à de faible distance et lorsque la distance augmente, les FST augmentent jusqu'à de fortes valeurs de structuration. Le FST maximum atteint une valeur supérieure à celle attendue à notre échelle. Sous l'effet de l'isolation par la distance, le FST maximum attendue en région Centre était de 0,07 à la distance maximale (Rousselet in prep). Or on observe un FST de 0,25 correspondant au niveau de structure entre des populations distantes de 900 km (Rousselet in prep ).

Le FST entre la maille 002 et 014 est égal à 0,12 et le FST entre la maille 002 et 251 est égale à 0,20. Ces deux FST sont supérieures au FST maximale attendu, la maille 002 présente une forte structuration avec les mailles du coin Sud Est et les mailles du coin Nord Ouest. Cependant le FST entre la maille 002 (coin Sud-Ouest) et 202 (coin Nord Est) est égal à 0,07. Ce FST est proche de celui attendu et montre un niveau de structuration correspondant à l'échelle d'étude.

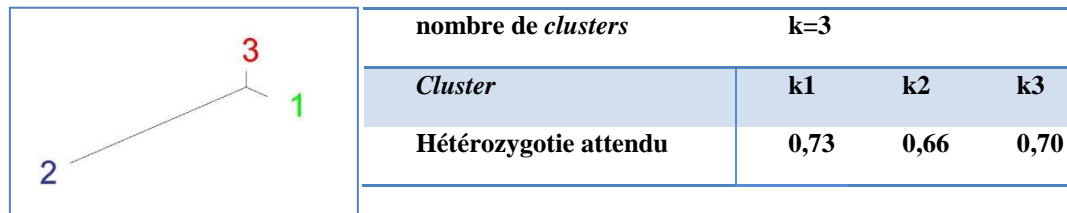
**Tableau I** Double demi-matrice représentant le Fst et la distance de Nei (1978) entre les populations des mailles de la grille 64 x 64km. La diagonale donnent les valeurs de Fis au sein de chaque maille, au dessus de la diagonale on trouve les valeurs de la distance de Nei et en dessous on trouve les valeurs de Fst. (NS) indique les valeurs qui sont non significativement différentes de 0.  $\alpha$  indique les mailles de la grille 64 x 64 km comprises dans la région Centre.

FST\NEI		002	006 $\alpha$	010	014	066 $\alpha$	070 $\alpha$	074 $\alpha$	078 $\alpha$	133 $\alpha$	194	198 $\alpha$	202	251
002	15	0,02ns	0,20	0,34	0,33	0,05ns	0,17	0,21	0,17	0,18	0,33	0,16	0,19	0,62
006 $\alpha$	15	0,08	0,06ns	0,15	0,27	0,15	0,05	0,07	0,23	0,08	0,40	0,08	0,02ns	0,75
010	15	0,11	0,07	0,12	0,05	0,22	0,23	0,15	0,13	0,14	0,27	0,13	0,18	0,46
014	17	0,12	0,11	0,02ns	0,12	0,35	0,34	0,30	0,18	0,23	0,30	0,18	0,32	0,35
066 $\alpha$	3	0,02ns	0,08ns	0,08ns	0,13	-0,01ns	0,13	0,11	0,19	0,13ns	0,32	0,21	0,12	0,55
070 $\alpha$	15	0,06	0,02ns	0,09	0,13	0,05ns	0,10	0,09	0,21	0,11	0,47	0,14	0,07	0,90
074 $\alpha$	13	0,08	0,03ns	0,06ns	0,12	0,04ns	0,04ns	0,06ns	0,13	0,02ns	0,28	0,07	0,06	0,63
078 $\alpha$	14	0,07	0,10	0,05	0,08	0,08ns	0,08	0,06	0,09	0,10	0,22	0,20	0,27	0,45
133 $\alpha$	15	0,06	0,04ns	0,05	0,09	0,04ns	0,04ns	0,01ns	0,04	0,06ns	0,18	0,04ns	0,07	0,45
194	15	0,10	0,14	0,10	0,11	0,10	0,14	0,10	0,08	0,06	0,05ns	0,14	0,39	0,16
198 $\alpha$	11	0,06	0,04ns	0,05ns	0,07	0,06ns	0,05ns	0,03ns	0,07	0,01ns	0,05ns	0,01ns	0,04ns	0,41
202	15	0,07	0,01ns	0,07	0,12	0,04ns	0,03ns	0,02ns	0,10	0,02ns	0,12	0,02ns	0,02ns	0,77
251	5	0,20	0,25	0,18	0,15	0,21	0,25	0,22	0,18	0,16	0,07ns	0,15ns	0,23	0,14ns

### 3.2 Création de *cluster*

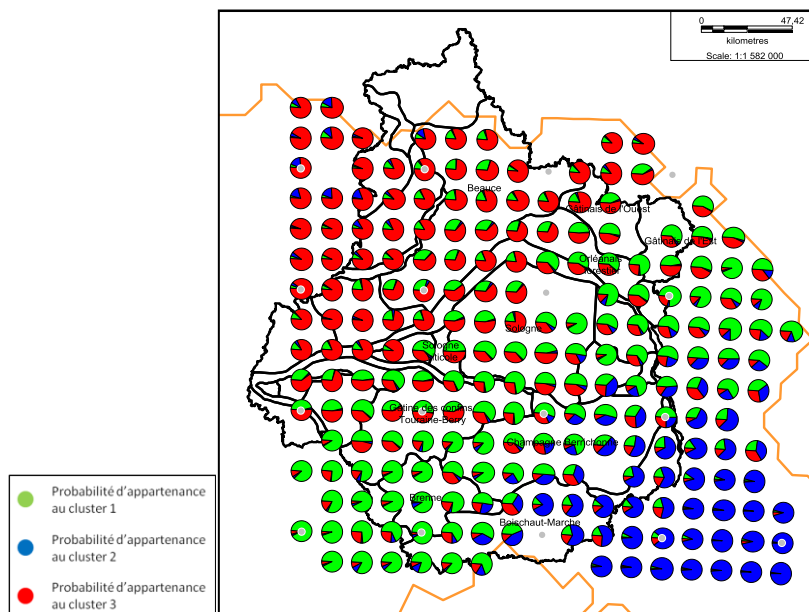
La courbe des estimations du  $\ln P(X|K)$  montre un optimum pour  $K=6$ , ce qui signifie que  $K=6$  est le plus haut niveau de structure du scénario testé. A ce niveau les *clusters* ont des différenciations maximales (annexe 4). Pour le reste des analyses le nombre de *clusters*

retenus est  $K=3$ , les arbres des autres  $K$  ( $K=2$  à  $K=6$ ) ainsi que leur projection se trouve en annexe 6



**Figure 3.** Arbre représentant le niveau de différenciation entre les *clusters* ainsi que le tableau de l'hétérozygotie attendu au sein des *clusters* pour  $K=3$ .

L'arbre est construit à partir du niveau de différenciation entre les *clusters* (figure 3). La taille des branches est liée à la différenciation entre les *clusters*. Plus cette taille augmente, plus la différenciation est importante. Cet arbre montre qu'il y a une différenciation faible entre les *clusters* 1 et 3 car la taille des branches est petite, alors que la grande taille de la branche du *cluster* 2 illustre une différenciation importante du *cluster* 2 avec les *clusters* 1 et 3.



