



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

Faculté des Sciences et Technologies
Ecole doctorale Ressources Procédés Produits Environnement
Département de Formation Doctorale Sciences Agronomiques et Forestières, Biologie et Ecologie,
Biotechnologies
Spécialité Biologie Végétale et Forestière

Thèse

Présentée pour l'obtention du titre de

Docteur de l'université Henri Poincaré, Nancy I

Par

Constance XHAARD

Influence des processus démographiques sur la structure et les caractéristiques génétiques des champignons phytopathogènes, cas de l'agent de la rouille du peuplier *Melampsora larici-populina*

Date prévisionnelle de soutenance le 23 juin 2011

Composition du jury

Jean-François COSSON
Richard HAMELIN
Eric GELHAYE
Lydia BOUSSET
Benoit MARCAIS

Directeur de Recherche
Directeur de Recherche
Professeur
Chargée de Recherche
Directeur de Recherche

Rapporteur
Rapporteur
Examineur
Examinatrice
Directeur de thèse

Membres invités

Pascal Frey
Fabien Halkett

Chargé de Recherche
Chargé de Recherche

Co-directeur
Co-directeur

UMR 1136 INRA / UHP, Interactions Arbres / Microorganismes, Ecologie des Champignons
Pathogènes Forestiers
Centre INRA de Nancy, Champenoux

REMERCIEMENTS

Mes plus sincères remerciements vont à Pascal Frey et à Fabien Halkett, ils m'ont donné la possibilité d'effectuer cette thèse et m'ont encadrée et supportée durant ces quatre années. Je remercie Pascal pour son dynamisme, sa bonne humeur et ses connaissances sur Mlp ; et Fabien pour sa disponibilité, son partage de connaissances sur la génétique des populations et d'avoir maintenu mon moral grâce à quelques tablettes de chocolat.

Je remercie également Jean-François Cosson et Richard Hamelin pour avoir accepté le rôle de rapporteurs et Lydia Bousset et Eric Gelhaye qui ont bien voulu participer à mon jury de thèse.

J'adresse aussi mes remerciements aux membres de mes comités de thèse, Cyril Dutech, Etienne Klein, Jérôme Enjalbert et Pierre Gladieux, qui, grâce à leurs regards extérieurs, m'ont fourni de bons conseils.

Une collaboration pour un travail de modélisation a été initiée avec Frédéric Fabre et Samuel Soubeyrand, je les remercie vivement de leur accueil chaleureux à Avignon et de m'avoir fait profiter de leurs talents de modélisateurs.

Il m'est impossible d'oublier tous les membres de l'équipe de pathologie forestière qui m'ont accueillie pendant la durée de cette thèse, ceux qui m'ont accompagnée au cours des tournées dans la Durance, ainsi que ceux qui ont accepté de garder Stan pendant mes absences. Je remercie donc chaleureusement Benoît, Claude, Marie-Claude, Olivier, sans oublier Béranger et Christine. Je remercie plus spécialement Axelle sans qui une partie de l'acquisition des données n'aurait pas été possible et Bénédicte pour m'avoir embauchée comme secrétaire et pour toutes les discussions autour d'un thé.

Je n'oublie pas non plus les autres doctorants, stagiaires ou contractuels, dont certains sont devenus des amis.

La bonne humeur installée dans cette équipe a largement contribué au bon déroulement de cette thèse.

Je tiens également à remercier mes amis lorrains qui m'ont accueillie dans leur région, en particulier grâce à Kiwi qui m'a aidé à me créer de nombreuses amitiés équestres. Et les non-lorrains qui ne m'ont pas oubliée pour autant. Une attention spéciale à ma famille, mes parents et mes sœurs, pour leur soutien inconditionnel, malgré qu'ils aient souvent été témoins de mes sautes d'humeurs.

Enfin, je remercie Germain pour tout ce qu'il représente.

SOMMAIRE

INTRODUCTION GENERALE.....	4
1. Biologie des populations des agents phytopathogènes.....	4
1.1. L'épidémiologie.....	4
1.2. Les apports de la génétique des populations pour l'étude des champignons phytopathogènes.....	5
1.2.1. Principes de la génétique des populations.....	5
1.2.2. La génétique des populations chez les champignons phytopathogènes....	6
2. Le pathosystème Rouille / Peuplier.....	9
2.1. <i>Melampsora larici-populina</i> , l'agent de la rouille du peuplier.....	9
2.1.1. Biologie de <i>M. larici-populina</i>	9
2.1.2. Cycle de vie de <i>M. larici-populina</i>	9
2.2. L'hôte écidien, le mélèze.....	10
2.3. L'hôte télien, le peuplier.....	11
2.3.1. Peuplier noir.....	11
2.3.2. Peupliers cultivés.....	11
2.4. Etat des connaissances sur la structure des populations de <i>M. larici-populina</i>	12
2.5. Le site de la vallée de la Durance.....	13
OBJECTIFS DE LA THESE.....	15
MATERIEL ET METHODES.....	17
1. Introduction : méthodes de biologie moléculaire et principes des analyses de génétique des populations réalisées.....	17
2. Développement de nouveaux marqueurs microsatellites.....	19
ARTICLE 1 : CHARACTERIZATION OF 41 MICROSATELLITES LOCI DEVELOPED FROM THE GENOME SEQUENCE OF THE POPLAR RUST FUNGUS, <i>MELAMPSORA</i> <i>LARICI-POPULINA</i>	19
Article publié en 2009 dans la revue <i>Conservation Genetics Resources</i> 1 : 21-25	
3. Les assignations Bayésiennes.....	23
3.1. Définir des populations.....	23
3.2. Les programmes d'assignation.....	24
3.3. Méthodes alternatives.....	27
3.4. Programmes utilisés.....	27
RESULTATS.....	29
Chapitre 1 : Influence de la présence de l'hôte cultivé sur la structure génétique de <i>Melampsora larici-populina</i> sur son hôte sauvage en France.....	29
ARTICLE 2 : THE GENETIC STRUCTURE OF THE PLANT PATHOGENIC FUNGUS <i>MELAMPSORA LARICI-POPULINA</i> ON ITS WILD HOST IS EXTENSIVELY IMPACTED BY HOST DOMESTICATION.....	30
Article accepté dans <i>Molecular Ecology</i>	

Chapitre 2 : Structure génétique spatiale de <i>Melampsora larici-populina</i> dans la zone de sympatrie peuplier-mélèze.....	60
ARTICLE 3 : THE WEST SIDE STORY OF A PLANT PATHOGEN: LANDSCAPE GENETICS OF THE POPLAR RUST FUNGUS IN THE FRENCH ALPS.....	61
Article en préparation	
Chapitre 3 : Mise en évidence de deux épidémies simultanées de rouille du peuplier dans la vallée de la Durance.....	78
ARTICLE 4 : DISENTANGLING PATHOGEN ORIGINS FROM WILD AND CULTIVATED STANDS AT THE END OF AN EPIDEMIC.....	79
Article en préparation	
Chapitre 4 : Déroulement de l'épidémie dans la vallée de la Durance.....	99
1. Introduction.....	99
2. Méthodologie pour le suivi de l'épidémie.....	100
3. Résultats.....	101
4. Perspectives sur la modélisation du déroulement de l'épidémie.....	101
Chapitre 5 : Conséquences génétiques de la dispersion de la rouille du peuplier dans la vallée de la Durance.....	102
1. Introduction.....	102
2. Matériel et Méthodes.....	103
3. Résultats.....	104
4. Discussion.....	105
5. Conclusion et perspectives.....	107
DISCUSSION GENERALE.....	108
1. Conséquences génétiques et démographiques d'un contournement de résistance.....	108
1.1. A l'échelle de la France.....	109
1.2. A l'échelle régionale : apport de l'étude de la vallée de la Durance..	110
1.3. Stabilité des groupes génétiques et enseignement pour les stratégies de déploiement des cultivars résistant.....	111
2. Couplage de l'épidémiologie et de la génétique des populations..	112
CONCLUSION	114
BIBLIOGRAPHIE	115

Introduction

INTRODUCTION GENERALE

1. Biologie des populations des agents phytopathogènes

Les maladies des plantes sont aussi vieilles que l'agriculture. En effet, dès son origine, l'agriculture créa des conditions favorables aux maladies des plantes cultivées en concentrant des populations de plantes sur des surfaces limitées et en effectuant des cultures successives de la même plante (Lepoivre 2003). Depuis très longtemps, l'homme a cherché à comprendre l'origine des maladies des plantes cultivées afin de mieux les éviter ou les combattre. En effet celles-ci sont à l'origine de pertes de rendement qui entraînent des pertes économiques, voire des famines dans les cas les plus dramatiques. Le cas le plus tristement célèbre est celui de l'introduction en Europe en 1844 de *Phytophthora infestans*, agent du mildiou de la pomme de terre, puis de la Grande Famine Irlandaise de 1845 à 1852, qui fit près d'un million de morts et poussa deux millions d'Irlandais à émigrer vers d'autres continents (Fry 2008). Encore aujourd'hui, la principale maladie du riz, la pyriculariose causée par *Magnaporthe oryzae*, causes des pertes mondiales annuelles équivalentes à l'alimentation de 60 millions de personnes (Dean *et al.* 2005). De même, une nouvelle souche de rouille noire du blé apparue en Ouganda en 1999 représente une nouvelle menace de famine dans certains pays d'Asie (Stokstad 2007).

Historiquement, l'étude des maladies des plantes, ou phytopathologie, a concerné principalement l'étiologie des maladies et leur épidémiologie. Plus récemment, grâce au développement de la biologie moléculaire, la phytopathologie a intégré des approches de génétique des populations, qui, de manière complémentaire à l'épidémiologie, permet de déterminer l'origine des épidémies et distinguer les structures génétiques spatiales des populations d'agents phytopathogènes (Gladieux *et al.* 2011).

1.1. L'épidémiologie

En phytopathologie, les recherches s'intéressent à la fois à la sélection de plantes résistantes aux maladies et aux agents pathogènes, ainsi qu'aux maladies qu'ils propagent. Quantifier les fluctuations démographiques des agents pathogènes, leur dispersion et la transmission des maladies qu'ils propagent sont les principaux objectifs de l'épidémiologie

(Sache 2003). Les maladies sont définies à la fois par les caractéristiques phénotypiques des agents pathogènes et par les symptômes qu'ils induisent sur les hôtes, ainsi la quantification des maladies s'effectue soit de manière directe en déterminant l'abondance de l'agent pathogène lui-même, soit de manière indirecte en quantifiant l'incidence de la maladie sur l'hôte (à travers la quantité de symptômes qu'il développe) (Sache 2003).

Toutefois ces études en se cantonnant à des analyses démographiques ou phénotypiques sont souvent impuissantes pour démêler l'histoire évolutive des agents pathogènes. Les progrès dans les techniques et analyses moléculaires permettent d'ajouter une dimension en connaissant notamment l'origine génétique des individus étudiés (Archie *et al.* 2009). Ainsi, les problématiques majeures ont pu évoluer en particulier grâce à l'apport des approches moléculaires (De Meeûs *et al.* 2007). Dans ces analyses plus modernes sur la biologie des populations des agents phytopathogènes, un rôle majeur est tenu par des approches de génétique des populations.

1.2. Les apports de la génétique des populations pour l'étude des champignons phytopathogènes

1.2.1. Principe de la génétique des populations

La population est l'unité d'étude de la génétique des populations. Ce terme peut être défini de plusieurs manières, mais en général on considère qu'une population est représentée par un groupe d'individus appartenant à la même espèce, situés au même endroit et à la même période. Le principe de la génétique des populations réside dans l'étude des fréquences alléliques portées par les individus (définissant la diversité génétique) et de ses fluctuations au sein et entre populations d'individus collectés puis caractérisés à l'aide de marqueurs génétiques. L'utilisation de méthodes de génétique des populations permet de mettre en évidence les processus évolutifs qui façonnent la présence ou l'absence de structures génétiques au sein des populations d'une espèce. Ces processus évolutifs se divisent en cinq facteurs : la mutation, la dérive génétique, les flux de gènes, la recombinaison et la sélection. Leur importance est relative selon les cas étudiés et leur action agit sur la formation, le maintien ou l'absence de structure génétique.

1.2.2. La génétique des populations chez les champignons phytopathogènes

Les champignons phytopathogènes sont soumis aux mêmes forces évolutives que l'ensemble des autres espèces (McDonald & Linde 2002), pourtant les études de génétique des populations qui leurs sont consacrées sont assez rares. En effet l'acquisition de données sur ces micro-organismes est parfois rendue délicate en raison de leur petite taille, ce qui rend leur observation et leur échantillonnage plus fastidieux. L'observation à l'aide de méthodes indirectes a facilité l'étude d'organismes de petite taille tels que les champignons phytopathogènes (De Meeûs *et al.* 2007). Ces progrès ont été rendus possibles par l'apport des outils de la biologie moléculaire, notamment par les techniques d'amplification (PCR) des loci d'intérêt qui permettent d'obtenir des données même à partir de quantités minimales d'ADN.

Bien que l'approche soit moins aisée par rapport aux macro-organismes, les intérêts des études de génétique des populations chez les champignons phytopathogènes n'en sont pas moins riches, en partie en raison de la complexité de leur biologie et de leur cycle de vie et des pressions de sélection qu'ils subissent (McDonald & Linde 2002). Nous nous sommes intéressés à trois thèmes particuliers illustrant le couplage de l'épidémiologie et de la génétique des populations chez les champignons : la diversité des modes de reproduction, la sélection (ou l'adaptation) et la quantification de la dispersion.

1.2.2.1. Modes de reproduction

Les champignons présentent une très large diversité de modes de reproduction (Taylor *et al.* 1999 ; Burdon & Roelfs 1985, Samils *et al.* ; 2001 ; Jin *et al.* 2010). Certaines espèces se reproduisent de manière sexuée, asexuée ou alternent des phases de reproduction sexuée et asexuée durant leur cycle biologique. Au sein d'une même espèce, on peut mettre en évidence une dualité des modes de reproduction, avec des lignées uniquement clonales et d'autres qui s'engagent de façon plus ou moins importante dans la reproduction sexuée (Goyeau *et al.* 2007 ; Duan *et al.* 2010). Les approches de génétique des populations permettent de déterminer le type de reproduction. Par exemple l'existence de lignées uniquement clonales peut être mise en évidence grâce à l'utilisation d'indices de génétique des populations tels que le nombre de génotypes répétés, un grand nombre de paires de loci en déséquilibre de liaison et une valeur de F_{IS} très négative (= forte hétérozygotie) (Balloux *et al.* 2003 ; Halkett *et al.* 2005). Ces indices permettent de distinguer entre différentes formes de clonalité, il est ainsi possible de faire la différence entre une lignée uniquement clonale et une lignée qui se multiplie de manière asexuée seulement pendant une phase de son cycle biologique, par

Tableau 1 : Interaction entre l'agent pathogène et la plante hôte dans le cadre de la relation gène-pour-gène. La présence d'un allèle de résistance (dominant par rapport à l'allèle de sensibilité) chez la plante et d'un allèle d'avirulence correspondant (dominant par rapport à l'allèle de virulence) chez l'agent pathogène aboutit à l'échec de l'infection (ou interaction incompatible). Dans les trois autres combinaisons, il y a interaction compatible entre la plante et l'agent pathogène et donc infection de l'hôte.

Phénotype de l'agent pathogène	Phénotype de la plante	
	Résistant	Sensible
Virulent	Compatible	Compatible
Avirulent	Incompatible	Compatible

exemple une propagation clonale pendant une phase épidémique (Bourassa *et al.* 2007). Dans ce dernier cas, une fois les génotypes répétés éliminés, les signatures indiqués ci-dessus et révélant une empreinte profonde de la reproduction asexuée ne seront pas retrouvés (retour à l'équilibre d'Hardy-Weinberg et absence de déséquilibre de liaison). A l'inverse l'utilisation de marqueurs génétiques peut aussi détecter la reproduction sexuée à l'aide de preuves d'évènements de recombinaison (Mboup *et al.* 2009). Ces caractérisations peuvent compléter les expériences menées au laboratoire (par exemple Jin *et al.* 2010) en venant corroborer l'existence ou l'absence de reproduction sexuée dans les populations naturelles.

1.2.2.2. Adaptation et Sélection

L'étude de la structuration génétique des populations d'un agent pathogène peut permettre de mettre en évidence l'adaptation des individus à leur hôte. En agronomie, la sélection de plantes résistantes aux maladies a parfois entraîné une différenciation génétique forte entre des populations d'agents pathogènes adaptées à ces plantes résistantes et d'autres non spécialisées (Giraud *et al.* 2010). L'amélioration variétale a souvent conduit à la sélection de plantes porteuses de gènes de résistance qualitative, c'est-à-dire que les agents pathogènes ne sont pas capables d'infecter ces plantes résistantes. Le mécanisme le plus communément accepté pour expliquer l'adaptation des agents pathogènes aux plantes porteuses de résistances qualitatives est la relation gène-pour-gène (Flor 1971). L'interaction est incompatible et ne conduit pas au développement de la maladie tant que l'agent pathogène ne possède pas la virulence correspondante (Tableau 1). Dans ce cas le produit du gène d'avirulence est reconnu par la plante hôte qui déclenche alors des réactions de défense en cascade. Le passage d'avirulence à virulence s'effectue par un changement génétique (de la mutation ponctuelle à la délétion d'une région chromosomique) conférant une perte de reconnaissance par la plante hôte. Ceci mène au champ à un contournement de résistance pour lequel les agents pathogènes s'adaptent à leur hôte résistant (Stukenbrock & McDonald 2009).