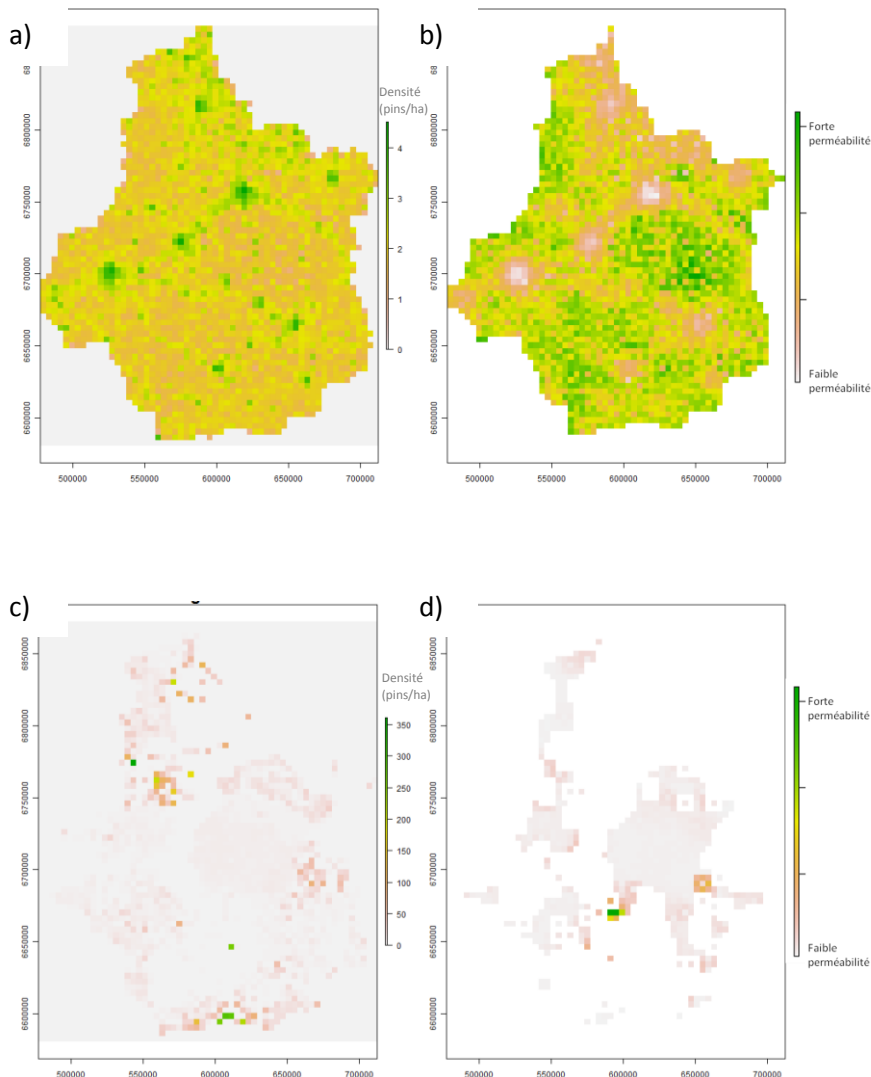
**Figure 4.** Grille d'échantillonnage couvrant le bassin parisien avec une maille de 16 km de côté. Les diagrammes circulaires représentent la probabilité des individus de la maille d'appartenir à un *cluster*. Cette probabilité est calculée à l'aide de *Structure* selon les paramètres suivants : fréquences indépendantes et 100 000 permutations. Un diagramme représente une maille, une couleur représente un *cluster* créée par *Structure*. Les points gris représentent les mailles de la grille 64km x 64 km.

Le logiciel *Structure* a inféré 3 *clusters* bien structurés spatialement. Dans le coin Sud-Est de la grille la probabilité d'appartenir au *cluster* « bleu » est forte. De même, la probabilité d'appartenir au *cluster* « vert » et au *cluster* « rouge » est plus forte respectivement dans la Sud-Ouest et dans le Nord-Ouest de la grille. Les zones de contact entre les *clusters* sont des zones où la probabilité d'appartenir à l'un ou l'autre est variable selon les mailles. Ce qui



traduit un certain degré de mélange des *clusters*. La surface d'admixture entre le *cluster* « rouge » et le *cluster* « vert » est plus étendue que la surface d'admixture entre le « vert » et le « bleu ». Le *cluster* bleu semble être cantonné au Sud-Est de la grille alors que les deux autres *clusters* se sont étendus plus largement.

### 3.3 Diversité paysagère et création de modèles de perméabilité



a) Carte des pins isolés de la région Centre obtenue à partir du modèle liant la présence des AHF à la distribution du bâti. Le vert représente les plus fortes densités en arbres isolés, le blanc représente la densité la plus faible.

b) Modélisation de la perméabilité du paysage à partir de la densité en pin obtenu par la carte a) et obtenue par la formule  $F(x) = 1/(V_{\max}(x) - X_1 + X_2)$  avec  $F(x)$  : taux de perméabilité de la cellule 1 vers la cellule 2 ;  $X_1$  : densité en pin de la cellule 1 ;  $X_2$  : densité en pin de la cellule 2 voisine ;  $V_{\max}$  : valeur maximum entre  $X_1$  et  $X_2$ . La modélisation ne prend en compte que les arbres isolés. La perméabilité est faible lorsque le pixel est blanc, et elle devient forte lorsque le pixel est vert.

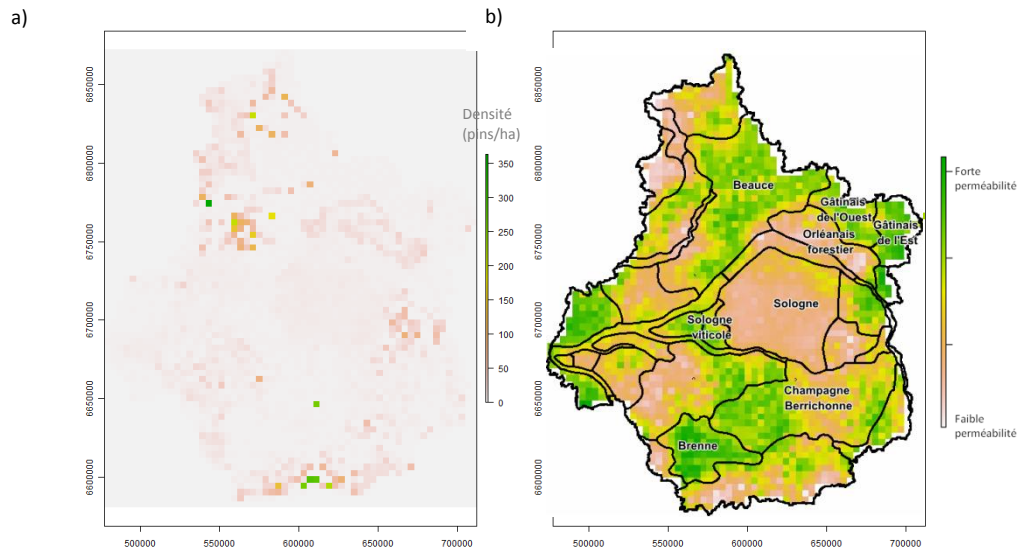
c) Carte de l'IFN des peuplements de pins supérieurs à 6,25ha en région Centre. La densité est représentée par l'échelle de couleur allant du blanc au vert. Le blanc traduit une absence de pin, le vert représente les fortes densités.

d) Carte de la modélisation de la perméabilité, utilisant la même formule que la carte b) à partir des données de la carte IFN 3.

**Figure 5** Représentation des pins et modèle de perméabilité AHF et IFN.

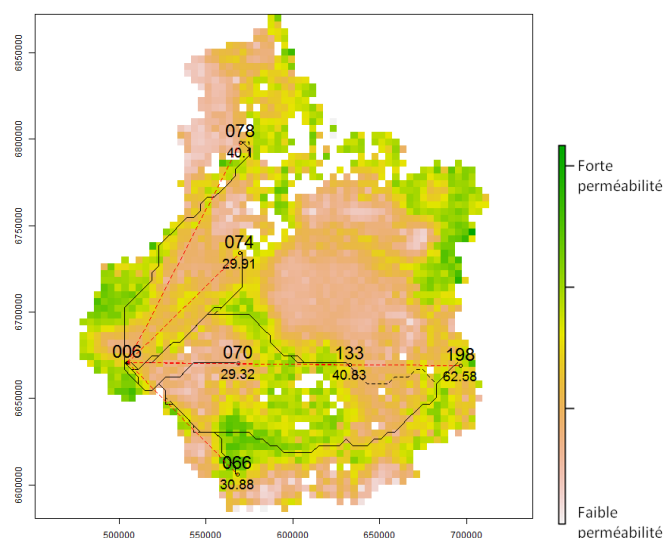
La carte des AHF est calculé à partir de la densité du bâti. Selon ce modèle les grandes villes ont donc une densité en pin importante, ce qui les fait apparaître en vert. La modélisation montre une faible quantité de pin distribuée sur toute la région. Le modèle de perméabilité étant basé sur les densités de pin, la carte obtenue montre une progression facile en zone rural (pixel vert) et une progression ralentie voire nul en ville.

La carte de l'IFN représente le Pin Sylvestre, le Pin Laricio, le Pin Noir, le Pin à Crochets, le Pin Cembro et le Pin Mugo. Les peuplements de pins couvrent la région Centre sur une surface moins importante que les AHF. Il y a moins d'information sur la perméabilité du paysage permettant de comprendre la progression de la chenille processionnaire.



**Figure 6.** Représentation des pins et modélisation de la perméabilité et associant les AHF et les données de l'IFN. La carte (a) est une carte complète de la répartition en pin dans la région Centre. Cette carte est obtenue par combinaison des deux cartes de densité précédentes. Ainsi, les peuplements de pins et les AHF sont représentés sur la même carte. La carte (b) est une carte de modélisation de la perméabilité, obtenue de la même façon que précédemment avec la projection des écopaysages définie par le bureau d'étude Biotope. La modélisation prend en compte à la fois les AHF et les peuplements de pins.

Le croisement des deux cartes de densité de pin, donne des informations plus précises. La perméabilité calculée met en évidence des corridors (pixel vert) et des barrières à la progression de la PP. La carte des écopaysages représente des régions qui ont une homogénéité de sol, de climat et de végétation. Certaines délimitations de ces écopaysages coïncident parfaitement avec la modélisation. Des écopaysages d'openfield et de prairie (Beauce, Brenne et Gâtinais) sont modélisés comme étant des paysages fortement perméables (corridors) alors que les grands massifs forestiers tels que la Sologne ou la forêt d'Orléans apparaissent comme des barrières. De plus, des paysages de transition comme la Sologne viticole sont également en accord avec la modélisation. La Sologne viticole est un paysage alternant bosquets et zones cultivées. Il s'agit d'un paysage de transition entre le grand massif forestier de la Sologne et les terres cultivées de la vallée de la Loire. La modélisation fait apparaître la Sologne viticole plus perméable que la Sologne forestière. La Champagne Berrichonne, paysage de grande culture, apparaît comme étant une zone perméable coupée en deux par une barrière qui demande à être mieux documentée.



**Figure 7.** Modélisation de la perméabilité du paysage à la progression de la PP avec les trajectoires de moindre coût. L'origine est représentée par le point noir plein, les destinations sont représentées par les cercles noirs. Les lignes noires représentent les trajectoires de moindre coût pour la processionnaire. Les lignes en pointillées rouges représentent la distance la plus courte. Les noms de maille sont donnés par les grands caractères, le coût de déplacement entre la maille d'origine à celle d'arrivée est donné en petits caractères.

Dans un milieu homogène (perméabilité identique dans toutes les cellules), la trajectoire de moindre coût correspond à la plus courte distance, c'est-à-dire une trajectoire en ligne droite. Or les trajectoires de moindre coût calculées à partir du modèle de perméabilité diffèrent nettement de la ligne droite ce qui illustre la forte hétérogénéité du paysage vis-à-vis de la dispersion de la PP.

**Tableau II** Matrice du moindre coût entre deux mailles de la grille 64 x 64

Arrivée\départ	006	066	070	074	078	133	198
006		31,95	32,9	31,34	44,23	41,34	64,14
066	30,88		20,02	32,5	52,95	19,72	35,81
070	29,32	17,15		20,15	40,92	17,68	39,18
074	29,91	31,14	21,59		23,44	26,88	49,68
078	40,1	48,28	39,02	20,41		44,02	53,29
133	40,83	18,69	19,83	27,81	48,26		22,8
198	62,58	35,18	41,63	50,07	57,48	22,26	

Pour arriver à la maille 070, le coût le moins élevé est lorsque le départ de la trajectoire est dans les mailles 066 ou 133. De même, le coût est faible lors des trajectoires 066 vers 133 et 133 vers 066. La distance paysagère est plus faible entre ces mailles, leur connectivité est importante.

Lorsque les trajectoires partent des mailles 006 et 074 pour aller vers la maille 070, les coûts sont plus élevés. La maille 070 est moins bien connectée avec les mailles 006 et 074.

### 3.4 Génétique du paysage

**Tableau III.** Test de Mantel confrontant la matrice de distance génétique (calculée avec 1026 individus) avec la matrice de distance paysagère et la matrice de distance euclidienne.

Le seuil de significativité est fixé à  $p = 0,05$ . \* :  $p < 0,05$ (significatif), \*\* :  $p < 0,01$  (très significatif), avec  $H_0$  : les matrices ne sont pas corrélées et  $H_1$  : les matrices sont corrélées.

	R statistique	Significativité
Matrice de distance génétique de Nei / Matrice de distance paysagère	0,08459	0,037*
Matrice de distance génétique de Nei / Matrice de distance euclidienne	0,2439	0,001**

La matrice de distance génétique de Nei est significativement corrélée avec la matrice de distance paysagère et la matrice de distance euclidienne. La corrélation entre distance génétique et distance paysagère devient non significative quand on réalise le test de Mantel partiel M. génétique versus m. paysagère / m. distance euclidienne (tableauIV). Ceci indique que la relation avec le paysage est indirecte et ne traduit donc qu'un effet global de la distance géographique sur la distance génétique.

**Tableau IV.** Test partiel de Mantel entre la distance génétique et la distance paysagère en contrôlant l'effet de la distance géographique. Le seuil de significativité est fixé à  $p = 0,05$ . NS :  $p > 0,05$  (non significatif).

	R statistique	Significativité
Matrice de distance génétique de Nei / Matrice de distance paysagère	0,0328	0,276 NS

**Tableau V.** Test de Mantel effectué en regroupant les individus par maille. La matrice de distance génétique est calculée selon les *clusters* déterminés par *Structure*.

	R statistique	Significativité
Test de mantel avec tous les <i>clusters</i>	0,1738	0,001**
Test partiel de mantel avec tous les <i>clusters</i>	0,07511	0,017
Test de Mantel <i>cluster vert</i>	0,1133	0,002***
Test partiel de Mantel <i>cluster vert</i>	0,06196	0,043*
Test de Mantel <i>cluster bleu</i>	0,01322	0,39NS
Test partiel de Mantel <i>cluster bleu</i>	-0,05164	0,826NS
Test de Mantel <i>cluster rouge</i>	0,2307	0,001***
Test partiel de Mantel <i>cluster rouge</i>	0,1374	0,001***

La probabilité d'appartenance au *cluster* vert ainsi que celle d'appartenir au *cluster* rouge sont fortement corrélées avec la matrice de distance paysagère. La probabilité d'appartenance au *cluster* diminue progressivement avec les coûts de traversée du paysage.

La probabilité d'appartenance au *cluster* bleu n'est pas corrélée avec la matrice de distance paysagère. Dans ce cas, le paysage n'influence pas la distribution du *cluster* bleu.

La confrontation des distances génétiques avec les distances paysagères montre que la structuration des populations à l'échelle de la grille n'est pas liée au paysage. La présence de trois lignées crée une forte structuration à cette échelle. L'analyse de la probabilité d'appartenance aux *clusters* en fonction de la perméabilité du paysage révèle que deux *clusters* sont distribués à large échelle et ont donc bien dispersé dans le paysage. Le troisième couvre une surface plus restreinte, sa dispersion dans l'espace est faible quelque soit le paysage rencontré.