A la suite d'un contournement de gène de résistance, une différenciation entre populations virulentes et avirulentes peut apparaître chez l'agent pathogène (Guérin *et al.* 2007). Les individus virulents sont souvent moins diversifiés génétiquement que les individus avirulents. Ceci est dû à l'effet de fondation qui suit le goulot d'étranglement apparaissant au moment du contournement. En effet, la population virulente subit les effets combinés de la dérive (petit nombre d'individu initial) et de la pression de sélection exercée par l'hôte

(Burdon & Thrall 2008 ; Gladieux *et al.* 2008). La différenciation peut être aggravée lorsque la reproduction sexuée a lieu sur l'hôte exerçant la pression de sélection (Giraud *et al.* 2010).

Ainsi l'étude de populations infectant soit des hôtes sauvages ou cultivés (Stukenbrock & McDonald 2008) soit des hôtes cultivés de différentes variétés (Montarry *et al.* 2006 ; Goyeau *et al.* 2007) permet de suivre à la fois les conséquences génétiques (structuration) et démographiques de l'adaptation des populations d'agents phytopathogènes à leurs hôtes.

1.2.2.3. Flux de gènes et dispersion

Les spores des champignons phytopathogènes dispersées par le vent peuvent parcourir de très grandes distances (Brown & Hovmøller 2002). Une partie des études sur les flux de gènes chez les champignons phytopathogènes s'intéresse aux voies de dispersion pendant les phases d'expansion et aux distances parcourues pendant la migration (Rivas *et al.* 2004 ; Prospero *et al.* 2008 ; Barrès *et al.* 2008 ; Goss *et al.* 2009). De plus, en raison de leurs fortes capacités de dispersion, les champignons phytopathogènes peuvent envahir de nouvelles régions, ce qui peut avoir pour conséquence l'émergence de nouvelles maladies si les conditions environnementales sont favorables (Desprez-Loustau *et al.* 2007).

Le modèle couramment utilisé en génétique des populations pour estimer les distances de dispersion est le modèle d'isolement par la distance, qui corrèle la différenciation génétique (résultant d'une balance entre migration et dérive) avec la distance géographique entre paires de populations (Rousset 1997 ; 2000). Sous un certain nombre d'hypothèses, la pente de la droite de régression peut être utilisée pour inférer la distance de dispersion. Cependant on se retrouve souvent dans l'impossibilité de dissocier le produit $D\sigma^2$ (D densité en individus et σ^2 variance de dispersion entre deux générations) proportionnel à la valeur de la pente. De plus ce modèle est sensible aux déséquilibres génétiques résultant des événements démographiques qui peuvent avoir lieu dans les populations (Leblois *et al.* 2004). C'est pourquoi il devient nécessaire de prendre en compte la démographie car ses conséquences peuvent influencer les structures génétiques. Ainsi, il est possible d'inférer les processus de dispersion (par exemple dispersion graduelle *versus* stochastique) et les voies de migration, d'après les conséquences génétiques des événements de colonisation (Barrès *et al.* 2008, Halkett *et al.* 2010). Ceci n'est malheureusement pas toujours compatible avec les problématiques courantes en épidémiologie végétale, car les études ne s'effectuent pas aux mêmes échelles de temps ou d'espace, les échelles spatiales considérées étant souvent beaucoup plus larges en génétique des populations.

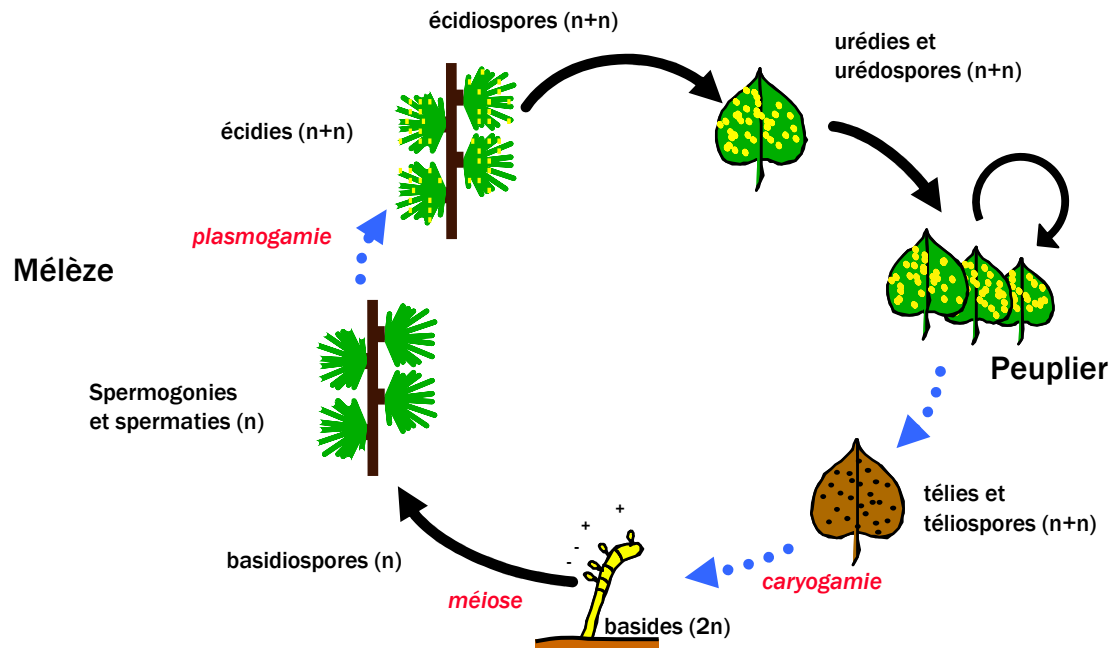


Figure 1: Cycle biologique de *M. larici-populina* incluant ses deux hôtes : mélèze et peuplier. Les flèches noires et bleues correspondent respectivement à des événements de dispersion ou de changement d'état.

La combinaison d'approches couplées d'épidémiologie et de génétique des populations permet d'obtenir une vision plus complète pour comprendre les interactions entre les processus démographiques et les caractéristiques génétiques des populations d'une espèce. La rouille du peuplier provoquée par l'agent pathogène *Melampsora larici-populina* représente un pathosystème particulièrement riche pour ce type d'études.

2. Le pathosystème Rouille / Peuplier

2.1. *Melampsora larici-populina*, l'agent de la rouille du peuplier

2.1.1. Biologie de *M. larici-populina*

Melampsora larici-populina, l'agent responsable de la rouille du peuplier, est un champignon biotrophe obligatoire, appartenant aux basidiomycètes, de la classe des téliomycètes et de l'ordre des Pucciniales (autrefois appelées Urédinales). Le genre *Melampsora* compte plus d'une cinquantaine d'espèces dont 17 sont pathogènes sur les peupliers (Vialle *et al.* soumis). Sur les neuf espèces présentes en Europe, celle qui cause le plus de dégâts dans les peupleraies françaises et européennes est *M. larici-populina* (Pinon & Frey 2005), notre objet d'étude.

2.1.2. Cycle de vie de *M. larici-populina*

Le cycle de vie de *M. larici-populina* est caractéristique d'une rouille hétéroïque (c'est-à-dire qu'elle accomplit son cycle en infectant deux plantes-hôtes d'espèces différentes) et macrocyclique (c'est-à-dire que cinq types de spores sont produits au cours du cycle) (Fig. 1). L'alternance d'hôtes coïncide avec l'alternance des modes de reproduction, la reproduction sexuée a lieu une fois par an sur l'hôte mélèze (l'hôte écidien) au début du printemps, suivie de plusieurs cycles de multiplication asexuée qui ont lieu du printemps à l'automne sur l'hôte peuplier (l'hôte télien).

A la sénescence automnale des feuilles, des croûtes noires se forment sous l'épiderme de la face supérieure des feuilles de peuplier, ce sont les télies contenant des télisporos dicaryotiques. Il s'agit d'une forme de résistance sous laquelle *M. larici-populina* passera l'hiver. Au début du printemps, la reproduction sexuée débute par la caryogamie, qui a lieu au sein des télisporos pour former des basides diploïdes. Ces basides subissent une méiose et

produisent des basidiospores haploïdes qui infectent des aiguilles de mélèze. Quelques jours plus tard, ces infections donnent naissance à des spermogonies. La reproduction sexuée se poursuit sur le mélèze par plasmogamie entre une spermatie produite par une spermogonie d'un signe sexuel et une spermogonie du signe sexuel opposé. De nouvelles fructifications dicaryotiques naissent de cette plasmogamie, les écidies, qui apparaissent sous forme de petites pustules jaunes sur la face inférieure des aiguilles de mélèze. Les écidies produisent des écidiospores dicaryotiques qui sont disséminées par le vent et vont infecter des feuilles de peuplier. Une dizaine de jours après l'infection, des urédies, pustules de couleur jaune, apparaissent sur la face inférieure des feuilles de peuplier. C'est alors que débute la phase de multiplication asexuée, les urédies produisant de très grands nombres d'urédiospores dicaryotiques capables d'infecter à nouveau des feuilles de peuplier. Cette phase se déroule de fin mai à octobre et de nombreux cycles de multiplication asexuée peuvent avoir lieu, à l'origine des épidémies de rouille, ceci jusqu'à la chute des feuilles automnale.

En climat tempéré, la phase de reproduction sexuée est obligatoire car la survie asexuée sur feuilles de peuplier est rendue impossible par la chute de celles-ci. La phase de reproduction sexuée annuelle assure un brassage génétique important.

Les hôtes écidien et télén de *M. larici-populina* sont deux espèces phylogénétiquement très éloignées (Gymnospermes / Angiospermes) et présentent des caractéristiques biologiques particulières à chacune d'elles, ayant des conséquences sur la biologie des populations de *M. larici-populina*.

2.2. L'hôte écidien, le mélèze

Le mélèze (*Larix* sp.) est l'un des rares conifères à feuilles caduques. L'aire de répartition naturelle du mélèze d'Europe (*Larix decidua*) est assez limitée, comprenant les massifs montagneux d'Europe centrale (incluant l'arc Alpin, les Carpates, les Sudètes et les Tatras) ainsi que les plaines du nord de la Pologne (Riou-Nivert 2001). Cependant l'aire de répartition du mélèze s'est étendue en raison de son utilisation en sylviculture ou en ornement, depuis le dix-huitième siècle, si bien qu'actuellement le mélèze est présent sur l'ensemble du territoire français (bien que moins densément que dans les Alpes) (Riou-Nivert 2001). L'infection des aiguilles de mélèze par *M. larici-populina* n'a aucune incidence sur la croissance ou le développement des arbres infectés.

2.3. L'hôte télien, le peuplier

En France comme dans le reste de l'Europe, les peupliers sont à la fois présents à l'état sauvage (*Populus nigra*, *P. alba* et *P. tremula*) et à l'état cultivé sous la forme de cultivars (aussi appelés clones car multipliés par voie végétative) le plus souvent hybrides. Il s'agit d'hybrides euraméricains (*P. deltoides* x *P. nigra*) et interaméricains (*P. deltoides* x *P. trichocarpa*) issus du croisement entre espèces de peupliers européens et américains. Les peupliers cultivés et le peuplier noir (*P. nigra*) sont soumis aux infections de *M. larici-populina*, mais la maladie n'a pas le même impact pour ces deux types d'hôtes (Frey *et al.* 2005).

2.3.1. Le peuplier noir

Le peuplier noir est présent dans toute l'Europe ainsi que dans l'ouest de l'Asie. En France, cette espèce héliophile se trouve principalement au sud de la Loire. Les peupliers aiment croître sur des sols humides et profonds caractéristiques des bords de cours d'eau. Sa reproduction à la fois par voie sexuée et asexuée en fait un bon colonisateur, en particulier des berges. L'infection par *M. larici-populina* n'a en général qu'une faible incidence sur les feuilles de peuplier noir (Frey *et al.* 2005, Legionnet *et al.* 1999).

2.3.2. Peupliers cultivés

Le peuplier représente une des premières essences de feuillus en terme de récolte grâce à sa croissance rapide. De ce fait, la populiculture française revêt une importance économique non négligeable. Les grandes régions populicoles françaises sont le Nord-Pas-de-Calais, la Picardie, la Champagne-Ardenne, le Centre, l'Ile-de-France et l'Aquitaine. Les peupleraies cultivées sont monoclonales (un seul cultivar) et équiennes (âge identique) à l'échelle de la parcelle, et souvent pauciclones dans une même région, donc représentent une très faible diversité génétique.

La résistance aux maladies, notamment à la rouille, est un des principaux critères de sélection des cultivars de peupliers. Deux types de résistance se distinguent : la résistance qualitative et la résistance quantitative. La résistance qualitative empêche l'infection de la plante par l'agent pathogène, selon le modèle de la relation gène-pour-gène proposée par Flor (1971). La sélection de peupliers a longtemps favorisé des cultivars porteurs de résistance qualitative ; en revanche aucune résistance qualitative n'a été mise en évidence chez le

peuplier noir (Legionnet *et al.* 1999, Pinon & Frey 2005). Le déploiement de cultivars de peuplier à résistance qualitative a rapidement conduit au contournement de ces résistances par des individus de *M. larici-populina* virulents. L'étude des individus de *M. larici-populina* contournant ces résistances qualitatives a permis d'identifier huit facteurs de virulence (ou virulences) (Pinon & Frey 2005). Il en résulte que la composante qualitative du pouvoir pathogène d'un individu de *M. larici-populina* est étudiée par l'inoculation au laboratoire d'une gamme différentielle de huit clones de peupliers, portant chacun une des résistances qualitatives. Le phénotype (ou pathotype) de l'individu correspond à la combinaison des virulences obtenues.

L'impact de la rouille est plus important dans les peupleraies cultivées que dans les peupleraies sauvages. En effet, une fois la résistance qualitative contournée, la plupart des cultivars de peupliers possèdent une faible résistance quantitative et sont donc très sensibles aux attaques d'individus de *M. larici-populina* porteurs de la virulence correspondante. Ces infections entraînent des chutes précoces des feuilles, donc une réduction de croissance et par conséquent d'importantes pertes économiques. Il en est de même en Amérique du Nord, où les peuplements sauvages de *P. deltoides* en ripisylve sont moins infectés par *M. medusae* f. sp. *deltoidae* que les plantations commerciales (Prakash & Thielges 1987, Chen & Harrington 2006).

2.4. Etat des connaissances sur la structure des populations de *M. larici-populina*

Les études précédentes se focalisant sur la structure génétique des populations de *M. larici-populina* se sont généralement intéressées à trois axes majeurs, tirant parti des caractéristiques particulières du pathosystème.

Tout d'abord, la présence jointe d'hôtes sauvages et cultivés sur le territoire a permis l'étude des interactions et la comparaison entre des populations de *M. larici-populina* infectant des peupliers sauvages ou cultivés (Pinon & Frey 1997 ; Gérard *et al.* 2006). Les premières études ne s'intéressaient qu'à la comparaison des profils phénotypiques en ne regardant que la composition en virulences (Pinon & Frey 1997). Ces premiers résultats indiquaient une plus faible diversité en virulences chez les populations prélevées sur peupliers sauvages et une plus grande richesse avec une accumulation de plusieurs virulences, parfois inutiles, chez les individus collectés dans des peupleraies cultivées. Plus récemment, des

informations génétiques sont venues compléter les informations données par les profils phénotypiques (Gérard *et al.* 2006). Cependant ni l'utilisation de marqueurs RAPD, ni celle de marqueurs microsatellites n'a révélé de distinction franche entre les profils génétiques des individus infectant les hôtes sauvages ou cultivés (Gérard *et al.* 2006, Barrès *et al.* 2008), ceci alors que des différences de diversité et de complexité des profils de virulence sont toujours observées (Gérard *et al.* 2006, Barrès *et al.* 2008).

D'autre part, l'étape de reproduction sexuée sur mélèze a soulevé des questions sur l'influence de la proximité entre mélèzes et peupliers. L'hypothèse d'une plus grande diversité génétique des individus collectés sur des peupliers proches de mélèzes a été testée, mais aucune différence n'a été constatée entre les individus infectant des peupliers proches ou éloignés de mélèzes (Gérard *et al.* 2006). En revanche, une plus grande diversité en pathotypes a été trouvée dans les populations de *M. larici-populina* collectées sur peupliers à proximité de mélèzes (Frey *et al.* 2005)

Enfin, ces études ont permis de mettre en évidence que les spores de *M. larici-populina* possédaient des fortes capacités de dispersion, ce qui avait déjà été soupçonné en raison de la présence de *M. larici-populina* dans la plupart des pays pratiquant la populiculture en dehors de l'Europe, et des effets de fondation ont été mis en évidence pour des populations de *M. larici-populina* collectées en Islande et au Canada (Barrès *et al.* 2008). Le signal d'isolement par la distance détecté à l'échelle continentale indique que la dispersion s'effectue de manière graduelle (Barrès *et al.* 2008).