## IV. Discussion

### 4.1 Diversité génétique

Nous avons observé un niveau de structuration important à une échelle relativement réduite de 256 x 256 km avec un  $F_{st}$  maximum entre paires de population de 0,25 (sites 251 et 70), alors qu'à l'échelle de la France entière, Kerdelhué *et al.* (2006) n'ont pas obtenue de valeurs de  $F_{st}$  supérieures à 0,23. Ces auteurs ont montré qu'à grande échelle la structure génétique de la processionnaire est caractérisée par un phénomène d'isolement par la distance (IBD). Ce patron est dû à une dispersion diffusive limitée dans l'espace conduisant à une différenciation génétique entre paire de populations proportionnelle à la distance géographique qui les séparent (Rousselet *et al.*, en préparation). Si on se base sur cette relation distance génétique / distance géographique mise en évidence par ces auteurs, il faudrait que les sites d'échantillonnage soient distants d'environ 900 km pour obtenir une valeur de  $F_{st}$  de 0,25, alors que le niveau de différenciation attendu (au sein d'une même lignée) pour une distance de l'ordre de 300 km (diagonale de notre grille) était de 0,07. Seule la diagonale Sud-Ouest / Nord-Est, correspondant à l'axe principal d'expansion, montre des valeurs compatibles avec un phénomène d'isolement par la distance seul. Le niveau de structuration, bien plus élevé qu'attendu, et a été étudié à l'aide du logiciel *Structure* qui permet d'explorer différents niveaux de regroupements de populations. Les estimations du nombre de clusters (K) données par structure montrent une structuration maximale lorsque K est de 6. Nous avons donc cherché à déterminer la pertinence biologique des clusters définis, de K=2 à K=6, en

considérant leur distribution spatiale à l'aide du logiciel MapInfo. La distribution spatiale des différents clusters est cohérente avec la mise en évidence dans cette région de trois lignées maternelles caractérisées par des mitotypes différents, dont deux appartenant au même mitogroupe (Rousselet *et al.*, 2010 ; Simonato *et al.*, comm. pers.). L'un des mitogroupes correspond à l'expansion d'une lignée maternelle distribuée dans l'ouest de l'Europe (mitotype 14, Rousselet *et al.*, 2010), l'autre mitogroupe d'une lignée présente dans l'Est (mitotype 1 et mitotype 5), les deux mitogroupes arrivant en contact dans la zone d'étude. Le coin Sud-Est de notre grille correspond à la zone d'expansion du mitotype 1 du sud de la France vers le nord. Le mitotype 5 est quant à lui trouvé dans le coin Nord-Ouest où il aurait été introduit près d'Alençon depuis les Alpes. Il a été trouvé à proximité d'une grande pépinière et correspondrait une colonie pionnière isolée qui a été rejointe par le front d'expansion vers le nord. L'existence de cette population isolée a été rapportée dès la fin des années 1970 par Abgrall (2001). C'est donc probablement ce mélange de lignées maternelles très différenciées qui est à l'origine de la très forte structuration observée à une si faible échelle d'étude.

Ainsi, la distribution des clusters rend compte de la distribution pour  $K=2$  des deux mitogroupes, pour  $K=3$  des 3 mitotypes, et pour  $K=4, 5$  et  $6$  de différents niveaux d'admixture entre les lignées maternelles d'origine (présentes dans la région avant l'expansion due au réchauffement climatique mais à distribution disjointe). Nous avons donc choisi pour nos analyses de nous baser principalement sur  $K=3$ .

La distribution de ces 3 clusters en région centre rend compte du point de départ de leur extension :

- La probabilité d'appartenance au *cluster* vert est proche des 100% dans le coin Sud Ouest de la grille et diminue progressivement en remontant vers le Nord. La progression de ce *cluster* se fait depuis l'Ouest du Massif centrale vers le Nord-Est.
- La probabilité d'appartenance au *cluster* bleu est proche de 100% dans le Sud Est, et diminue très peu en remontant vers le Nord. La progression de ce *cluster* semble être freinée.
- La répartition des probabilités d'appartenance au *cluster* rouge montre une progression depuis le Nord Ouest de la région Centre. En effet, l'origine du *cluster* est au Nord Ouest de la région Centre où la probabilité d'appartenance est voisine de 100%. Cette probabilité diminue progressivement en redescendant vers le Sud. Le mélange se fait progressivement avec la remonté vers le Nord du *cluster* vert. La présence de ce *cluster* ralentit (ou masque) la progression du *cluster* vert vers le Nord Ouest.

A la jonction entre le *cluster* vert et le *cluster* rouge, la probabilité d'appartenance tend vers les 50%. Il existe un fort taux d'introgression entre ces *clusters*. Cela traduit une forte admixture entre le *cluster* rouge et le *cluster* vert, ce qui montre que ces 2 *clusters* ont une bonne dispersion à travers le paysage et peu d'éléments ne perturbent leur rencontre. Un individu vert et un individu rouge ont de bonnes probabilités de rencontre.

Au Sud-Est de la région Centre, l'introgression est plus faible. A la jonction des *clusters* vert et bleu, la probabilité d'appartenance diminue rapidement. L'admixture entre le *cluster* bleu et le *cluster* vert est faible. Ce qui traduit une probabilité de rencontre faible et ralentit la progression.

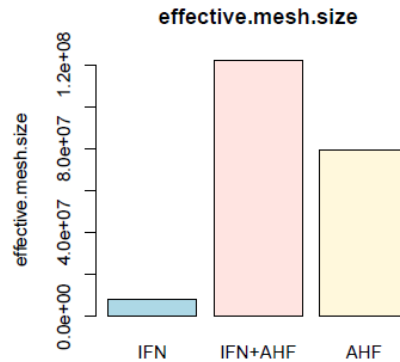
La projection des valeurs de  $k=4$ ,  $k=5$  et  $k=6$  (annexe) confirme l'introgression observée. Les *clusters* supplémentaires (jaune violet et cyan) se placent à la jonction entre les *clusters* (annexe 6). Le *cluster* nouvellement créé pour  $k=4$  et représente l'admixture entre les *clusters* rouge et vert. Pour  $k=5$ , le logiciel *Structure* crée 2 *clusters* supplémentaires entre le *cluster* vert et le *cluster* rouge selon la contribution de chaque *cluster* à l'admixture. Enfin pour  $k=6$  ; le nouveau *cluster* créé représente l'admixture entre le *cluster* bleu et le *cluster* vert.

*Structure* a calculé deux niveaux d'admixture à la jonction rouge et vert avant d'en déterminer un à la jonction vert – bleu. Le niveau d'introgression peut être jugé plus fort entre les verts et les rouges qu'entre les bleus et les verts.

## 4.2 Ecologie du paysage

Nous nous sommes basés sur une cartographie des AHF construite par simulation à partir d'un modèle reliant la probabilité d'occurrence d'un arbre hôte isolé en fonction de la distance du bâti le plus proche. Le modèle est construit à partir de données collectées en Beauce et a été extrapolé à l'ensemble de la région centre. Lorsque la densité en bâti augmente, la probabilité de rencontrer un pin isolé augmente, c'est pourquoi la densité des pins isolés est plus importante en centre ville. Pour améliorer cette carte, des échantillonnages en ville et dans les zones de transition urbain/rural sont en cours pour affiner le modèle.

La carte de l'IFN prise pour les différents calculs est la carte répertoriant les hôtes préférentielles de la PP (pins noirs, pins sylvestres) et non tous les hôtes probables. Ces pins sont également les pins les plus représentés dans la région Centre. La faible densité des autres hôtes probables et la faible fréquence de colonisation de ces hôtes par la PP dans notre région ont permis de les négliger.



**Figure 8.** Histogramme représentant la valeur de l'indice *effective mesh size* en fonction de la représentation des pins. Plus l'indice est élevé, plus la connectivité est importante.

L'association de la carte des AHF avec la carte de l'IFN, permet une meilleure représentation des pins dans la région Centre en particulier car elle permet de prendre en compte des zones au paysage ouvert avec peu d'arbre. L'*effective mesh size* (Jaeger 2000) est un indice de fragmentation de l'habitat qui traduit la probabilité que deux points tirés au hasard soient dans une surface recouverte par des pins. Plus la probabilité sera forte, moins la fragmentation est importante. La fragmentation constitue le principal facteur jouant sur la connectivité du paysage. Une fragmentation importante du paysage provoque une faible connectivité. Lorsque l'*effective mesh size* est calculé avec les informations fournies par l'IFN, la valeur est très faible. L'IFN répertorie des surfaces boisées de 2,25 ha, dans notre cas, seuls les peuplements de pin sont pris en compte. Le paysage apparaît alors fortement fragmenté avec peu de connectivité entre les *patches* de pins. La carte AHF répertorie tous les arbres de façon ponctuelle. Cela ne rajoute pas de grande surface de pins mais multiplie de façon importante la présence de pins dans la région. Les pins sont alors éparpillés dans toute la région. La densité de ces pins isolés est suffisamment importante pour augmenter la probabilité que deux points soient sur un pin, l'*effective mesh size* est alors plus fort dans cette représentation du paysage.