2.5. Le site de la Vallée de la Durance

Une majeure partie de ce travail de thèse s'est focalisée sur une région présentant une situation écologique particulière, la vallée de la Durance. Elle a retenu notre attention car tous les thèmes présentés ci-dessus peuvent y être abordés : la relation entre individus infectant des peupliers sauvages et cultivés, la contribution de la reproduction sexuée à l'évolution des caractéristiques génétiques des populations et enfin les capacités de dispersion de *M. larici-populina* pendant une phase épidémique et ses conséquences le long de la vallée.

Cette vallée est située dans les Alpes, dans l'aire de répartition naturelle des mélèzes. La partie la plus en amont, le Briançonnais, constitue une zone de sympatrie entre peuplier et mélèze qui confine la reproduction sexuée de *M. larici-populina* uniquement dans cette zone. En aval de cette zone, la Durance, qui s'écoule vers le sud, est bordée de manière quasi-continue d'une ripisylve constituée principalement de peupliers noirs sur environ 200 km. Ce système forme un corridor écologique qui facilite le suivi des épidémies de rouille du peuplier le long de la Durance. Cependant la présence de quelques rares peupleraies cultivées à la fois dans la zone de sympatrie peuplier – mélèze et dans le reste de la vallée complexifie l'étude de ce système. Des observations épidémiologiques réalisées depuis 2001 nous ont permis de noter que les peupliers présents dans la vallée de la Durance étaient annuellement infectés, probablement à partir d'un inoculum provenant de la zone de sympatrie peuplier – mélèze en amont de la vallée.

Objectifs de la thèse

OBJECTIFS DE LA THESE

Ce travail s'articule autour de plusieurs problématiques majeures, mais toutes ont pour principal objectif de mettre en lumière l'influence relative de différents facteurs sur la structure génétique et la dynamique des populations de *M. larici-populina*. Dans ce but, nous avons mêlé à la fois des approches de génétique des populations et d'épidémiologie. Une attention particulière a été portée à l'échelle géographique du site de la vallée de la Durance. Cette vallée constitue un corridor écologique favorable à l'étude conjointe de la démographie de *M. larici-populina* et de ses conséquences génétiques.

Une première étape a permis d'élaborer des outils moléculaires efficaces pour la réalisation d'études précises de génétique des populations. Ainsi le premier objectif de ce travail a été d'enrichir le panel de marqueurs microsatellites déjà existant (Barrès *et al.* 2006) afin de disposer d'un grand nombre de marqueurs suffisamment polymorphes pour pouvoir discriminer clairement des groupes génétiques lors de tests d'assignation.

Dans un second temps, nous avons souhaité connaître l'état actuel de la structure génétique des populations de *M. larici-populina* infectant l'hôte sauvage à travers la France avant d'entreprendre des études à des échelles plus fines. Des études précédentes ont montré que les populations de *M. larici-populina* ont été fortement impactées par le déploiement du cultivar 'Beaupré', porteur de la Résistance 7, dans le paysage populiicole français. Les différentes étapes du contournement de la Résistance 7 suivies par l'expansion des individus porteurs de la Virulence 7 ont été très bien décrites (Pinon *et al.* 1998). Notre problématique était de déterminer quel est l'impact de la présence des hôtes cultivés, en particulier du cultivar 'Beaupré', sur la composition génétique et phénotypique des populations de *M. larici-populina* infectant l'hôte sauvage, *P. nigra*. Les groupes génétiques sont-ils structurés spatialement en fonction des grandes zones de populiculture en France ?

Dans le troisième chapitre, une étude à une échelle plus restreinte a été réalisée, dans une région des Alpes où peupliers et mélèzes évoluent en sympatrie. Dans cette région, où le relief est fortement marqué, une barre montagneuse, le Massif des Ecrins, traverse

Objectifs

transversalement le site d'étude. Précédemment des individus appartenant à des groupes génétiques distincts avaient été détectés dans cette zone d'étude. Dans ce chapitre, nous avons cherché à savoir de quelle manière s'effectuait la démarcation entre les domaines spatiaux des groupes génétiques. S'agit-il d'une délimitation continue avec une zone de coexistence ou d'une discontinuité franche reliée à un aspect du paysage ?

Grâce aux conclusions des premiers chapitres, nous avons pu étudier l'épidémie de *M. larici-populina* dans la vallée de la Durance. Cette étude s'est déroulée en plusieurs étapes. Tout d'abord comme ce système est complexifié par la présence de peupliers sauvages et cultivés, nous avons cherché à connaître la composition génétique des individus de *M. larici-populina* qui forment l'épidémie, c'est-à-dire si ces individus étaient homogènes ou s'ils appartenaient à des groupes génétiques différents.

Dans un second temps, nous avons cherché à décrire le plus finement possible le déroulement de l'épidémie, suivie par une étape de modélisation du déroulement de l'épidémie.

Enfin, dans un dernier temps ; nous sommes focalisés sur l'évolution des caractéristiques génétiques des individus au cours de la phase épidémique. Nous nous sommes principalement intéressés aux conséquences génétiques que pouvaient avoir eu la dispersion à la fin de l'épidémie.

Matériel et Méthodes

MATERIEL ET METHODES

1. Introduction : méthodes de biologie moléculaire et principe des analyses de génétique des populations réalisées

Ce chapitre a pour but de présenter le développement de marqueurs génétiques neutres et leur utilisation dans les analyses de génétique des populations. Une attention particulière a été portée sur la description des tests d'assignation basés sur des algorithmes Bayésiens.

Pour caractériser des populations, deux types de marqueurs peuvent être utilisés, des marqueurs neutres et des marqueurs sélectionnés. Les marqueurs sélectionnés révèlent le phénotype des individus. Dans le cas de *M. larici-populina*, on peut utiliser par exemple le profil de virulences (ou pathotype), qui détermine l'aptitude des individus à infecter ou non des hôtes porteurs de la résistance qualitative correspondante. A l'inverse l'utilisation de marqueurs neutres permet d'éviter dans une certaine mesure l'influence de la sélection (bien que ce ne soit pas toujours vrai, voir Selkoe *et al.* 2006) pour se concentrer sur les conséquences des événements démographiques (balance entre la dérive et la migration).

Depuis une quinzaine d'années, l'utilisation des marqueurs microsatellites est devenue quasiment incontournable (Vignal *et al.* 2002). D'une part, les techniques de génétique moléculaire ont subi de nombreuses améliorations, en particulier avec l'usage de la PCR (Polymerase Chain Reaction) qui permet l'amplification des marqueurs à partir d'une faible quantité d'ADN initial (Jarne & Lagoda, 1996 ; Sunnucks 2000 ; Selkoe *et al.* 2006). D'autre part, les logiciels d'analyses sont devenus plus performants en particulier grâce à l'apport des méthodes Bayésiennes et des avancées en informatique.

Pour caractériser les populations de *M. larici-populina*, quinze marqueurs microsatellites avaient été développés précédemment par notre équipe au cours de la thèse de Benoît Barrès selon la méthode des banques enrichies (Barrès *et al.* 2006). Mais ceux-ci n'étaient pas suffisamment polymorphes et trop peu nombreux pour mettre en évidence des structures génétiques fines. Notamment ce lot de marqueurs ne permettait pas de discriminer correctement des individus (faible puissance des tests d'assignation), alors qu'une différenciation significative se dégageait entre deux groupes formés par des individus aux profils de virulence contrastés (Barrès 2006). C'est pourquoi le développement de nouveaux

marqueurs microsatellites a représenté une première étape importante de ma thèse et un pré-requis pour pouvoir mieux caractériser les populations de *M. larici-populina*. Le séquençage du génome de *M. larici-populina* par le Joint Genome Institute (Département de l’Energie Américain) en 2007 (Duplessis *et al.* 2011) a grandement facilité cette étape et a permis de développer un très grand nombre de marqueurs microsatellites très polymorphes.

La première partie de ce chapitre est consacrée à la description du développement de nouveaux marqueurs microsatellites. Cette partie se présente sous la forme d’un article publié dans la revue *Conservation Genetics Resources*. La seconde partie est consacrée aux analyses en génétique des populations, plus particulièrement aux analyses Bayésiennes, récemment développées et bien adaptées à nos questions biologiques.

2. Développement de nouveaux marqueurs microsatellites

ARTICLE 1 : CHARACTERIZATION OF 41 MICROSATELLITES LOCI DEVELOPED FROM THE GENOME SEQUENCE OF THE POPLAR RUST FUNGUS, *MELAMPSORA LARICI-POPULINA*

Article publié en 2009 dans la revue *Conservation Genetics Resource* 1 : 21-25.

Constance XHAARD, Axelle ANDRIEUX, Fabien HALKETT, Pascal FREY.

INRA, Nancy-Université, UMR1136 Interactions Arbres – Microorganismes, IFR 110, F-54280 Champenoux, France.

Keywords : Basidiomycete, *Melampsora larici-populina*, poplar rust fungus, genome sequence, microsatellites, M13-tailed primer method.

Correspondence: P. Frey, Fax: +33 383 394069; E-mail: frey@nancy.inra.fr

Abstract

Forty-one microsatellite markers were developed for the poplar rust fungus *Melampsora larici-populina*. The genome sequence of *M. larici-populina* strain 98AG31 was screened *in silico* for microsatellite loci using MAGELLAN software. Thirty-nine loci, evaluated on a sexual population of 35 *M. larici-populina* isolates sampled on larch trees in the French Alps, were identified to be highly polymorphic. These markers revealed two to 34 alleles, with an average of 7.23 alleles per locus. The estimated heterozygosity ranged from 0.029 to 0.958 across the 41 microsatellite loci. These markers will be highly useful for population genetics studies of *M. larici-populina*.

Table 1 Characterization of the 41 microsatellite loci developed for *Melampsora larici-populina*

Locus name	Repeat motif	Genome position ^a	Primer sequences (5'-3')	Expected size (bp) ^b	No. of individuals genotyped	No. of alleles	Size range (bp) ^c	H _e	H _o
MLP40	(AAAACA)24	29/398457	F: M13-GTGTGACATCACACCAAAATGG R: CACCTGTTTTCTCTTTGATCC	467	35	19	391–559	0.863*	0.629
MLP41	(TTGTTC)24	11/105864	F: M13-CAAACCGATCTTGAGTAGTCACC R: GAACCTATGTTGAACCGATTTCC	285	35	26	168–432	0.941	0.857
MLP42	(ACAATC)19	1/3608849	F: M13-CTAAAGATTTCACACCAGTTGAACC R: GTTGAAAAGTAGGCACCAATCC	344	35	13	244–428	0.885*	0.486
MLP43	(TGTCAT)15	4/1205881	F: M13-GGCAAGAATAAAAAGGACAGTGC R: TCCTTTTCTACTGCCCTTAGATGG	227	35	34	256–522	0.958	0.943
MLP48	(GATT)18	38/711722	F: M13-AAAATGAAGAAAAGAGTCATGAAATG R: AAGAAAGAAGAAGTGGTGATGATTG	209	34	13	212–292	0.726	0.559
MLP49	(GAT)18	37/803561	F: M13-TATGACTGAATCATCCGTCGAAC R: GACACTTTTACACCGATCAATCC	340	34	12	368–425	0.814	0.794
MLP50	(TCT)16	44/484187	F: M13-CAGCGTCTTGATCTTCTTCTCC R: TACCATTACCATTACCACCAACC	279	34	5	300–312	0.496	0.441
MLP54	(ATG)14	8/807623	F: M13-GTTTGCTGGTTTACCTTCTGC R: ACCACCCTTCTCCTCTTTCC	142	35	4	146–155	0.680	0.743
MLP55	(ATC)15	1/2804879	F: M13-GCAAGGAGATGAGTCAAGACG R: CAGTCAGCTCGGTTATCATCC	153	35	5	170–182	0.610	0.571
MLP56	(AAC)7	5/2550898	F: M13-GAGGTGTGAGTCGAGAGTTGG R: GATAGTGTGGAAGAGATTGGTGACC	292	35	7	291–318	0.639	0.714
MLP57	(CAA)13	2/1943755	F: M13-TTGAAGGCTGAGACTGAAACC R: ACCCGATCAGTACTTAATTCAACC	169	35	6	172–202	0.264	0.286
MLP58	(AAG)12	3/728772	F: M13-TCATCAGCAGTGTGTTGTGTTGG R: CAACTCAATGGACTCAAGATGG	251	35	6	258–276	0.701	0.714
MLP59	(ATC)13	8/1544938	F: M13-ATCAAGATCTCAAACCAATCC R: TTGATGTTGAGGTTGATGTGG	322	35	7	337–355	0.777	0.514
MLP60	(ATC)13	12/201001	F: M13-GGAGAAAGGATTATTGGGAAAAGG R: GGAGGGCACTGTACAATCTAACC	164	31	5	162–186	0.624*	0.355
MLP61	(ATG)11	15/1174383	F: M13-TGGGTTTGATTGGGTATAACG R: CATTCCCAATTCAAGCAGACC	248	35	3	273–279	0.343	0.371
MLP62	(TATATT)13	75/26274	F: M13-AAGTGACAGCACAAAGTGACAGG R: CATTGGACAATGAGTGAGTGC	466	33	12	411–517	0.874	0.849
MLP64	(TGT)12	64/490100	F: M13-ACTTCAAGTGTGGAAGGATCC R: TGAATGATGAAGAGTTGCTTGG	401	35	6	409–427	0.520*	0.286

The Eurasian poplar rust fungus, *Melampsora larici-populina*, is a heteroic and macrocyclic pathogen, which has its sexual (aecial) stage on larch (*Larix* sp.) and its asexual (uredinial) stage on poplar (*Populus* sp.). During the last decades, this pathogen has caused important economic losses to poplar cultivation in Europe due to the breakdown of several major resistance genes (Gérard *et al.* 2006). Fifteen microsatellite loci have already been developed for *M. larici-populina* using enriched libraries (Barrès *et al.* 2006). Eleven of these loci proved very useful to characterize the genetic structure of *M. larici-populina* populations at continental and intercontinental scales (Barrès *et al.* 2008). Despite substantial polymorphism these markers revealed insufficient to resolve fine-scale population genetics studies.

Completion of the *M. larici-populina* genome sequence by the US-DoE Joint Genome Institute provided an alternative method for microsatellite discovery via computational whole-genome scan. The genome sequence of *M. larici-populina* strain 98AG31 (available at <http://genome.jgi-psf.org/Mellp1/Mellp1.home.html>) was analysed for microsatellite content using MAGELLAN software (<http://www.medfac.usyd.edu.au/people/academics/profiles/dcarter.php>). Detection criteria were constrained to motifs of length 2-6 bp and a minimum repeat number of 10, 8, 6, 6, and 6, respectively. Using these criteria, 654 microsatellite loci were identified in the 102 Mb genome sequence of *M. larici-populina*, of which 58 loci were further developed. Primers were designed using PRIMER3 software (Rozen & Skaletsky 2000). The 58 primer pairs were tested on a population of 35 *M. larici-populina* isolates collected in May 2007 at Prelles, French Alps (Barrès *et al.* 2008). This population was sampled at the aecial stage on larches, just after sexual reproduction, to avoid clonal interference,

Microsatellite loci were amplified in a GeneAmp PCR System 9700 Thermocycler (Applied Biosystems) under the following conditions: an initial denaturing step of 5 min at 95°C; 35 cycles including 30 s initial denaturation at 94°C, 90 s annealing at 60°C and 60 s extension at 72°C; and a 30 min final extension step at 60°C. PCR was carried out in 20 µL final reaction volumes containing 2 µL DNA, 2 µL of 10X reaction buffer, 1.5 mM MgCl₂, 0.2 mM dNTP, 0.5 U *Taq* polymerase (Sigma), 0.2 mM reverse primer, 0.02 mM M13-tailed forward primer and 0.2 mM fluorescence-labelled M13 primer (5'-TGTAACGACGGCCAGT-3', Proligo). Amplified PCR products were separated, sized and analysed on a CEQ8000 Genetic Analysis System (Beckman Coulter).

Out of the 58 loci tested, 41 (71%) amplified correctly, two were monomorphic (MLP75 and MLP84), and 39 were polymorphic with two to 34 alleles and an average of 7.23

Table 1 continued

Locus name	Repeat motif	Genome position ^a	Primer sequences (5'-3')	Expected size (bp) ^b	No. of individuals genotyped	No. of alleles	Size range (bp) ^c	H _e	H _o
MLP66	(GAT)12	6/2141452	F: M13-CTGGGAAATGGATGGTATGC R: CATTCACGTCAACATATAGCAAGG	195	35	5	205–220	0.688	0.686
MLP68	(TTGA)12	51/296526	F: M13-TGGATCTCAGGATGAAATTGG R: AACTTTGACCTAAGATTGGAGACG	411	34	3	433–445	0.262	0.235
MLP70	(TA)12	7/191822	F: M13-CGTCGATAAAAAACCCCTTAGC R: TGAAGGAAGAATAGAAGGTGAAGG	362	35	6	374–386	0.474*	0.200
MLP71	(AG)14	13/1266972	F: M13-CAACCATCAACTGCTTTATCAGG R: GTTGAATTTTGACCGATGAGG	369	34	7	385–397	0.662*	0.471
MLP73	(AC)14	38/809210	F: M13-ACCAAACCAAGAAAGCATGG R: GCCTGATCAACTTTCTTTTCC	434	33	12	442–474	0.870	0.636
MLP74	(AT)13	45/307816	F: M13-GCAATCAACATGCCTTTATCG R: TGTGCATAGAATCTGGGTTC	377	33	5	394–408	0.594	0.606
MLP77	(AT)10	15/629856	F: M13-AAAGAGCCCTACGAGAATTGG R: TCAACCTTTTGGAGAAGATGG	437	35	3	456–460	0.465	0.429
MLP80	(GAT)8	34/344557	F: M13-CATCCCAAGAACAAGAAACC R: TCTACTGCTCGTCATTCTACCG	404	35	2	414–417	0.029	0.029
MLP81	(CTA)8	118/40239	F: M13-GGTCCTGATTGGTAGCAGAAGC R: TCCTTCTAGCTCAGCCTGTCC	146	34	3	160–169	0.513	0.559
MLP82	(TAC)10	15/1199792	F: M13-CTGATCATCAACCACTTCAACC R: TGAACCTCTTGATGGTGTGG	149	34	7	163–202	0.724	0.735
MLP83	(ATG)10	46/418977	F: M13-GCAGGTGAAGGTAACATTCTATTATCC R: ACCAAGAGCTCAAGCAGTCG	159	33	5	169–181	0.573	0.424
MLP86	(TGA)10	26/747564	F: M13-GGTTTTTGGTCTTGATGAATGG R: CAAACCCTTCCAGAAGTTTACC	348	34	2	371–374	0.029	0.029
MLP87	(TGT)8	11/834087	F: M13-AGAAGCTGGATCTGGAATCG R: GAGCTAATTGAGGCCATAAACC	350	35	4	365–392	0.478	0.371
MLP88	(AGT)9	47/569731	F: M13-GCGAGTGAATTAACCAAAACC R: CAGAATGAGAAGAACATGTGTGG	443	34	4	436–451	0.506*	0.294
MLP89	(ACC)9	4/1469693	F: M13-AGCCCAATGGGATAAGAAGG R: CCCTTTACCAAAACCATTTCC	376	35	3	394–451	0.084	0.086
MLP91	(GTT)10	40/444668	F: M13-CCACATCTTGTGGTGAGAAGC R: TTGCTTTGATCGTGTCTTCG	137	35	2	151–154	0.135	0.143
MLP92	(TTG)11	26/241476	F: M13-ATTGGACTTGATGATGGTTCG R: GAGAGGATTCGGAATGAGTGC	336	35	3	353–362	0.402	0.429

alleles per locus (Table 1). Expected and observed heterozygosities were computed using GENETIX (Belkhir *et al.* 2000); Hardy-Weinberg equilibrium and linkage disequilibrium were tested using GENEPOP (Raymond & Rousset 1995). Tests for the presence of null alleles were performed in MICRO-CHECKER 2.2.3 (van Oosterhout *et al.* 2004). No significant linkage disequilibrium was found between all locus pairs following sequential Bonferroni correction for multiple comparisons. Seven loci showed significant deviation from Hardy-Weinberg equilibrium. This homozygote excess was in all cases due to the presence of null alleles. Cross-amplification in related species was not assessed since the rate of cross-amplification was shown to be low in *Melampsora* spp. (Barrès *et al.* 2006) and more generally in fungi (Dutech *et al.* 2007).

The microsatellite markers developed in this study will be complementary to those previously developed (Barrès *et al.* 2006) and will allow a better resolution for fine-scale population genetics analyses. Furthermore, the 596 remaining loci detected *in silico* in the genome sequence of *M. larici-populina* are currently being validated in order to build a genetic map of the fungus.

References

- Barrès B, Dutech C, Andrieux A, Caron H, Pinon J, Frey P (2006) Isolation and characterization of 15 microsatellite loci in the poplar rust fungus, *Melampsora larici-populina*, and crossamplification in related species. *Molecular Ecology Notes*, **6**, 60–64.
- Barrès B, Halkett F, Dutech C, Andrieux A, Pinon J, Frey P (2008) Genetic structure of the poplar rust fungus *Melampsora larici-populina*: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, **8**, 577–587.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, 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).
- Dutech C, Enjalbert J, Fournier E, Delmotte F, Barrès B, Carlier J, Tharreau D, Giraud T (2007) Challenges of microsatellite isolation in fungi. *Fungal Genetics and Biology*, **44**, 933–949.

Table 1 continued

Locus name	Repeat motif	Genome position ^a	Primer sequences (5'-3')	Expected size (bp) ^b	No. of individuals genotyped	No. of alleles	Size range (bp) ^c	H _e	H _o
MLP93	(GTT)11	31/73498	F: M13-CAACCGAAACACTTCATCACC R: CCTATGATGAGAGCGATGTCC	145	34	4	146–164	0.644	0.559
MLP94	(TTC)10	1/236499	F: M13-AACTCAATGGTGGTTCAGAGG R: GCAAACAACAGAAAAGAGCAAC	454	34	5	479–491	0.693	0.588
MLP95	(ACA)9	51/476411	F: M13-GGGTGGAAGTAATCGACAGG R: GAAACTTCACAACCCACAACC	450	35	4	465–480	0.593	0.543
MLP96	(TGG)9	22/789309	F: M13-GTGGTGGAATGAGAAGAATGG R: AGTTGATATTCAGGGGGTAAAGC	415	35	4	422–435	0.574	0.486
MLP97	(GATT)9	5/606572	F: M13-TTGAGGTAGACAGTTTGATTGAGG R: GAGAATGATTTGCGGGATTTGC	406	35	5	404–428	0.574	0.486
MLP75	(TC)12	5/545291	F: M13-ACGGATTGAGAAGTGGAGAGC R: AGTTCATCCAAACTGCAATGG	370	35	1	371	–	–
MLP84	(TCA)10	11/1352911	F: M13-GAAGCTAGATCGAAACCAAACG R: TTTCCCCCTTCATTCTTCTTGG	345	34	1	364	–	–

^a Scaffold/sequence position according to the *M. larici-populina* genome sequence v1.0 (<http://genome.jgi-psf.org/Mellp1/Mellp1.home.html>) retrieved on December 1, 2008. ^b Expected size is given accordingly to the genome sequence. ^c Size range actually obtained with the M13-tailed primers. H_e, unbiased expected heterozygosity; H_o, observed heterozygosity; * significant deviation from Hardy–Weinberg equilibrium after Bonferroni correction

- Gérard PR, Husson C, Pinon J, Frey P (2006) Comparison of Genetic and Virulence Diversity of *Melampsora larici-populina* Populations on Wild and Cultivated Poplar and Influence of the Alternate Host. *Phytopathology*, **96**, 1027-1036.
- Raymond M, Rousset F (1995) GENEPOP (Version-1.2) - Population-genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248-249.
- Rozen S, Skaletsky H (2000) PRIMER3 on the WWW for general users and for biologists programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386.
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535-538.

Acknowledgements:

This research was supported by grants from ECOGER (AIP 00295-298, INTERPOPGER program), ANR (ANR 07-BDIV-003, EMERFUNDIS program), INRA, and a fellowship from the French Ministry of Education and Research (MESR) to CX. We would like to thank the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>) and the *Melampsora* Genome Consortium (<http://mycor.nancy.inra.fr/genomeResources.php>) for access to the *M. larici-populina* genome sequence prior to publication. We are grateful to Sébastien Duplessis, Francis Martin, Jessy Labbé and Marie-Pierre Oudot-Le Secq (INRA, Nancy) for fruitful discussions.

3. Les assignations Bayésiennes

3.1. Définir des populations

Le terme de population est un concept central dans de nombreux champs disciplinaires de la biologie, tels que l'écologie, la biologie évolutive ou la biologie de la conservation (Waples & Gaggiotti 2009). Cependant, il n'existe pas de définition consensuelle de ce terme et chaque discipline possède sa propre définition (Pritchard *et al.* 2000). Selon un point de vue tourné vers l'écologie, la population se définit comme un ensemble d'individus de la même espèce présents en même temps dans un environnement donné et vivant en interaction. Du point de vue de la biologie évolutive, une population est constituée d'individus pouvant se reproduire entre eux, ce qui suppose également qu'ils vivent au même endroit et qu'ils sont en interaction (Waples & Gaggiotti 2009). En génétique des populations, la population constitue l'unité de base des études, et un des principaux objectifs de cette discipline est justement d'identifier des populations (ou des groupes génétiques homogènes) et d'évaluer leurs interrelations – notamment leur structuration spatiale ou temporelle – afin de mettre en évidence les facteurs qui façonnent les caractéristiques génétiques de ces populations et influencent les flux de gènes entre elles.

Avant l'émergence des méthodes d'assignation Bayésiennes, l'ensemble des individus collectés dans un même endroit était le plus souvent considéré comme formant l'unité « population » pour les analyses génétiques. Mais ce regroupement *a priori* des individus pouvait ne pas correspondre à bon nombre de situations écologiques générant un mélange (par exemple par migration) d'individus d'origines génétiques différentes. La non-prise en compte de ce partitionnement au sein des individus collectés localement peut induire de graves biais dans les analyses génétiques (Whitlock & McCauley 1999). Depuis une dizaine d'années, le développement des méthodes d'assignation Bayésiennes a permis de résoudre ce problème en définissant les populations *a posteriori* en regroupant les individus de manière à constituer des entités génétiques homogènes (Pritchard *et al.* 2000 ; Beaumont & Rannala 2004 ; Chen *et al.* 2007). Ces méthodes, très rapidement adoptées par la communauté des généticiens des populations, ont révolutionné la manière de déterminer l'appartenance des individus à des populations d'origine non connues au préalable. Ces approches, basées sur des algorithmes Bayésiens, emploient uniquement l'information portée par les fréquences alléliques des individus. La méthode consiste à minimiser le déséquilibre d'Hardy-Weinberg et les déséquilibres de liaison au sein d'un nombre (non déterminé *a priori*) d'entités génétiques ; le

meilleur partitionnement des données génétiques indiquant en retour le nombre et la nature de ces entités, définissant ainsi les « vraies » populations d'origine. L'identification du nombre et de la composition des populations à partir d'un échantillon d'individus est devenue un pré-requis incontournable aux études de génétique des populations, et de nombreux programmes d'assignation ont été développés (Excoffier & Heckel 2006).

3.2. Les programmes d'assignation

3.2.1. Diversité des programmes d'assignation

L'utilisation des programmes d'assignation s'est largement popularisée dans les études de génétique des populations (Shoemaker *et al.* 1999). Ils se sont révélés être des outils incontournables pour inférer des structures génétiques sans information *a priori* (Beaumont & Rannala 2004). Le programme STRUCTURE développé par Pritchard *et al.* (2000) a été un des programmes pionniers pour les analyses d'assignation Bayésiennes et il est devenu actuellement une référence pour ce type d'analyse (Kaeuffer *et al.* 2007). Depuis, pour répondre à des problématiques sensiblement différentes, de nombreux autres algorithmes ont été développés (Manel *et al.* 2005), tels que BAPS (Corander *et al.* 2004), GENELAND (Guillot *et al.* 2005b) ou TESS (Chen *et al.* 2007). Tous ces programmes poursuivent le même objectif, à savoir reconstruire des populations génétiquement homogènes. La majorité des algorithmes ont pour principe de recréer un équilibre d'Hardy-Weinberg et de diminuer les déséquilibres de liaison au sein des populations (Manel *et al.* 2005).

Toutefois, tous les algorithmes d'assignation ne sont pas identiques, et différentes options sont disponibles selon le programme utilisé (Excoffier & Heckel 2006). Par exemple STRUCTURE peut fonctionner de 16 manières différentes selon les options choisies (François & Durand 2010). L'apparition d'une de ces options a constitué une amélioration marquante, il s'agit de la prise en compte d'individus dont le génotype est un mélange des génotypes de plusieurs groupes parentaux, appelée « admixture » dans les programmes et que nous pouvons traduire par « ancestralité mixte ». Historiquement, les plus anciens programmes réalisaient des tests d'assignation en considérant que les individus ne pouvaient dériver que d'une seule population ancestrale. Dans ce cas le génotype entier est considéré comme l'unité de base de l'algorithme et il est assigné ou non à une population d'origine (Guillot *et al.* 2009). A l'inverse, un algorithme autorisant l'ancestralité mixte permet qu'un individu dérive de plusieurs populations d'origine en fragmentant son génotype en autant de portions

nécessaires. La valeur du coefficient de chacune de ces portions reflète l'appartenance de l'individu à ces différentes populations parentales (Chen *et al.* 2007, François & Durand 2010).

Cependant, dans les deux types de modèles, on ne peut pas savoir si un individu non-assigné clairement à un groupe génétique peut être réellement qualifié d'hybride. Même en utilisant la méthode avec ancestralité mixte, un individu dont le génome serait assigné pour moitié à une population parentale et pour moitié à une autre reste difficilement interprétable. S'agit-il bien d'un hybride ou d'un individu dont l'origine n'est pas définissable ?

C'est pourquoi un programme a été spécialement développé pour la détection des individus hybrides, il s'agit de NEWHYBRIDS (Anderson & Thompson 2002). Ce programme, qui ne nécessite pas de connaître les fréquences alléliques des individus parentaux, permet de distinguer les individus parentaux, les hybrides de première ou de seconde générations et les individus issus de rétrocroisements (« backcross ») (Anderson & Thompson 2002). Cependant la détermination des individus de seconde génération ou ceux issus de rétrocroisements nécessite l'utilisation d'un très grand nombre de marqueurs, environ 50, ou une très forte différenciation entre les populations parentales (Vähä & Primmer 2006). Avec un nombre plus réduit de marqueurs, on se limite à la détection des individus appartenant aux populations parentales et aux hybrides de première génération.

3.2.2. Ajout de populations de référence

Parfois les programmes permettent d'ajouter des informations sur l'origine de certains individus, c'est le cas de GENECLASS2 (Piry *et al.* 2004) ainsi qu'une des options de STRUCTURE. Dans ce cas, certains individus dont la population d'origine est connue forment des populations de référence. Les autres individus de l'échantillon sont répartis entre les populations de référence. De plus, dans le cas où les groupes génétiques sont très faiblement différenciés, une nouvelle option de STRUCTURE permet d'ajouter l'information d'un lieu de collecte commun ou non pour les échantillons afin d'améliorer la détection d'une structuration génétique (Hubisz *et al.* 2009).

3.2.3. Ajout de la dimension spatiale

Une autre spécificité implémentée dans certains programmes d'assignation est la prise en compte de l'information sur l'origine géographique des individus échantillonnés. Ces programmes d'assignations spatiales (GENELAND, TESS...) ont contribué à l'essor de la génétique du paysage (Manel *et al.* 2003 ; Guillot *et al.* 2005a, 2009 ; Storfer *et al.* 2010).

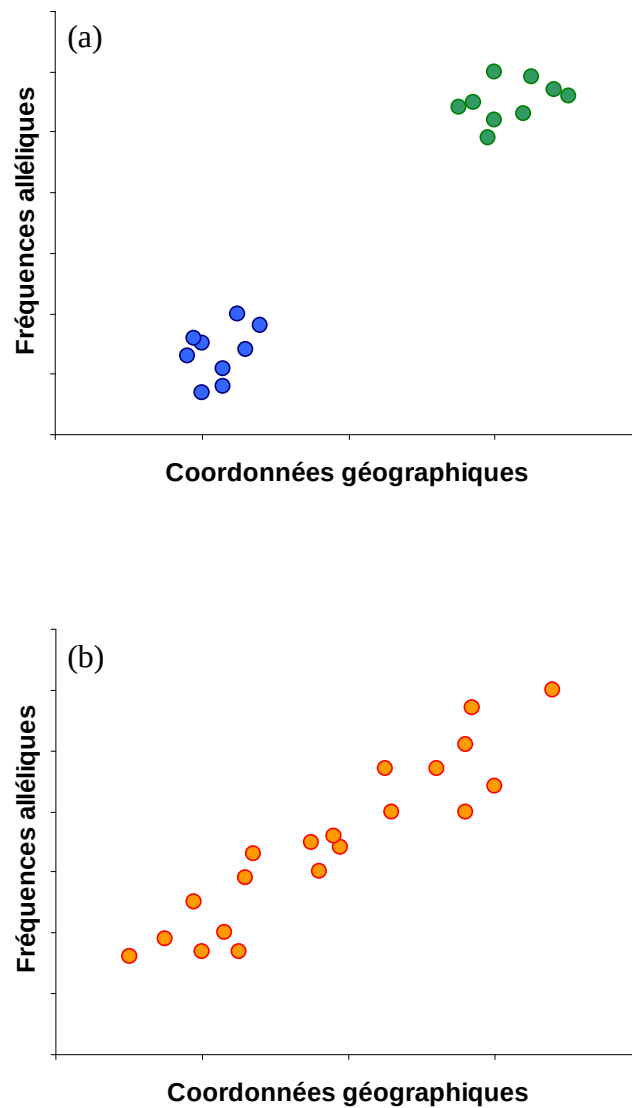


Figure 1 : Exemple théorique des patrons d'évolution des fréquences alléliques, (a) cas où deux populations se différencient, (b) cas où la population évolue sous le modèle de l'isolement par la distance (D'après Guillot *et al.* 2009).