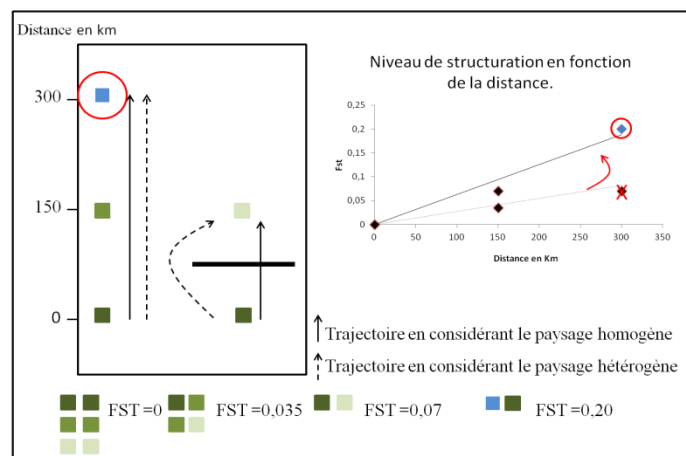
Ces résultats indiquent clairement l'importance des pins isolés du point de vue de la connectivité de l'habitat de la PP. Nous avons utilisé ces cartes de distribution d'hôtes dans l'ensemble de nos analyses de génétique du paysage. Nos résultats indiquent qu'en modélisant la perméabilité de la progression de la PP au travers du paysage, la présence d'AHF augmente fortement la perméabilité, les faisant apparaître comme des corridors alors que les grands massifs forestiers comme la Sologne et la forêt d'Orléans apparaissent comme des barrières à la progression.



Le modèle peut cependant être amélioré. L'hypothèse qui estime que les zones avec une forte densité de pin sont des barrières à la dispersion va à l'encontre de la biologie de l'espèce. Sous le terme barrière à la dispersion se cache bien souvent le concept d'habitat défavorable (ou faible en ressources) (Bergès 2010), ce qui n'est pas le cas des grands massifs forestiers avec une densité de pin importante. Or au sens strict la définition d'une barrière n'aborde pas les habitats défavorables. En effet, une barrière à la dispersion est une zone qui ralentit ou interrompt la progression, c'est-à-dire un élément dans lequel la perméabilité est faible. Un habitat hostile freine la progression. Mais dans le cas d'une espèce aux mœurs plutôt sédentaires, une zone favorable avec une forte densité en pin présente peu de raison de disperser loin. Si la dispersion est à de faible distance, la progression se fait alors petit à petit. Ainsi la traversée d'un élément dense en pin est lente, ce qu'il le rend peu perméable. Au contraire lorsque la densité en pin est faible, les individus ont tendance à disperser plus loin. La progression est alors plus rapide et l'élément se présente très perméable.

Les couts affichés sur la figure 7 sont les moindres. Il aurait été intéressant de connaître les coûts d'une trajectoire en ligne droite (distance kilométrique la plus courte) pour pouvoir les comparer avec les moindres coûts. Il est pour le moment impossible de le faire directement avec les fonctions du *package gdistance*.

### 4.3 Génétique du paysage mais en avant les corridors



**Figure 9.** Schéma représentant la conséquence des forts niveaux de structuration dus à la présence de lignes différentes. Les carrés verts représentent la structuration au sein d'une lignée au travers du paysage. Le carré bleu représente l'introduction d'une autre lignée.

La première confrontation génétique/paysage montre une corrélation à la fois entre la matrice de distance génétique et la matrice de distance paysagère et entre la matrice génétique et la matrice de distance euclidienne. Mais il n'y a plus de corrélation lorsque l'on applique le test partiel de Mantel. L'hétérogénéité du paysage n'explique donc pas la structure génétique. Il

s'avère en effet, que la région Centre est occupée par trois lignées distinctes se trouvant chacune à un coin de la grille. Ces lignées sont donc à la fois distantes géographiquement et génétiquement. Ces fortes valeurs de distance génétique entre des populations éloignées géographiquement et les faibles valeurs des populations proches noient les faibles variations que le paysage engendre. De plus, ces trois lignées forment des zones d'agrégats d'individus au sein desquels la différenciation est faible, mais la différenciation inter groupe est forte. Une telle situation est connue pour être mal prise en compte par le test de Mantel (Legendre et Trousselier 1988). La présence de plusieurs lignées en mélange (fortement différenciées) masque, s'il existe, l'effet du paysage qu'il aurait été possible de mesurer à travers les variations du niveau de structuration au sein d'une seule et même lignée. On s'attendait à ce que les barrières à la dispersion créent localement plus de structure qu'attendue sous un phénomène d'IBD seul, mais cet effet sur les distances génétiques, s'il existe n'est pas détectable ici car probablement trop faible en comparaison de la structuration créée par l'introduction et/ou l'expansion de plusieurs lignées. L'influence du paysage est donc plutôt à rechercher dans le niveau d'admixture entre ces 3 groupes dont on peut faire l'hypothèse qu'il est lié à la perméabilité du paysage.

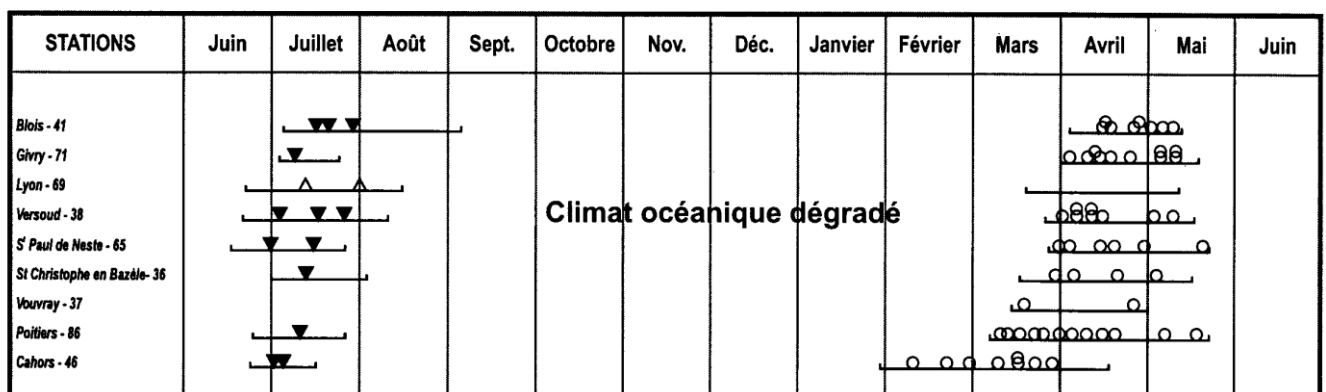
La création de *cluster* a tout d'abord permis de visualiser la distribution de la structure mais également permis de séparer les individus en trois grands groupes. En prenant les groupes indépendamment pour les tests de Mantel cela permet d'éliminer les agrégations. Chaque individu a une probabilité d'appartenir à un *cluster*. A partir de ces probabilités, le taux d'admixture peut être déduit. Une admixture maximale est atteinte lorsque la probabilité d'appartenance est également réparti au sein des *clusters* en présence. Dans les cas d'une rencontre entre deux *clusters* l'admixture maximum est obtenue lorsque l'individu à une probabilité d'appartenance de 50% pour les deux *clusters*. L'admixture est alors un bon indice pour caractériser la rencontre entre *clusters*. La rencontre est facilitée lorsque les flux de gènes entre *cluster* sont importants. Le test de Mantel a prouvé que la qualité des flux de gènes pouvait dépendre de la perméabilité du paysage à travers l'admixture. Dans le cas des *clusters* rouge et vert le paysage joue un rôle important dans leur rencontre. Les corridors sont des lieux à forte mixité alors que les barrières séparent des individus plus différents.

Dans le cas du *cluster* bleu, la rencontre avec les autres *clusters* ne dépend pas du paysage. Il s'avère en regardant la distribution des probabilités d'appartenance des individus que le *cluster* bleu est très rarement associé à un autre *cluster*. Il couvre une surface moins importante. De plus, ce *cluster* est très peu présent en région Centre (annexe 6), il est présent

essentiellement dans le coin Sud Est de la grille qui n'est pas pris en compte dans la matrice paysagère.