Cette discipline combinant l'écologie du paysage et la génétique des populations a pour but de détecter l'influence relative des caractéristiques du paysage ou de l'environnement sur les discontinuités génétiques ou au contraire sur les connectivités géniques, reflétant respectivement des barrières ou des corridors à la dispersion (Storfer *et al.* 2007).

La structuration par le paysage peut être plus ou moins marquée (François & Durand 2010). Dans le cas le plus extrême, les groupes génétiques peuvent être séparés physiquement par un obstacle ou une barrière, les individus seront alors bien assignés à leurs populations d'origine. Lorsque les aires de répartition des populations d'origine présentent une zone de contact avec un gradient d'appartenance des individus d'une population à une autre, on parle de clines. Dans ce cas on trouvera des hybrides entre les deux populations dans la zone de contact. Le troisième type de structure spatiale est l'isolement par la distance (Wright 1943). L'isolement par la distance se base sur le fait que la reproduction s'effectue entre des individus géographiquement proches et que les distances de dispersion des individus sont toujours inférieures à l'aire de répartition totale de cette espèce (Wright 1943 ; Rousset 1997, 2000).

Dans le cas de populations présentant un patron d'isolement par la distance, les fréquences alléliques évoluent de manière graduelle (Fig. 1) (Guillot & Santos 2009 ; Guillot *et al.* 2009 ; François & Durand 2010). Ainsi l'utilisation d'algorithmes d'assignation qui regroupent « de force » les individus va générer arbitrairement des groupes génétiques qui n'auront pas de réalité biologique.

3.2.4. Evaluer le nombre de populations

L'évaluation du nombre de groupes génétiques constitue une étape très importante dans l'utilisation de ces méthodes, car par définition le nombre de groupes n'est pas connu initialement. Certains programmes donnent une estimation directe du nombre de populations (GENELAND). En revanche ce n'est pas le cas pour d'autres programmes tels que STRUCTURE ou TESS.

Lorsqu'on utilise STRUCTURE, à la fin de chaque simulation on obtient une valeur de la probabilité postérieure des données sachant le nombre de groupes génétiques testés (notée $\ln P(D)$). En général, cette valeur augmente en fonction du nombre (noté K) de groupes génétiques testés. La valeur de K la plus probable est celle pour laquelle on constate un décrochement dans l'augmentation des valeurs de $\ln P(D)$ (Evanno *et al.* 2005). Cependant l'évaluation du nombre de groupes génétiques de cette manière est parfois difficile car les valeurs de $\ln P(D)$ continuent souvent d'augmenter avec la valeur de K . C'est pourquoi

Evanno *et al.* (2005) ont développé une méthode, se basant sur les différentes valeurs de $\ln P(D)$ selon les valeurs de K , mais en les transformant légèrement. Cette méthode rend la détection du nombre de groupes plus aisée et plus juste. La procédure pour l'évaluation du nombre de populations à l'aide du programme TESS est relativement similaire par rapport à celle de STRUCTURE mais moins ambiguë (Chen *et al.* 2007). Pour chaque simulation on obtient une statistique, le DIC (Deviance Information Criterion). Cet indice diminue avec K , et comme pour l'indice $\ln P(D)$ de STRUCTURE la valeur la plus probable de K est celle pour laquelle on observe un décrochement dans la diminution des valeurs de DIC (Chen *et al.* 2007). Il reste néanmoins à noter que le nombre de populations détectées dépend de la taille de l'échantillon et du nombre de marqueurs; et il peut aussi varier selon le programme utilisé (François & Durand 2010).

3.3. Méthodes alternatives

Bien que les programmes d'assignation soient largement utilisés, d'autres méthodes existent pour évaluer la structure génétique d'un échantillon d'individus. Citons principalement les analyses discriminantes dont les analyses en composantes principales (ACP). Ces analyses ont l'avantage d'être plus visuelles, moins coûteuses en temps de calcul, même dans le cas d'analyses de jeux de données de très grande taille (Patterson *et al.* 2007). Autre avantage, les groupes génétiques déterminés peuvent être en déséquilibre d'Hardy-Weinberg (HW) ou présenter des déséquilibres de liaison entre loci (Jombart *et al.* 2010).

3.4. Programmes utilisés

Il est recommandé par plusieurs auteurs de tester différents programmes au cours d'une même analyse afin de tester la robustesse des résultats obtenus quelles que soient les options utilisées, par exemple avec ou sans prendre en compte l'information spatiale (Guillot *et al.* 2009, François & Durand 2010). Ainsi de nombreuses études ont comparé différents programmes d'assignation (Latch *et al.* 2006 ; Chen *et al.* 2007 ; Rowe & Beebee 2007 ; Frantz *et al.* 2009) dans le but de mettre en évidence leurs points de divergence ou de convergence. Une seconde manière de vérifier la robustesse des résultats est d'associer les tests d'assignations avec d'autres méthodes telles que les analyses discriminantes ou les autocorrélations spatiales (François & Durand 2010). Enfin, pour arbitrer entre des nombres de populations différents, il ne faut pas perdre de vue la cohérence des populations

déterminées avec d'autres informations issues de la biologie des organismes (Guillot *et al.* 2009).

Dans ce travail, les différentes analyses génétiques ont été réalisées à partir d'un nouveau panel de 25 marqueurs microsatellites issus en grande majorité des travaux décrits en section 1.2 et tous choisis pour leur fort taux de polymorphisme. D'une manière générale, des tests d'assignation ont été effectués à l'aide de différents programmes choisis selon les besoins spécifiques des études. Pour chaque étude, nous avons utilisé systématiquement plusieurs approches. Nous avons utilisé en premier lieu des programmes sans *a priori*, tels que STRUCTURE, GENELAND et TESS (en fournissant pour ces deux derniers des coordonnées spatiales fictives pour ne pas prendre en compte l'information spatiale réelle). D'autres tests, utilisant les coordonnées géographiques réelles, ont également été réalisés. Nous avons aussi utilisé NEWHYBRIDS car sa spécificité nous a permis de tester l'origine hybride de certains individus présentant un profil d'appartenance mixte (à deux populations parentales). Toutes ces différentes analyses seront détaillées plus spécifiquement au cours des chapitres suivants.

Résultats

Chapitre 1

Influence de la présence de l'hôte
cultivé sur la structure génétique de
Melampsora larici-populina sur son
hôte sauvage en France

Chapitre 1

Influence de la présence de l'hôte cultivé sur la structure génétique de *Melampsora larici-populina* en France sur son hôte sauvage.

Dans ce premier chapitre, nous nous sommes intéressés à la composition génétique des individus de *M. larici-populina* collectés à travers la France. Notre but était de savoir si une structuration génétique des individus de *M. larici-populina* existe à cette échelle et dans quelle mesure cette structuration résulte de l'influence de la populiculture. Cette étude s'intègre dans la thématique des interactions entre champignons pathogènes vivant au sein de compartiments sauvage ou cultivé. En France, les deux types d'hôtes (peupliers sauvages et cultivés) sont disponibles pour l'agent pathogène. Toutefois, les peupliers cultivés sont souvent porteurs de gènes de résistance qualitative, ce qui exerce une pression de sélection particulière sur les populations de *M. larici-populina*. Seuls les individus porteurs de la virulence correspondante pourront infecter ces peupliers, car le spectre de virulences que porte l'agent pathogène conditionne la gamme d'hôtes disponibles pour l'infection.

Nous avons testé l'effet du déploiement massif d'un cultivar en France, le cultivar 'Beaupré' porteur de la résistance R7, sur la structure génétique et phénotypique des individus de *M. larici-populina*. Contrairement à des études précédentes (Pinon & Frey 1997, Gérard *et al.* 2006, Barrès *et al.* 2008), nous n'avons testé que la présence de la virulence correspondante (Vir7) et non pas l'ensemble du spectre de virulences. Afin que les résultats ne soient pas biaisés par le type d'hôte sur lequel les individus ont été échantillonnés, les collectes de *M. larici-populina* ont été restreintes à l'hôte sauvage *P. nigra*, qui n'exerce aucune pression de sélection qualitative sur le champignon (Légionnet *et al.* 1999, Pinon & Frey 2005). De plus, grâce à la reproduction sexuée sur mélèze, les individus de *M. larici-populina* collectés dans différentes régions de France reflètent l'ensemble de la composition génétique et phénotypique locale.

Au cours de cette étude, nous avons utilisé les marqueurs microsatellites nouvellement développés pour effectuer les analyses génétiques en utilisant des méthodes d'assignation Bayésiennes permettant (i) de tester l'existence d'une structuration génétique en identifiant les groupes génétiques homogènes au sein de l'échantillonnage et (ii) en corollaire, de définir l'appartenance des individus à ces groupes génétiques. Ce travail se présente sous la forme d'un article accepté dans la revue *Molecular Ecology*.

ARTICLE 2 : THE GENETIC STRUCTURE OF THE PLANT PATHOGENIC FUNGUS *MELAMPSORA LARICI-POPULINA* ON ITS WILD HOST IS EXTENSIVELY IMPACTED BY HOST DOMESTICATION

Authors: C. Xhaard¹, B. Fabre¹, A. Andrieux¹, P. Gladieux², B. Barrès³, P. Frey^{1*} and F. Halkett^{1*}

* these authors co-directed this work.

Affiliations:

1- INRA, Nancy-Université, UMR 1136 Interactions Arbre-Microorganismes, F-54280 Champenoux, France.

2- Laboratoire Ecologie, Systématique et Evolution, Univ. Paris Sud, UMR8079, Orsay, France; CNRS, Orsay, France; AgroParisTech, Orsay, France.

3- INRA, UMR Biodiversité Gènes et Communautés, Domaine de Pierroton, F-33662 Cestas, France

Corresponding author:

Fabien Halkett

INRA-Nancy Université, UMR 1136 Interactions Arbres-Microorganismes, F-54280 Champenoux, France

Phone: 33 (0) 3 83 39 40 55

Email: halkett@nancy.inra.fr

Running title: Host domestication impact on fungal pathogens.

Keywords: Selection, demographic history, disease spread, gene flow, clustering analysis, rust fungus.

Abstract:

Wild and cultivated plants represent very different habitats for pathogens, especially when cultivated plants bear qualitative resistance genes. Here, we investigated to what extent the population genetic structure of a plant pathogenic fungus collected on its wild host can be impacted by the deployment of resistant cultivars. We studied one of the main poplar diseases, poplar rust, caused by the fungus *Melampsora larici-populina*. A thousand fifty individuals sampled from several locations in France were phenotyped for their virulence profile (ability to infect or not the most deployed resistant cultivar ‘Beaupré’), and a subset of these was genotyped using 25 microsatellite markers. Bayesian assignment tests on genetic data clustered the 476 genotyped individuals into three genetic groups. Group 1 gathered most virulent individuals and displayed evidence for selection and drastic demographic changes resulting from breakdown of the poplar cultivar ‘Beaupré’. Group 2 comprised individuals corresponding to ancestral populations of *M. larici-populina* naturally occurring in the native range. Group 3 displayed the hallmarks of strict asexual reproduction, which has never previously been demonstrated in this species. We discuss how poplar cultivation has influenced the spatial and genetic structure of this plant pathogenic fungus, and has led to the spread of virulence alleles (gene swamping) in *M. larici-populina* populations evolving on the wild host.

Introduction

Human activities have become a major source of environmental and ecological changes. A particular example of this is the domestication of animals and plants, one of the most important technological innovations in human history (Purugganan & Fuller 2009). The domestication process led to marked shifts in the evolutionary dynamics of host-pathogen interactions in both animals (Daszak *et al.* 2000; Diamond 2002) and plants (Stukenbrock *et al.* 2007), and resulted in tremendous changes in pathogen population structure (Lebarbenchon *et al.* 2008; Stukenbrock & McDonald 2008; Giraud *et al.* 2010).

Some of the most devastating epidemics on agricultural crops were caused by pathogens that shifted from a (more or less related) wild host to a domesticated one (Anderson *et al.* 2004; Parker & Gilbert 2004; Desprez-Loustau *et al.* 2007). Examples of fungal (*sensu lato*, *i.e.* including oomycetes) plant diseases caused by host-shifts include rice blast (*Magnaporthe oryzae*; Couch *et al.* 2005), potato late blight (*Phytophthora infestans*; Gómez-Alpizar *et al.* 2007), wheat leaf blotch (*Mycosphaerella graminicola*; Stukenbrock *et al.* 2007) and leaf scald of barley (*Rhynchosporium commune*; Zaffarano *et al.* 2008). Alternatively, some pathogens appear to have followed their host along the domestication course through a process called host-tracking (e.g. *Ustilago maydis* on maize: Munkacsı *et al.* 2008, *Venturia inaequalis* on apple tree: Gladieux *et al.* 2008, 2010a). A drastic reduction in gene flow accompanying the host-shift and subsequent specialisation on the domesticated plant can lead to the rapid formation of new pathogen species (Giraud *et al.* 2010). Even in the case of host-tracking, where wild and cultivated hosts are closely related, a strong genetic differentiation between populations from wild and cultivated hosts can be promoted by environmental differences existing between natural and domesticated habitats (Lebarbenchon *et al.* 2008; Stukenbrock & McDonald 2008).

In this context of pathogen evolution during domestication of cultivated plants, numerous studies have aimed to retrace the domestication routes of crop pathogens (Couch *et al.* 2005; Munkacsı *et al.* 2008; Zaffarano *et al.* 2008). However, only few studies interested in host-tracking pathogens have focused on gene flows occurring between pathogens from wild and cultivated plants (but see Kiros-Meles *et al.* 2011), certainly because of the difficulty to sample on wild host plants (Gladieux *et al.* 2008). Wild host plants can play the role of a pathogen reservoir (Wisler & Norris 2005; Lê Van *et al.* 2011), especially when both cultivated and wild hosts are sympatric. The wild hosts can be the source of new virulences

that can cause drastic epidemics on cultivated plants (Couch *et al.* 2005; Gomez-Alpizar *et al.* 2007; Zaffarano *et al.* 2008) and also facilitate connectivity between crops. Conversely, pathogens from cultivated plants can spread onto wild plants and compete with pathogens from wild hosts, which can lead to extensive gene flow, including pathogenicity-related genes, and in turn an increase of the damages caused by pathogens on the wild plants (Rizzo *et al.* 2005).

Agricultural landscapes, especially since the development of intensive agricultural practices, offer new and more favourable environments for the development of plant pathogens (Yarwood 1970; Hansen 1987). Contrary to wild host populations, which are generally characterized by low plant density and high genetic diversity (Gilbert 2002; Burdon 1993; Burdon & Thrall 2008; Stukenbrock *et al.* 2007), human-managed agroecosystems are characterized by high plant density and a reduced genetic diversity, which enhances pathogen multiplication by favouring transmission between infected and uninfected plants (Burdon & Chilvers 1982; Stukenbrock & McDonald 2008). Therefore, plant-pathogen interactions in the wild have often been depicted as benign associations in which serious epidemics rarely occur, whereas differences in the physical properties and biology of human-managed environments can lead to devastating epidemics in agricultural landscapes (Burdon 1993). Together with adaptation to host plants, differences in population dynamics between wild and domesticated ecosystems can not only impact pathogen population structure but also leave specific genetic signatures resulting from the demographic fluctuations; as was observed in meta-population systems with different host life spans in the anther smut fungus *Microbotryum violaceum* (Bucheli *et al.* 2001).