Cependant, il est également possible qu'une barrière biologique masque sa dispersion dans le paysage. En effet, en sachant que les mâles ont une dispersion plus lointaine que les femelles (Demolin 1969a), il est probable que les mâles dispersent loin de leur origine (50km) (Demolin 1969a). Dans ce cas les seules femelles disponibles sont celles dont l'origine est proche, c'est-à-dire d'un *cluster* différent. Ayant des origines différentes l'émergence des mâles et des femelles peut être décalée, on parle alors de rupture de coïncidence phénologique provoquant un isolement reproductif. Une étude sur les pics de vols montre qu'il existe un décalage phénologique de quelques jours selon la latitude et l'altitude (Abgrall et al 2001). Abgrall se base sur une étude de Demolin, montrant une précocité des individus soumis à un climat Aquitain (hiver doux). Le climat Méditerranéen et le climat océanique dégradé provoquent un retard du cycle de développement. La région Centre est soumise à un climat océanique dégradé qu'ils définissent comme ayant une tendance plus ou moins continentale. C'est-à-dire, une tendance à avoir des hivers plus froids et des étés plus secs.

Des stations de cette étude ont été placées en région Centre (Vouvray, Blois et St Christophe de Bazelle) ainsi qu'à Givry en Haute Saône. Abgrall observe des différences dans les périodes de vols entre Givry et Blois. A Blois, la période de vol s'étale de juillet à début septembre avec des pics de vol dans la dernière quinzaine de juillet. A Givry la période de vol est plus courte, le pic est début juillet (figure 10). Cela est typique d'une région ayant des premières gelées en début d'automne. La PP est capable de résister au premier gel lorsqu'elle a atteint le stade L3. Pour les individus exposés à un hiver précoce, le cycle de développement doit atteindre le stade L3 avant les premières gelées. Si la période de vol est trop tardive, les larves n'ont pas le temps d'atteindre le stade L3 avant les premières gelées. La date d'émergence des adultes serait donc conditionnée par les premiers gels. Dans ces régions la date d'envol est plus tôt dans la saison que dans les régions aux hivers plus doux.



**Figure 10.** Adaptations bioécologiques de la PP. Variations de cycle de développement en fonction du climat (Abaque « terrain »). Abgrall et al

- Périodes d'émergence des adultes
- Chenilles en procession de nymphe
- △ Dates des 50% d'émergence - Piégeage lumineux
- ▼ Dates des 50% d'émergence - Enfouissement forcé des chenilles
- Dates des des prélèvements de chenilles en procession de nymphe

Ainsi les populations de l'Ouest de la région, soumis à un climat plutôt océanique, émergeraient quelques jours après les populations de l'Est, soumis à un climat semi continentale. Si les mâles de l'Ouest peuvent disperser vers l'Est, ils trouvent peu de femelles pour se reproduire. De même les mâles issus des populations de l'Est auront des difficultés à trouver une femelle à l'Ouest qui émergent plus tard. Du fait de ce décalage phénologique, la rencontre entre les populations du *cluster* bleu et les populations du *cluster* vert est limitée et les flux de gènes faibles.

## V. Conclusion

Le niveau de structuration observé en Région Centre n'a pas pu mettre en évidence les flux de gènes en fonction du paysage. De trop forts taux de structuration, du à la présence de trois lignées distinctes ont noyé les faibles variations dues au paysage. Cependant la rencontre et le mélange entre les lignées peut mettre en évidence des corridors et des barrières. De plus, les résultats ont montré l'importance de prendre en compte les AHF dans la progression des ravageurs dans le paysage. De telles études peuvent aider à prévoir les voies de colonisation qui peuvent être utilisées par des espèces en expansion ou des espèces invasives.

## VI. Références

Abgrall, J.F. 2001. Le réseau de surveillance PP en France 1969-1989 : conception – historique – résultats. C.E.M.A.G.R.E.F, groupement de Nogent-sur-Vernisson.

Belkhir, K., P. Borsa., L. Chikhi L., N. Raufaste and Bonhomme F. 1996-2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Bergès 2010 Corridors écologiques et conservation de la biodiversité, intérêts et limites pour la mise en place de la Trame verte et bleue. Sciences Eaux & Territoires 03 : 34-39

Bialozyt, R., B. Ziegenhagen, and R.J. Petit. 2005. Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology* 19: 12–20.

Burban, C., E. Magnoux., J. Rousselet., C. Kerdelhué., 2012. Development and characterization of 13 new microsatellite markers in the pine processionary moth, *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae). *Molecular Ecology Resources Primer Development Consortium et al., Permanent genetic resources added to Molecular Ecology Resources Database* 1 August 2011–30 September 2011. *Molecular Ecology Resources*, 12 (1) : 185-189.

Demolin, G., 1969a Comportement des adultes de *Thaumetopoea pityocampa* Schiff. Dispersion spatiale, importance économique. *Annales des Sciences Forestières* 26 : 81-102.

van Etten, Jacob. 2012. R package gdistance: distances and routes on geographical grids (version 1.1-4)

Jaeger, Jochen A. G. 2000. Landscape Division, Splitting Index, and Effective Mesh Size: New Measures of Landscape Fragmentation. *Landscape Ecology* 15, 2 : 115-130.

Klein E.K., C. Lavigne and P.-H. Gouyon. 2006. Mixing of propagules from discrete sources at long distance: comparing a dispersal tail to an exponential. *BMC Ecology*, 6, 3.

Ibrahim K.M., R.A. Nichols and G.M. Hewitt. 1996. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, 77: 282-291.

Legendre, P. and M. Trousselier, 1988. Aquatic heterotrophic bacteria: Modeling in the presence of spatial autocorrelation. *Limnol. Oceanogr.* 33: 1055-1067

McRae, B. H., B. G. Dickson., T. H Keitt, and V.B. Shah. 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology* 89 : 2712–2724.

Manel, S., M.K. Schwartz., G. Luikart et P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* 18 : 189–197

Mantel, N. 1967. The detection of disease *clustering* and a generalized regression approach. *Cancer Res* 27 : 209–220.

Nei, M 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583-590.

Pritchard, J. K., M. Stephens and P. Donnelly. 2000a. Inference of population structure using multilocus genotype data. *Genetics*, 155:945–959.

Robinet, C., P. Baier., J. Pennerstorfer., A. Schopf and A. Roques. 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., J. Rousselet., C.E. Imbert., D. Sauvard., J. Garcia., F. Goussard and A. Roques. 2010. Le réchauffement climatique et le transport accidentel par l'Homme responsable de l'expansion de la chenille processionnaire du pin. *Forêt Wallonne* 108 : 19-27

Robinet, C., C.E. Imbert., J. Rousselet., D. Sauvard., J. Garcia., F. Goussard. and A. Roques. 2012. Human-mediated long-distance jumps of the pine processionary moth in Europe. *Biological Invasions* 14, 8: 1557-1569.

Rousselet, J. in prep (chapitre de livre Urticlim)

Rousselet, J., R. Zhao., D. Argal., M. Simonato., A. Battisti., A. Roques. and C. Kerdelhué. 2010. The Role of Topography in Structuring the Demographic History of the Pine Processionary Moth, *Thaumetopoea Pityocampa* (Lepidoptera: Notodontidae). *Journal of Biogeography* 37, 8: 1478–1490.

Rousselet, J., E. Magnoux and C. Kerdelhué. 2004. Characterization of five microsatellite loci in the pine processionary moth *Thaumetopoea pityocampa* (Lepidoptera Notodontidae Thaumetopoeinae). *Molecular Ecology Notes*, 4 : 213-214.

Santos, H., J. Rousselet., E. Magnoux., M.R. Paiva., M. Branco and C. Kerdelhué. 2007. Genetic isolation through time: allochronic differentiation of a phenologically atypical population of the pine processionary moth. *Proceedings of the Royal Society B: Biological Sciences*, 274 : 935-941.

Samalens, J.C. and J.P. Rossi. 2011. Does Landscape Composition Alter the Spatiotemporal Distribution of the Pine Processionary Moth in a Pine Plantation Forest?. *Population Ecology* 53, 2: 287-296.

Taylor. P., D. Lenore Fahrig, H. Kringen et M. Gray. 1993 Connectivity Is a Vital Element of Landscape Structure . *Oikos* 68, 3: 571.

Walther, G. R., E. Post, P. Convey, A. Menzel, C. Parmesan, T. J. C. Beebee, J. M. Fromentin, O. Hoegh-Guldberg, et F. Bairlein. 2002. Ecological Responses to Recent Climate Change . *Nature* 416, 6879: 389-395.

Weir B.S. et C.C COCKERHAM 1984. Estimating F-statistics for the analysis of population structure. *EvolU/ion* 38: 1358-1 370.

### **Remerciements :**

Je remercie Alain Roques de m'avoir accueilli au sein de son équipe ainsi que François Lieutier pour m'avoir permis d'effectuer ce stage de M2.

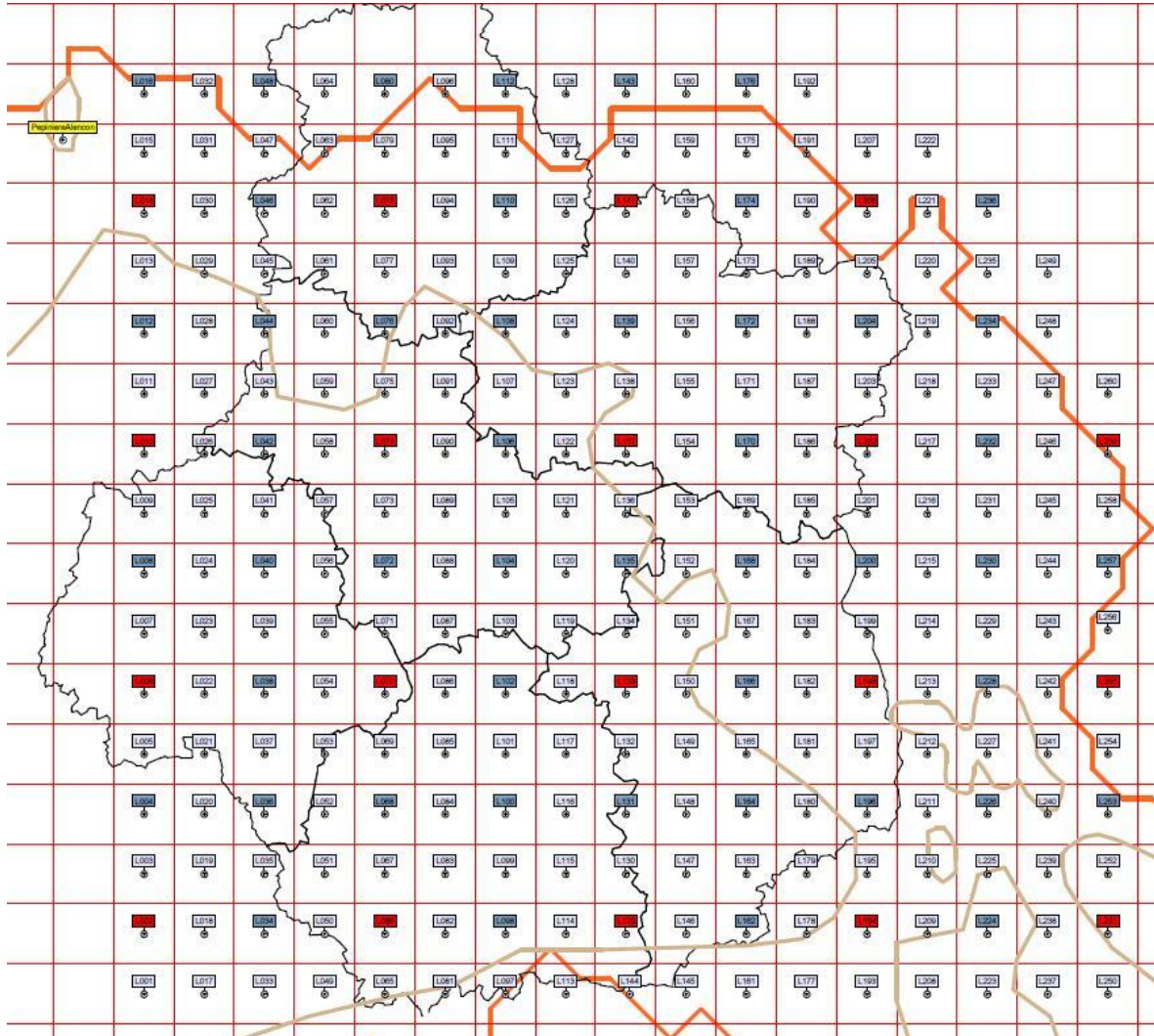
Je dis un grand merci à Jérôme Rousselet et Jean Pierre Rossi pour avoir rendu ce stage possible ainsi que pour m'avoir encadré et transmis leurs connaissances sur le sujet et parfois un peu plus.

Merci également à Emmanuelle Magnoux, Béatrice Courtial et Claudine Courtin pour leur connaissance de la biologie moléculaire.

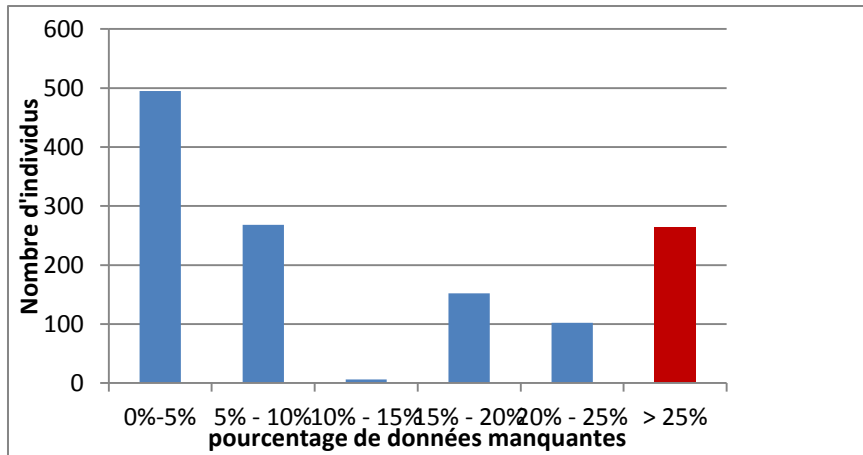
Sans oublier l'ensemble des thésard, des stagiaires et autre non permanent pour leur disponibilité et leur conversation scientifique ou non.

Je pense également au reste de l'unité de Zoologie forestières pour leur accueil.

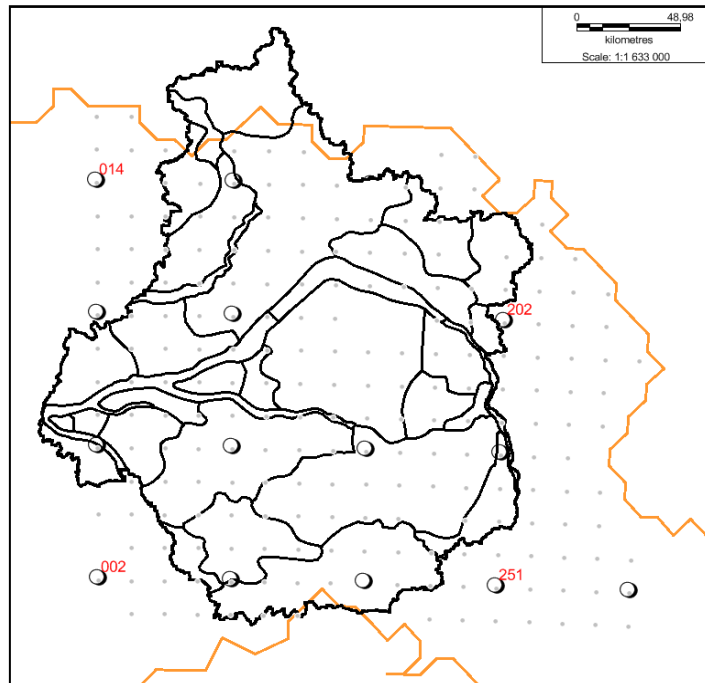
## Annexes :



**Annexe 1** Grille d'échantillonnage au niveau du front de l'aire de répartition de la processionnaire du pin dans le bassin parisien. Les lignes oranges représentent les limites du front. Chaque maille est libellée d'une étiquette. Dans les mailles avec une étiquette grises ou plus bleues 5 individus ont été prélevés, les étiquettes rouges représentent les mailles dans lesquelles 15 individus ont été prélevés.