Poplar rust is a suitable model with which to study the interactions between pathogens evolving on wild and cultivated hosts, in particular because wild and cultivated poplars grow in sympatry in most parts of the Northern Hemisphere. The sympatry of wild and cultivated hosts enables both historical and contemporary gene flow to be assessed in the absence of confounding effect of large geographical distances on pathogen dispersal. Assuming recurrent gene flow exists between wild and cultivated hosts it is thus possible to assess the consequences of the presence of cultivated hosts on the genetic structure of pathogen populations sampled on the wild host. Wild and cultivated poplars differ in their ecological characteristics, and thus in the nature of the niche they provide to the pathogen. On the one hand, the native Eurasian black poplar, *Populus nigra*, is a pioneer tree species of riparian forests that is distributed widely across Europe and Asia (Imbert & Lefevre, 2003). Qualitative resistance to the Eurasian poplar rust fungus *Melampsora larici-populina* has not

been detected so far in *P. nigra* (Pinon & Frey 2005), and rust incidence on wild poplar stands is generally low (Gérard *et al.* 2006). On the other hand, cultivated poplar stands are particularly exposed to the attacks of many pathogens because of their intensive monoclonal cultivation over several decades (Gérard *et al.* 2006). In Europe, rust caused by *M. larici-populina* is the main disease on cultivated poplars and causes severe damage and economic losses in poplar cultivation (Frey *et al.* 2005). Most of the rust-resistant poplar cultivars released so far by European poplar breeders have been shown to carry qualitative resistance genes, all of which have been overcome by *M. larici-populina* (Pinon & Frey 2005). The most damaging resistance breakdown occurred in the hybrid cultivar *Populus x interamericana* ‘Beaupré’, which carries the R7 resistance gene. This cultivar showed high yield performance and was grown successfully between 1980 and 1994. During this period, ‘Beaupré’ was the poplar cultivar planted most in north-western Europe, accounting for 80% of new plantations in northern France (Pinon *et al.* 1998). In 1994, breakdown of the R7 resistance was detected in Belgium and northern France, and virulent (Vir7) *M. larici-populina* individuals spread all over Western Europe in less than five years, causing very destructive epidemics (Pinon *et al.* 1998; Pinon & Frey 2005). Contrary to a meta-population system (or annual crop fields) which displays frequent extinction/recolonization by pathogens and rapid turn over of pathogen virulence (McDonald 2010), a perennial plant like a poplar tree exhibits a much higher longevity. Even if plantation of ‘Beaupré’ decreased after 1997 and totally ceased by 2005, this cultivar still predominates in poplar cultivated landscape in Northern France; thus exerting a long lasting and constant selection pressure on *M. larici-populina* populations through the deployment of R7 resistance.

The rust fungus *M. larici-populina* is an obligate biotrophic pathogen belonging to the order Pucciniales (Basidiomycetes). It is a macrocyclic heteroecious fungus that produces five spore stages during its life cycle and which infects two host species. The sexual stage occurs on larch needles (*Larix* sp.) in spring—the impact on larch being insignificant. Then, *M. larici-populina* infects poplar leaves, on which many rounds of asexual multiplication occur between spring and autumn until leaf fall. The fungus then develops an overwintering form on senescent poplar leaves. In spring, the life cycle restarts with the infection of newly formed larch needles. The sexual stage on larch ensures a huge reshuffling of genetic material among individuals each year. Although native to the mountains of Central Europe (Alps, Carpathians, Sudetes, Tatras) and lowlands in Northern Poland, the European larch (*Larix decidua*) has been widely planted in lowlands throughout Europe since the 18th century for timber and ornamental purposes. Given the overall distribution of larches, high levels of

genotypic diversity are observed in all locations where *M. larici-populina* is sampled, even when individuals were collected on poplar at the end of the epidemic phase (Gérard *et al.* 2006; Pei *et al.* 2007; Barrès *et al.* 2008), in accordance with a predominantly sexual reproductive mode (Halkett *et al.* 2005a). Similar results were obtained with the North-American poplar rust fungus *Melampsora medusae* f. sp. *deltoideae*, when individuals were collected in poplar-larch sympatry area (Bourassa *et al.* 2007). In addition, host alternation and wind-dispersed spores foster gene flow across host plants (including between wild and cultivated poplars) and locations. Previous population genetic studies on *M. larici-populina* revealed a very low level of genetic differentiation between locations, and demonstrated the lack of a relationship between geographical and genetic distances, both in France (Gérard *et al.* 2006; Barrès *et al.* 2008) and in the UK (Pei *et al.* 2007). Nonetheless, at a continental scale, isolation by distance was tentatively observed (Barrès *et al.* 2008), confirming a gradual dispersion of spores between locations. Even at this large spatial scale, F_{ST} values remained weak, indicating that, even if dispersal of the pathogen is gradual, there is substantial gene mixing over the European continent. Concerning phenotypic characteristics, no relationship between virulence profile and genotype, or between virulence profile and geographical location, has been evidenced (Gérard *et al.* 2006; Barrès *et al.* 2008). The lack of relationship between phenotype and genotype has been proposed to result from the extensive genetic reshuffling imposed by the sexual stage in larch.

Here we aimed to benefit from a very peculiar ecological situation (the large and long lasting selection pressure exerted by R7 cultivated poplars on the rust fungus *M. larici-populina*) to investigate the putative consequences of the impact of host domestication on population genetic structure of a plant pathogen on its wild host. We sampled more than twenty sites across France, where large differences in the acreage of R7 cultivated poplars exist between northern and southern locations. The individuals collected were genotyped using 25 polymorphic microsatellite markers and phenotyped for their Vir7 profile. The use of both putatively neutral and selected markers provided an opportunity to examine the relationship between population genetic structure and a key biological property (Barrett *et al.* 2008b); here the ability of *M. larici-populina* individuals to grow or not on R7 cultivated poplars. We first employed Bayesian clustering analyses to infer population structure in a set of *M. larici-populina* individuals collected on the wild host *P. nigra*. We then examined (i) whether the genetic differentiation matches the ability to grow on cultivated plants; (ii) whether identified genetic groups of the pathogen display a particular pattern of spatial distribution and how this relates to differences in cultivated host densities and finally (iii)

Table 1: Characteristics of sampling sites. The number in brackets after the number of genotyped individuals indicated the number of individuals with more than 6 loci that failed to amplify; *G/N*, Genotypic richness

Site No	Location	Latitude	Longitude	No of individuals phenotyped	No of individuals genotyped	G/N
1	Amance ^a	48°45' N	6°20' E	80	15 (1)	1.00
2	Angers	47°23' N	0°39' W	26	16 (0)	0.75
3	Aoste (Italy)	45°46' N	7°18' E	16	14 (0)	0.93
4	Changy	46°09' N	3°53' E	46	14 (2)	0.93
5	Charrey-sur-Saône ^a	47°05' N	5°09' E	30	16 (0)	1.00
6	Croix	50°41' N	3°08' E	21	11 (5)	1.00
7	Guémené-Penfao ^a	47°38' N	1°54' W	61	15 (1)	0.93
8	Mirabeau	43°41' N	5°04' E	108	83 (13)	0.81
9	Montpellier	43°41' N	3°52' E	30	24 (2)	0.63
10	Nancy	48°43' N	6°10' E	48	16 (0)	1.00
11	Nievroz	45°49' N	5°04' E	88	16 (0)	1.00
12	Prelles	44°51' N	6°35' E	99	90 (6)	1.00
13	Ramières 1	44°45' N	4°53' E	39	16 (0)	0.31
14	Ramières 2	44°44' N	4°58' E	52	15 (1)	0.80
15	Rennes	48°06' N	1°48' W	34	16 (0)	1.00
16	Roubaix	50°41' N	3°10' E	12	12 (0)	0.92
17	Seilh	43°42' N	1°22' E	16	11 (5)	0.82
18	Saint-Ay	47°52' N	1°46' E	108	15 (1)	0.33
19	Saint-Laurent-du-Var	43°40' N	7°12' E	15	14 (1)	0.21
20	Ungersheim	47°51' N	7°17' E	26	16 (0)	0.88
21	Veurey-Voroise	45°16' N	5°37' E	80	14 (2)	1.00
22	Villeneuve d'Ornon	44°47' N	0°34' W	15	15 (0)	0.47
23	Orléans ^a	47°54' N	1°54' E	120	-	-

^a Experimental poplar nursery.

whether genetic groups of the pathogen display differences in patterns of genetic variability consistent with a history of strong selection by host resistance and subsequent changes in demography.

Materials and Methods

* Sampling strategy

Poplar leaves infected with *M. larici-populina* were collected on *P. nigra* in 23 locations across France (and Aoste, Italy) between August and November 2009 (Table 1). Sampling was performed on wild riparian poplar stands, except for four locations (Amance, Charrey-sur-Saône, Guémené-Penfao and Orléans) where wild *P. nigra* trees were grown in experimental poplar nurseries. These poplar nurseries also contained many different poplar cultivars harboring eight distinct qualitative resistances. As no qualitative resistance to *M. larici-populina* has been reported so far in *P. nigra* (Pinon & Frey 2005), collecting on *P. nigra* – whatever the location type – avoided any specific selection pressure resulting from the qualitative resistance exerted by some cultivated poplars (Gérard *et al.* 2006). Hence, because of the extensive gene flow in this species, each collection site was assumed to reflect the local composition of *M. larici-populina* genotypes. At each location, 15 to 150 rust-infected leaves were collected. The highest sampling effort was conducted at Prelles (Table 1), located in the original distribution area of *M. larici-populina*, where larch and wild poplar are sympatric. In the following, samples from this location are referred to as the reference population.

* Individual isolation

One single uredinium (sporulating lesion of *M. larici-populina* producing urediniospores) per leaf was selected randomly in order to avoid excessive clonal replicates, and grown on fresh leaf discs of *P. x euramericana* ‘Robusta’ as described by Barrès *et al.* (2008). Mono-uredinial isolation ensures that a single genotype per sampled leaf is isolated and multiplied, so that both virulence phenotype assessment and molecular analysis can be performed on the same individual.

* Virulence phenotype assessment

All the *M. larici-populina* individuals (Table 1) were inoculated on two poplar

cultivars, one susceptible (*P. x euramericana* ‘Robusta’) and the second carrying the R7 rust resistance gene (*P. x interamericana* ‘Beaupré’), as described by Barrès *et al.* (2008). Phenotype of each individual was denoted Vir7 if both resistant and susceptible leaf discs were infected, and Avr7 if only the susceptible (control) leaf disc was infected. For each location, we calculated the proportion of Vir7 individuals. After virulence identification, leaf discs of ‘Robusta’ were frozen and stored at -20°C until DNA extraction.

* Microsatellite analysis

A subsample of 12 to 96 individuals per sampling site was defined randomly for microsatellite analysis (Table 1). Only 22 (out of the 23) locations were considered for genetic analyses. DNA was extracted using the BioSprint 96 DNA plant kit used in combination with the BioSprint 96 automated workstation (Qiagen) following the BS96-DNA-plant protocol. Microsatellite analyses were performed at 25 polymorphic loci: MLP12 (Barrès *et al.* 2006); MLP49, MLP50, MLP54, MLP55, MLP56, MLP57, MLP58, MLP59, MLP66, MLP68, MLP71, MLP73, MLP77, MLP82, MLP83, MLP87, MLP91, MLP92, MLP93, MLP94, MLP95, MLP96, MLP97 (Xhaard *et al.* 2009) and MLP100 (this study). Microsatellite loci were amplified in a GeneAmp PCR system 9700 thermocycler (Applied Biosystems) under the following conditions: an initial denaturing step of 5 min at 95°C; 32 cycles including 30 s of denaturation at 95°C, 90 s annealing at 60°C and 60 s extension at 72°C; and a final 30 min extension step at 60°C. PCR was carried out in a 10 µL final reaction volume containing 5 µL of Qiagen® Multiplex PCR kit, 2 µL of DNA, 0.2 µM of each reverse and forward primers. Forward primers were labelled with a fluorescent tag (FAM, NED, VIC or PET). Three multiplex PCR were run, comprising 9, 7, and 9 loci, respectively (Table 2). PCR products from the three multiplex reactions (3 µL PCR1, 4 µL PCR2 and 5 µL PCR3) were pooled and loaded on a ABI 3730 Genetic Analyzer (Applied Biosystems). Fragments were sized with a LIZ-1200 size standard, and alleles were scored using GENEMAPPER 4.0 (Applied Biosystems). Individuals for which more than six loci failed to amplify were omitted from further analyses.

* Data analysis

Identification of repeated multilocus genotypes

When pooling all multilocus genotypes from all locations, we identified individuals carrying repeated multilocus genotypes. Given the life cycle of *M. larici-populina*, such

repeated multilocus genotypes can result from clonal multiplication or can simply occur by chance given the allelic frequency in the population (Halkett *et al.* 2005a). To distinguish true clonal copies from individuals carrying the same multilocus genotype by chance, we used MLGsim software (Stenberg *et al.* 2003) that computes the probability of observing a multilocus genotype n times in a population and performed Monte Carlo simulations (10 000 iterations) to determine the probability threshold below which each repeated multilocus genotype can be significantly considered as a true clone (significance level was set to 0.05).

In order to minimize departure from Hardy-Weinberg (HW) expectations and linkage disequilibrium (LD) on which several analyses are based, we considered only one copy of each multilocus genotype (found to be significantly over-represented) for all subsequent genetic analyses. An exception is the computation of the genetic characteristics of genetic groups, for which summary statistics were calculated both with and without clonal copies.

Clustering analysis

We investigated population subdivision using the Bayesian clustering method implemented in GENELAND (Guillot *et al.* 2005b; Guillot 2008). GENELAND offers two types of analyses, taking –or not– information on the geographical origin of individuals into account to assign individuals to genetic groups. Assignment tests were conducted using both methods, which can be more or less sensitive to the within-site genetic admixture. Another strength of GENELAND is that it infers the best number of genetic groups, K , during a single run using the specific Reverse Jump MCMC (Markov chain Monte Carlo) algorithm, which enables the Markov chain to explore the whole range of putative cluster numbers at each step of the simulation run. The number of genetic groups that best explain the data (i.e. how confidently individuals are sorted) is given by the value of K in which the Markov chain spent most iterations (Guillot *et al.* 2005a,b). K was allowed to vary from 1 to 10, and initialised at 8. Three independent runs were performed to check for convergence of Markov chains. This first batch of runs with varying K enabled the best putative number of genetic groups to be assessed. Once the most likely K value was determined, three additional runs were performed, with the corresponding value of K fixed, to better infer the assignment level of each individual to the K genetic groups (Guillot *et al.* 2005b). For the three runs with varying K , we used the uncorrelated allele frequency model, and for those with fixed K we used the correlated allele frequency model according to Guillot (2008) with default settings, except for the number of nuclei, which was increased to 500 (in the non-spatial mode), given the number of individuals. Run length was set to 550,000 iterations with a burning-in period of 50,000

iterations. The threshold of assignment coefficient (q) for individual membership to a genetic group was set to 0.8.

Population genetics analyses

We computed summary statistics of genetic variability within (and its apportionment among) groups identified using Bayesian clustering methods. Expected and observed heterozygosities were calculated with GENETIX (4.05) (Belkhir *et al.* 2004). We calculated allelic richness (A_r) using FSTAT version 2.9.3 (Goudet 1995) that implements a rarefaction procedure to account for differences in sample size. Differentiation coefficient between populations (F_{ST}) and deviation from random mating within populations (F_{IS}) were estimated according to Weir and Cockerham (1984) using GENETIX. Linkage disequilibrium between loci and deviations from HW equilibrium expectation (heterozygote excess and deficit) were calculated using GENEPOP on the web (<http://genepop.curtin.edu.au>) (Rousset 2008). In order to adjust the P value for multiple tests for linkage disequilibrium, we used the false discovery rate (FDR) procedure (Benjamini & Yekutieli 2001), which controls the proportion of significant results ('false discovery rate') instead of controlling the chance of making even a single error. The resulting adjusted P values are called Q values. This procedure is implemented in the R package QVALUE (Storey & Tibshirani 2003).

Analysis of isolation by distance

Isolation by distance was assessed within the most widespread group (Group 1) identified using Bayesian clustering methods (after removing locations with less than five assigned individuals) using the continuous method (Rousset 2000) of the 'isalde' procedure implemented in GENEPOP (Rousset 2008, option 'Analysis of isolation by distance between groups'). The significance of Mantel tests between pairwise genetic distances –analogue of $F_{ST}/(1-F_{ST})$ ratios– and geographical distances were performed using the ISOLDE program (10,000 permutations). Geographic distances were log-transformed because *M. larici-populina* individuals evolve in a two-dimensional space (Rousset 1997). For geographical distances we considered Euclidian distances between pairs of sampling sites.

Test for loci under selection

We tested for signs of positive selection on each locus, using the F_{ST} -outlier approach (Lewontin and Krakauer 1973), as implemented in FDIST2 (Beaumont & Nichols 1996). This method uses the expected distribution of F_{ST} vs. H_E to identify outlier loci that have

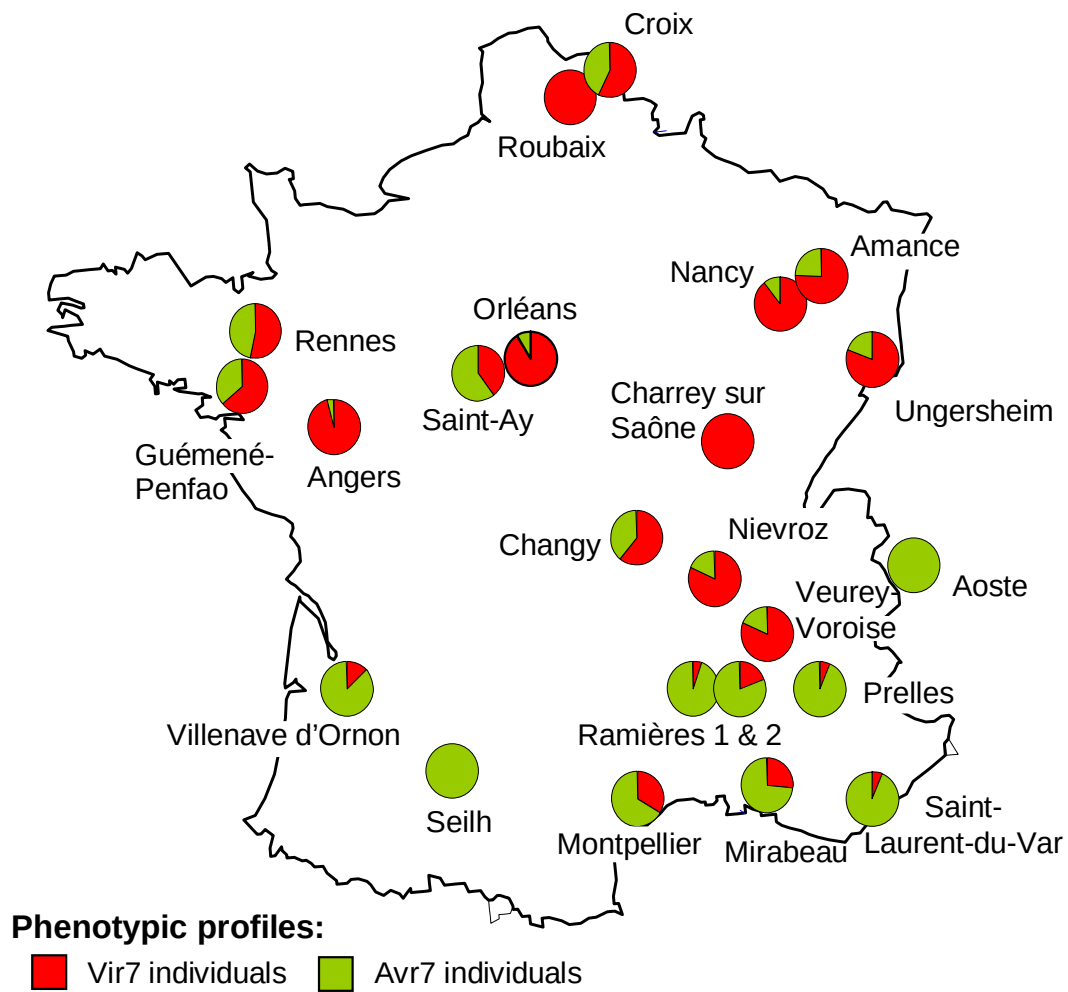


Figure 1: Geographical distribution of phenotypic profiles showing proportion of Avr7 (green) and Vir7 (red) individuals at each site.

excessively high (positive selection) or low (balancing selection) F_{ST} compared to neutral expectations. We ran 20,000 simulations to compute the 95% confidence intervals of the neutral value distribution of F_{ST} values under both the Stepwise Mutation Model (SMM) and the Infinite Allele Model (IAM).

Tests for demographic equilibrium

The demographic history of the groups of *M. larici-populina* identified using Bayesian clustering methods was assessed using two different methods designed to detect historical changes in population size, based on deviation from mutation-drift equilibrium. The first method relies on the fact that populations that have recently experienced a reduction in effective population size should exhibit larger values of gene diversity (i.e. heterozygosity level) than expected from the number of alleles at mutation-drift equilibrium (Cornuet & Luikart 1996). Conversely, expanding populations should exhibit deficits in heterozygosity levels. We used BOTTLENECK 1.2.02 (Piry *et al.* 1999) to assess the strength of the heterozygosity excess/deficit, i.e. expected heterozygosity (H_E) in the sense of Nei (1978), relative to the expected equilibrium heterozygosity (H_{eq}). The distribution of H_{eq} was obtained through simulation of the coalescent process given the observed allelic richness. The tests were performed under the stepwise mutation model (SMM), as well as under a two-phase model (TPM), allowing for 30% of multistep changes (IAM). We performed all analyses with 2,000 iterations of the coalescent process. Significance of the deviation from mutation-drift equilibrium was assessed using Wilcoxon signed rank tests, which are most powerful and robust when used with few polymorphic loci (Cornuet & Luikart 1996).

The second method is dedicated to loci, such as microsatellite loci, evolving under the SMM. Under this mutational model, a demographic scenario is likely to influence the distribution of allele sizes (Kimmel *et al.* 1998). In particular, for a population in expansion, the observed distribution of allele sizes should be less variable when compared to a population at demographic equilibrium. We used the imbalance index β_2 , which estimates the strength of the deviation between variance in allele size and expected heterozygosity, as it is the most powerful index to reveal population expansion (equation 8 in King *et al.* 2000). $\ln \beta$ is expected to be negative for a strictly growing population and positive for a population that has recently expanded following a bottleneck (Kimmel *et al.* 1998). 95% confidence intervals were obtained from the distributions of $\ln \beta_2$ estimates across loci, using the R package LOCFIT (Loader 1996).

Table 2: Characteristics of the microsatellite loci used. A , total number of alleles. H_E , expected heterozygosity calculated according to Nei (1978). H_O , observed heterozygosity. Probability test for Hardy-Weinberg (HW) equilibrium: significant levels of P value were corrected for multiple comparisons using false discovery rate (FDR) control (*, $Q < 0.05$; **, $Q < 0.01$; ***, $Q < 0.001$; ns, non significant). References: 1, Barrès *et al.* 2006; 2, Xhaard *et al.* 2009; 3, this study.

Locus	Repeat unit	Allele range	multiplex PCR No.	All individuals (n=476)				Prelles (n=90)				Reference
				A	H_E	H_O		A	H_E	H_O		
MLP12	complex (AAG)n	217-265	2	9	0.688	0.542	***	7	0.735	0.634	ns	1
MLP49	(GAT)18	344-407	3	19	0.732	0.731	***	13	0.766	0.762	ns	2
MLP50	(TCT)16	262-297	2	9	0.614	0.633	***	7	0.415	0.471	*	2
MLP54	(ATG)14	126-141	1	6	0.551	0.416	***	5	0.559	0.477	ns	2
MLP55	(ATC)15	133-167	1	11	0.521	0.517	***	6	0.614	0.682	***	2
MLP56	(AAC)7	269-306	2	11	0.625	0.594	***	9	0.661	0.640	ns	2
MLP57	(CAA)13	154-178	1	6	0.302	0.304	ns	6	0.402	0.438	ns	2
MLP58	(AAG)12	235-259	2	13	0.692	0.553	***	9	0.727	0.771	ns	2
MLP59	(ATC)13	305-335	2	12	0.760	0.625	***	11	0.808	0.440	ns	2
MLP66	(GAT)12	189-213	1	7	0.688	0.678	***	4	0.675	0.616	ns	2
MLP68	(TTGA)12	412-472	3	9	0.455	0.450	***	7	0.332	0.310	ns	2
MLP71	(AG)14	358-376	2	8	0.429	0.443	***	5	0.317	0.310	***	2
MLP73	(AC)14	419-449	3	17	0.842	0.829	***	11	0.822	0.738	ns	2
MLP77	(AT)10	431-437	3	6	0.374	0.336	***	4	0.506	0.451	ns	2
MLP82	(TAC)10	172-220	1	16	0.797	0.621	***	12	0.740	0.581	ns	2
MLP83	(ATG)10	145-163	1	7	0.476	0.453	ns	4	0.505	0.591	ns	2
MLP87	(TGT)8	338-353	3	6	0.465	0.370	***	3	0.464	0.427	ns	2
MLP91	(GTT)10	125-134	1	5	0.184	0.202	ns	3	0.055	0.056	ns	2
MLP92	(TTG)11	324-336	2	5	0.267	0.127	***	5	0.265	0.035	ns	2
MLP93	(GTT)11	125-149	1	7	0.572	0.629	***	6	0.601	0.636	***	2
MLP94	(TTC)10	452-477	3	10	0.666	0.592	***	6	0.686	0.605	ns	2
MLP95	(ACA)9	441-469	3	14	0.743	0.633	***	11	0.652	0.583	ns	2
MLP96	(TGG)9	403-413	3	5	0.558	0.478	***	3	0.500	0.434	ns	2
MLP97	(GATT)9	380-416	3	8	0.627	0.680	***	6	0.586	0.528	ns	2
MLP100	(ACA)18	178-231	1	14	0.706	0.509	***	12	0.597	0.455	ns	3

Results

* Virulence profiles

A total of 1,050 individuals of *M. larici-populina* from 23 locations across France (Table 1) were phenotyped for their ability to overcome (or not) the R7 qualitative resistance gene carried by *P. x interamericana* ‘Beaupré’. The frequency of individuals carrying the corresponding virulence (hereafter referred to as Vir7) ranged from 0 to 1 across sampling sites (Fig. 1), with a mean proportion of 0.53.

The spatial distribution of Vir7 rust individuals was non-random. Individuals from the northern half of France exhibited the highest proportion of Vir7 (up to 100% in Charrey-sur-Saône and Roubaix) (Fig. 1). The sole exception to this general trend was Saint-Ay, for which only 40% of individuals were Vir7. However, samples collected at a neighbouring site (Orléans nursery) were almost all Vir7 (91%). Unfortunately these individuals were not genotyped. A significant difference in the proportion of Vir7 between wild stands and poplar nurseries was found only in Saint-Ay vs. Orléans ($P < 0.05$); the other nurseries (Guéméné-Penfao, Amance, and Charrey-sur-Saône) displayed a Vir7 proportion similar to that found in wild poplar stands located nearby (Rennes, Nancy, and Nievroz, respectively) (Fisher exact test, $P = 0.83$).

* Summary of genetic and genotypic variability

Of the 514 individuals genotyped, 40 (8%) were excluded from further analyses because they failed to amplify or had too much missing data. These individuals were scattered across sampling sites (Table 1). The remaining 474 individuals exhibited only 2.9% missing data on average. Of the 474 individuals analyzed, 388 distinct multilocus genotypes were found. All the repeated multilocus genotypes were found to be significantly over-represented according to the analysis performed using MLGsim. Repetitions of a given multilocus genotype were thus considered to result from clonal reproduction and were further denoted clonal copies. Some clonal copies were sampled from different locations but most were restricted to a single site; only three repeated multilocus genotypes occurred in more than ten clonal copies. Subsequent analyses were performed considering a single copy of each repeated multilocus genotype.

All 25 microsatellite markers were polymorphic, with a number of alleles ranging from 5 to 19 (Table 2). The mean allele number across loci was 9.6 when pooling all samples

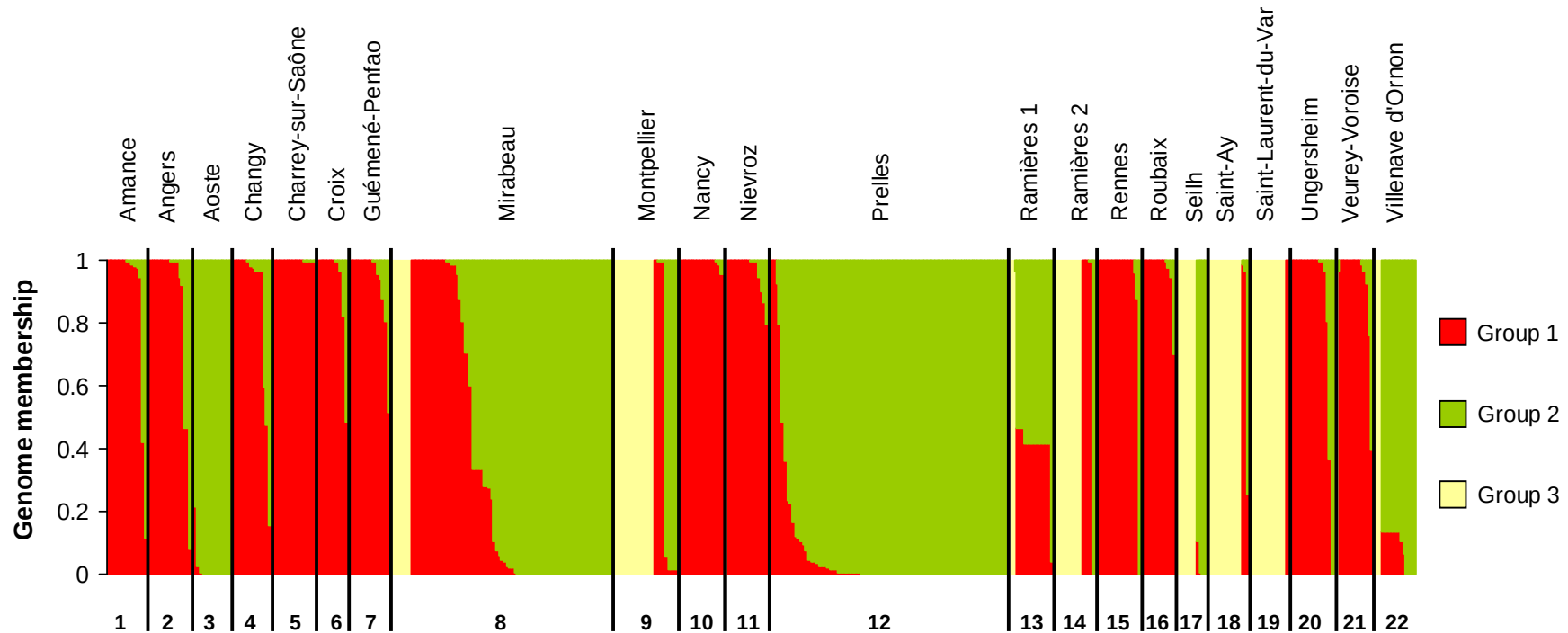


Figure 2: Assignment results of *Melampsora larici-populina* individuals. Each vertical line represents an individual whose genome is partitioned into K coloured segments (here $K = 3$ genetic groups). Segment lengths are proportional to the individual's mean membership coefficients in each genetic group (red: group 1; green: group 2; pale yellow: group 3). Vertical bars separate individuals sampled from different locations. Geographic locations are indicated underneath and numbered as in Table 1.

and 7.0 when considering only the reference population of Prelles, as in Xhaard *et al.* (2009). Gene diversity (i.e. expected heterozygosity, H_E) equalled 0.57 when considering all sampling locations and 0.56 for the Prelles sampling site. A highly significant HW deviation was found for almost all loci when considering all samples, which, together with the many pairs of loci in significant linkage disequilibrium observed (197/300) suggests some genetic structure (i.e. Wahlund effect). This trend was not observed when considering only the Prelles location, where nearly all loci did not depart significantly from HW equilibrium expectations.

* Genetic structure

GENELAND runs failed to consistently assess the number of genetic groups when taking into account the spatial information: two runs sorted the samples into two genetic groups (93% and 95% of iterations along the first and second Markov chain, respectively) and the third in three genetic groups (95% of iterations along the third Markov chain). Moreover, when fixing the value of K (three runs with $K = 2$ and three runs with $K = 3$), assignment results were incongruent across runs. Conversely, under the non-spatial mode, the three runs gave consistent results; with the three Markov chains converging rapidly on $K = 3$ (Markov chains stuck at $K = 3$ for the three runs). Setting a fixed value of $K = 3$ gave very consistent assignment results. In the following, we thus considered only those results obtained with the non-spatial mode. The membership coefficients q were averaged over the three runs with a fixed value of $K = 3$. Most individuals had high membership in a single genetic group (Fig. 2). Applying an assignment threshold of $q > 0.8$, more than 90% of multilocus genotypes (354 out of 390) could be assigned to a genetic group. Re-including multiple copies of the repeated multilocus genotypes, the three genetic groups comprised 192, 170 and 71 individuals, respectively. Some individuals from the same location belonged to different genetic groups (Fig. 2). This gathering of individuals with very different genetic profiles is likely to account for convergence failure of GENELAND when used with the spatial model, as under the spatial mode the same assignment coefficient (q) is set for all individuals sampled at a given location. The 43 non-assigned individuals were distributed evenly across the sampling sites, except in Ramières 1, where only three individuals could be confidently assigned to a genetic group (Fig. 2). Most individuals from this location derived from two clones displaying equal membership to genetic groups 1 and 2 (Fig. 2). The 43 non-assigned individuals were excluded from subsequent analyses.