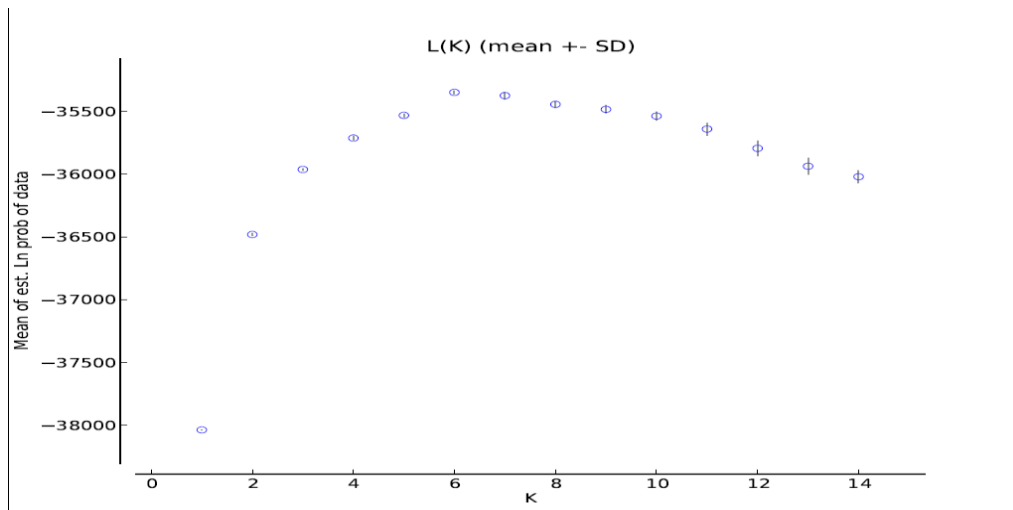


**Annexe 2** Répartition des individus selon le pourcentage de données manquantes sur les 11 marqueurs retenus. Au-delà de 25% (barre rouge) les individus n'ont pas été pris en compte.



**Annexe 3** Carte de la région Centre avec les délimitations des écopaysages. Les lignes oranges représentent le front de colonisation de la processionnaire du pin, les points gris représentent les populations ayant des individus génotypés, les cercles noirs représentent les populations de la grille 64 x 64km.

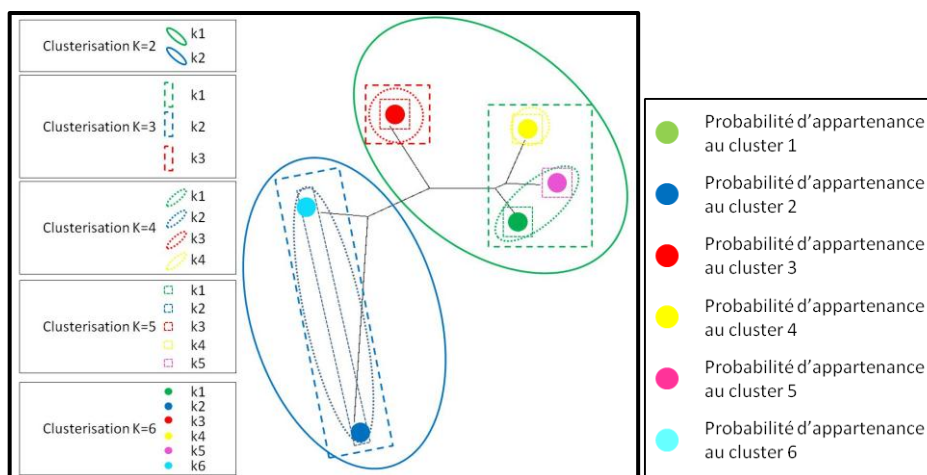


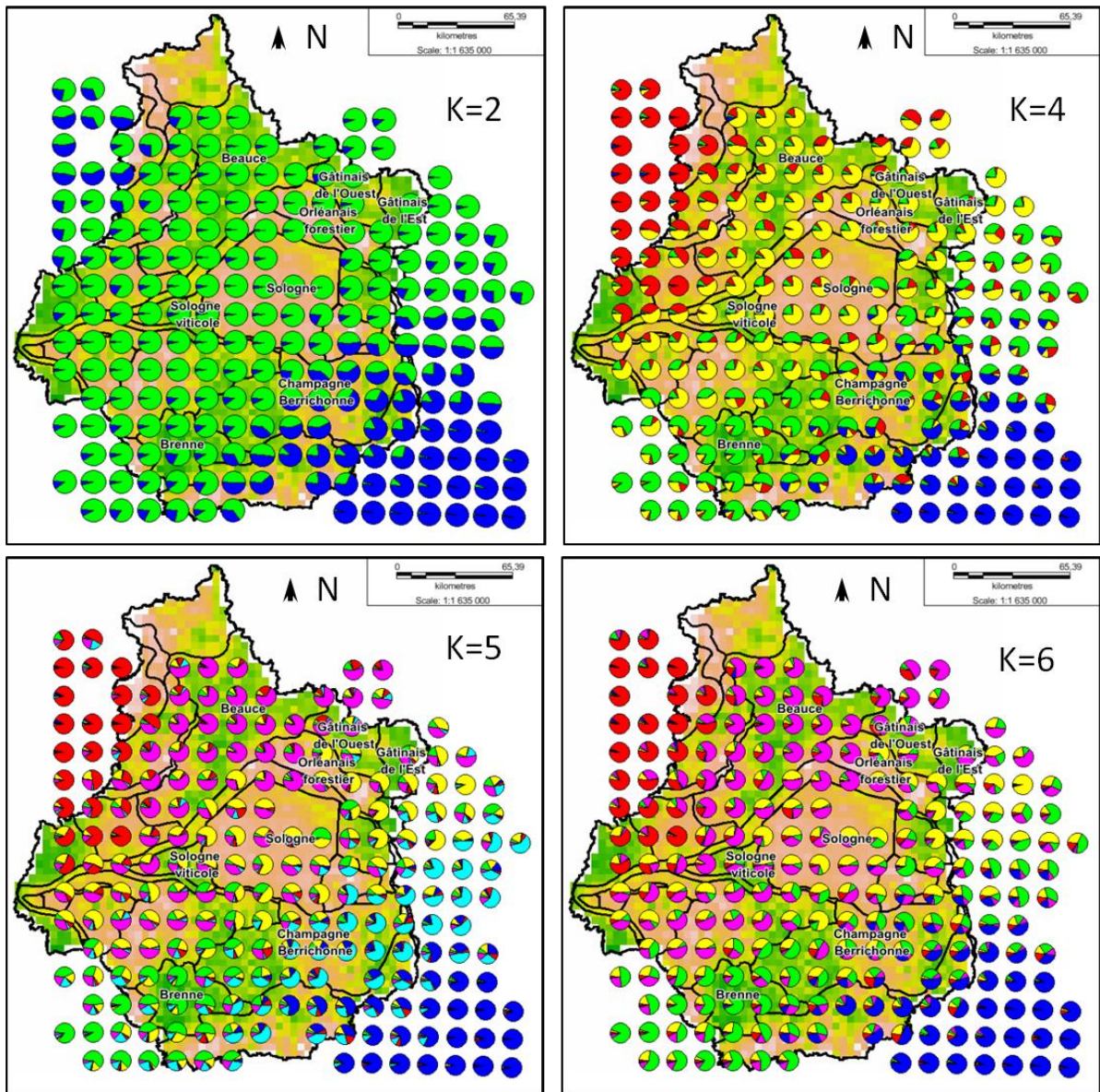


**Annexe 4** Estimation du  $\ln P(X|K)$  en fonction de  $K$ . Chaque point représente un run (1 iteration). Le plateau est atteint pour  $K=6$ .

nombre de cluster	k=2		k=4				k=5					k=6					
cluster	k1	k2	k1	k2	k3	k4	k1	k2	k3	k4	k5	k1	k2	k3	k4	k5	k6
<b>Hétérozygotie attendu</b>	0,72	0,69	0,74	0,70	0,69	0,65	0,71	0,74	0,69	0,64	0,69	0,73	0,70	0,63	0,69	0,75	0,69

**Annexe 5** Hétérozygotie attendu au sein des différents clusters, pour un  $k$  allant de 2 à 6.





**Annexe 6** : Arbre des distance génétique entre cluster pour un nombre de cluster de K=2, K=4, K=5 et K=6. ainsi que leur représentation spatiale des probabilités d'appartenance au cluster.