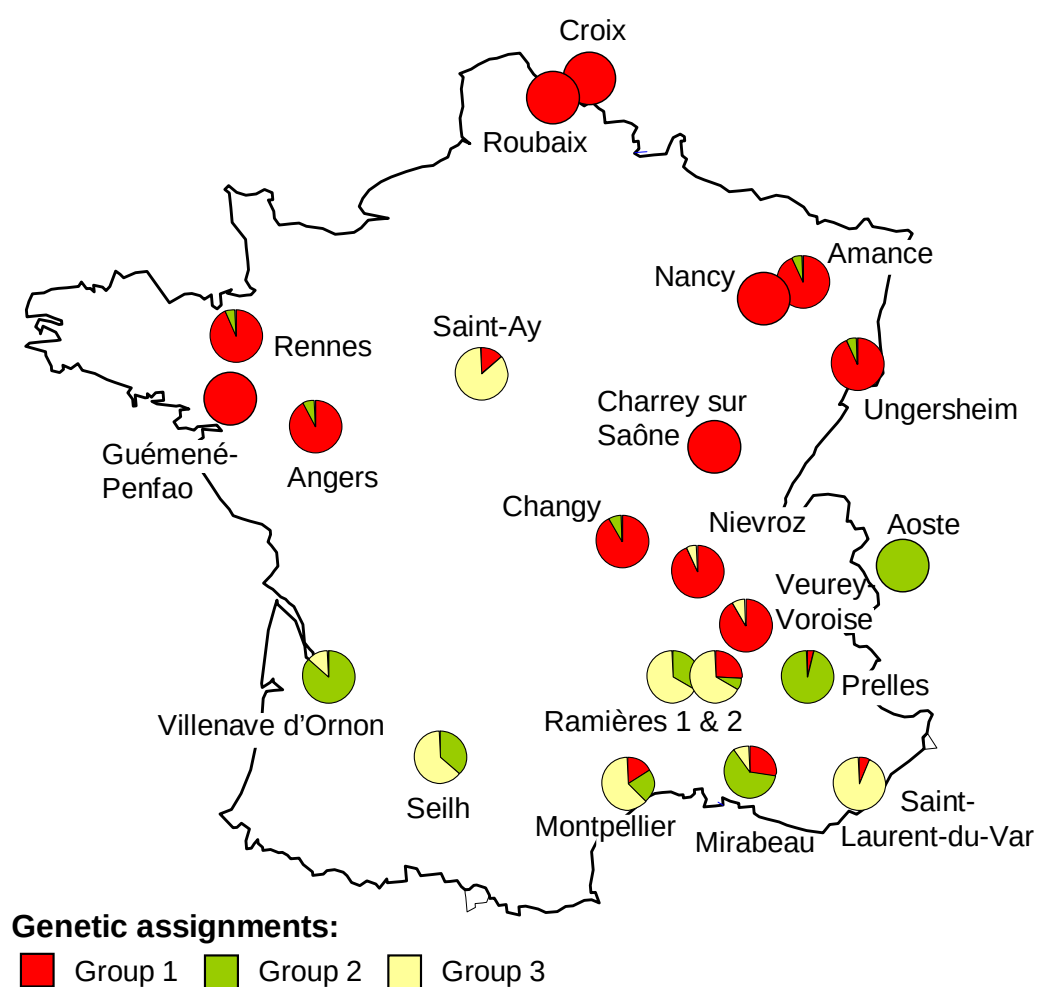


Figure 3: Geographical distribution of membership of the three genetic groups, showing proportion of individuals assigned to group 1 (red), group 2 (green), and group 3 (pale yellow) at each site.

Differentiation between the three groups was highly significant ($P < 0.001$), whichever comparison pair was considered, which is in accordance with the clear-cut clustering results. Genetic differentiation between groups 1 and 2 was lower ($F_{ST} = 0.04$) than comparisons including group 3 ($F_{ST} = 0.24$ and 0.20 with groups 1 and 2, respectively).

The spatial distribution of the three genetic groups is illustrated in Fig. 3. Individuals from group 1 occurred in nearly all locations (with the exception of Aoste, Ramières 1, Seilh and Villenave d'Ornon). However, group 1 occurred mostly in the northern part of France (with the exception of Saint-Ay, where this group accounted for only 13% of individuals). The second genetic group (group 2) was more abundant in the southern part of France, but again was non-evenly distributed and reached up to 100% of assigned individuals in Aoste, and nearly 97% in Prelles, both Alpine locations. Group 2 was present in all southern locations except Saint-Laurent-du-Var, possibly because of the limited sample size. In the northern part of France, individuals from group 2 were seldom sampled together with group 1. The third genetic group (group 3) was restricted to the southern part of France, with the notable exception of Saint-Ay, where group 3 accounted for 80% of the individuals (Fig. 2). All individuals belonging to this third genetic group were very well assigned ($q = 1$, Fig. 2) although they always coexisted with individuals belonging to the other two genetic groups (Fig. 3).

*** Genetic and phenotypic characteristics of the three genetic groups identified**

Group 1 exhibited the highest genotypic richness ($G/N = 0.96$ with N the number of individuals genotyped, and G the number of distinct multilocus genotypes) (Table 3). Only six repeated multilocus genotypes were found in this group, each occurring in two to three clonal copies, all sampled in the same location. Eighty one percent of the individuals from this group were Vir7, representing 86% of the total amount of Vir7 individuals. A highly significant correlation was found between individual membership coefficients to group 1 and their virulence profile (logistic regression, $P < 0.001$). Pooling all individuals regardless of their location of origin revealed a large heterozygote deficit (HW disequilibrium) and some pairs of loci exhibited significant linkage disequilibrium, suggesting some sub-structuring. Significant genetic differentiation was found among locations ($P < 0.001$, Table 3), but no isolation by distance was detected ($P = 0.159$).

Chapitre 1

Table 3: Characteristics of the three genetic groups. N , number of individuals; G/N , genotypic richness; $Vir7$, proportion of virulent individuals; H_E , unbiased estimate of gene diversity (Nei 1978); H_O , observed heterozygosity; probability test for heterozygote excess or deficit (***, $P < 0.001$; ns, non significant); A_r , allelic richness estimated for a minimal sample size of 71 (or 32 without clonal copies); F_{IS} values; LD, number of pairs of significant linkage disequilibrium after correction for multiple tests (for group 3 only 230 comparisons were calculated, because some loci were monomorphic in this group). The same calculations were conducted both with and without clonal copies. F -statistics (F_{ST} and F_{IS}) were calculated according to Weir and Cockerham (1984) taking into account the sampling location (within genetic groups). When available, standard errors across loci are given underneath in brackets.

	<i>N</i>	<i>G/N</i>	<i>Vir7</i>	<i>H</i> _E	<i>H</i> _O	Heterozygote		<i>A</i> _r	<i>F</i> _{IS}	LD	<i>F</i> -statistics	
						excess	deficit				<i>F</i> _{ST}	<i>F</i> _{IS}
Including clonal copies												
Group 1 ‘virulent-sexual’	192	0.96	0.81	0.536 (0.209)	0.501 (0.199)	ns	***	5.44 (3.17)	0.065 (0.100)	46/300	0.013 *** (0.004)	0.051 *** (0.016)
Group 2 ‘avirulent-sexual’	170	0.87	0.10	0.578 (0.203)	0.521 (0.181)	ns	***	5.69 (2.88)	0.100 (0.206)	244/300	0.093 *** (0.009)	0.033 ** (0.032)
Group 3 ‘avirulent-asexual’	71	0.45	0.13	0.378 (0.257)	0.561 (0.416)	***	ns	2.55 (1.16)	-0.486 (0.413)	150/230	0.116 *** (0.027)	-0.666 *** (0.038)
Without clonal copies												
Group 1 ‘virulent-sexual’	184	-	0.80	0.536 (0.198)	0.500 (0.194)	ns	***	5.48 (2.44)	0.068 (0.103)	5/300	0.009 *** (0.004)	0.057 *** (0.015)
Group 2 ‘avirulent-sexual’	138	-	0.09	0.578 (0.181)	0.512 (0.176)	ns	***	5.78 (2.36)	0.113 (0.194)	0/300	0.033 *** (0.004)	0.094 *** (0.030)
Group 3 ‘avirulent-asexual’	32	-	0.09	0.397 (0.248)	0.541 (0.388)	***	ns	2.75 (1.16)	-0.369 (0.381)	118/230	0.071 ** (0.034)	-0.477 *** (0.052)

Group 2 contained only 10% Vir7 individuals (Table 3). Differences with the genetic characteristics of group 1 included a lower genotypic richness but a larger genetic diversity (H_E) and allelic richness (A_r). Sub-structuring was also more pronounced, as shown by the higher number of loci in linkage disequilibrium and the higher F_{IS} and global F_{ST} values. The F_{IS} decreased markedly when taking into account sampling location (Table 3), although the heterozygote deficit was still significant. For both genetic groups, no pair of loci remained in linkage disequilibrium after removing clonal copies and taking into account sampling location. Of the seven repeated multilocus genotypes, two were more abundant (seven and ten clonal copies) and the five others were found at between two and four clonal copies. As in group 1, all the clonal copies of each repeated genotype were found in a single location.

Group 3 displayed very specific genetic characteristics, with lower allelic and genotypic richness, much more pairs of loci in linkage disequilibrium and larger heterozygote excess, which led to a very negative and significant value of F_{IS} ($P < 0.001$) (regardless of whether clonal copies were taken into account or not). The two most repeated genotypes were represented by 17 and 12 clonal copies. The other nine repeated multilocus genotypes were represented by two to four clonal copies. Only three repeated multilocus genotypes were found in different locations (including the most repeated genotype with a maximum resampling distance of 600 km). Group 3 was the only group to have clonal copies sampled at different locations. As with group 2, only a few individuals (9% of the multilocus genotypes) showed a Vir7 profile.

*** Evidence for positive selection in group 1**

Reduced genetic variability and a larger variance in heterozygosity across loci both suggested a possible history of strong natural selection in group 1 (Table 3). We further tested for selection by comparing the observed F_{ST} between groups 1 and 2 at each locus with the distribution expected under neutrality, conditional on levels of heterozygosity (Fig. 4). Tests were conducted only between genetic groups 1 and 2, since extensive evidence of long lasting asexual reproduction in group 3 violated the model assumptions. Simulations in FDIST2 were run with the following parameters: two demes, two populations, mean value of F_{ST} calculated between groups 1 and 2 ($F_{ST} = 0.04$) and the number of individuals without clonal copies, which comprised genetic groups 1 and 2 (322 individuals). Four microsatellite loci deviated from neutral expectations and displayed signs of positive selection (MLP54 and MLP59, $P < 0.05$ and MLP56 and MLP95 $P < 0.01$) (Fig. 4). No locus was detected under balancing

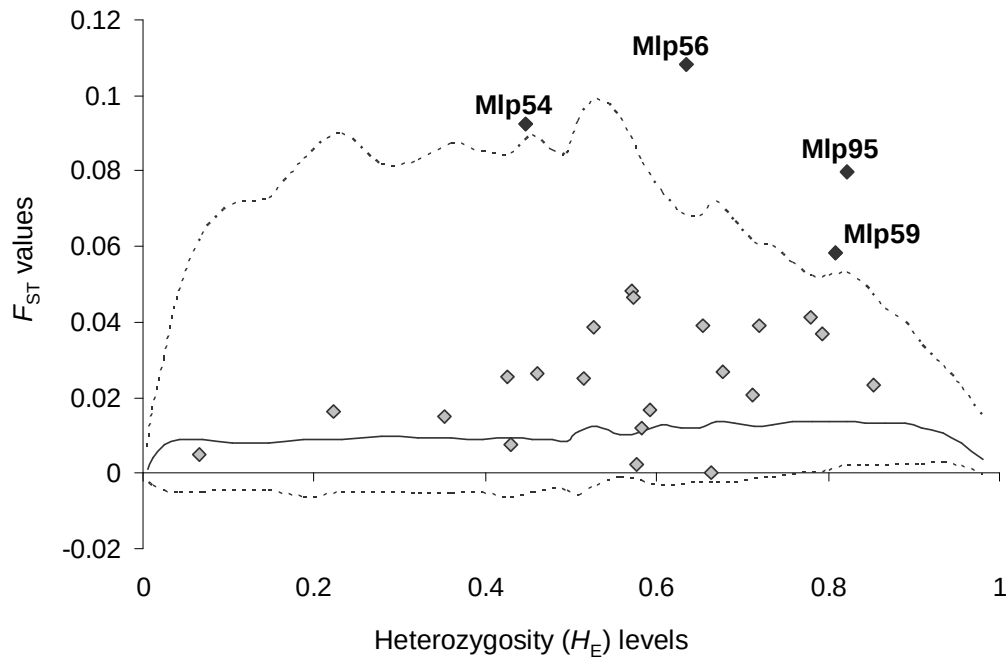


Figure 4: Test for selection between genetic groups 1 and 2. Observed F_{ST} values are plotted against heterozygosity levels for each locus. Solid and dashed lines denote median and 95% confidence interval of the simulated distributions under the Stepwise Mutation Model (SMM), respectively. Four loci were candidates for being under positive selection (black diamonds). The remaining loci were found with neutral expectations (grey diamonds).

Table 4: Tests for mutation-drift equilibrium. Bottleneck tests: number of loci exhibiting heterozygosity deficit (D) or excess (E) under the two mutation models (SMM, stepwise mutation model, or TPM, two-phase model). The probability of deviation from mutation-drift equilibrium was determined using a two-sided Wilcoxon test (**, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$). Imbalance index: mean estimates of $\ln \beta_2$ over loci and 95% confidence intervals (in brackets).

	Bottleneck tests (D/E)		Imbalance index
	SMM	TPM	$\ln \beta_2$
Group 1 'virulent-sexual'	20/1 ****	17/4 ***	2.83 (1.41-4.26)
Group 2 'avirulent-sexual'	20/1 ****	15/6 **	2.52 (0.90-4.14)
Prelles (reference population)	16/5 ***	11/10	2.52 (0.79-4.25)

selection. The same result was obtained with both mutation models, SMM and IAM. The four loci had lower levels of heterozygosity in group 1 than in group 2, suggesting that group 1 might have been exposed to stronger selective pressures than group 2.

*** Demographic history of genetic groups 1 and 2**

We used two complementary approaches to reveal the demographic history of *M. larici-populina* populations: a test for expected heterozygosity excess/deficit (Piry *et al.* 1999), and the value of the imbalance index (Kimmel *et al.* 1998). As a major assumption of these tests is that individuals reproduce sexually, they were performed on groups 1 and 2 only, not taking into account clonal copies. The tests were also performed on the reference population at Prelles located in the original poplar-larch sympatry area and thus expected to have experienced reduced demographic fluctuations. In addition, as loci under selection can bias estimates of demographic parameters, we retained only the 21 loci that did not significantly deviate from neutral expectations in analyses with FDIS2.

In analyses using BOTTLENECK, the two genetic groups exhibited a significant heterozygosity deficit (Table 4) whatever the mutation model considered (SMM or TPM), suggesting population expansion. The heterozygosity deficit was stronger under the SMM. Considering the reference population of Prelles mutation-drift disequilibria was observed only under the SMM.

The imbalance index was largely positive in all sub-samples considered (Table 4), suggesting that groups 1 and 2 of *M. larici-populina* have recently expanded following a bottleneck (Kimmel *et al.* 1998). On average, the value of the imbalance index was highest (and its density distribution narrowed) for group 1.

Discussion

*** Coexistence of three differentiated genetic groups on the same host plant species**

In the present study, we investigated the population genetic structure of a plant pathogenic fungus (*M. larici-populina*) on its wild host (*P. nigra*) in France, where both wild and cultivated poplars are sympatric in many locations. Bayesian clustering analyses revealed that *M. larici-populina* individuals could be clustered into three genetically distinct groups. These groups could coexist locally but predominated in different well-defined geographical

areas. The three groups displayed differences in terms of microsatellite variability and prevalence of virulent isolates able to infect cultivated poplar carrying resistance 7. Group 1 consisted mainly of virulent individuals while the two other genetic groups were mostly represented by avirulent individuals. Group 3 exhibited lower genotypic richness, extensive linkage disequilibrium across loci and a large heterozygote excess suggesting a predominantly asexual reproductive mode (Balloux *et al.* 2003; Halkett *et al.* 2005a). Conversely, genetic characteristics of group 1 and 2 were consistent with a predominantly sexual reproductive mode. In the following, group 1 is referred to as the ‘virulent-sexual’ group, group 2 as the ‘avirulent-sexual’ group and group 3 as the ‘avirulent-asexual’ group.

Numerous population genetic studies in phytopathogenic fungi highlighted strict host plant specialisation with large genetic differences among (not within) host plants (Enjalbert *et al.* 2005; Goyeau *et al.* 2007; Fournier & Giraud 2008; Zaffarano *et al.* 2008; Gladieux *et al.* 2010b, Giraud *et al.* 2010). Unlike these studies, a significant genetic structure was found among *M. larici-populina* samples collected on the same host plant, *P. nigra*, with individuals assigned to distinct genetic groups that can coexist locally. In most fungal plant pathogens, limited dispersal between selection on the host and mating favours host-specific differentiation because mutations for adaptation on a new host affect both fitness on the host and mating patterns (Giraud *et al.* 2006, 2010). However, as discussed in Barrès *et al.* (2008) and Gilabert *et al.* (2009), such processes of host-specific differentiation through host-specialisation are not expected in host-alternating species like *M. larici-populina*. Indeed, unless isolating mechanism(s) favouring assortative mating or linkage among loci involved in plant adaptation pre-exist (Bolnick & Fitzpatrick 2007), no possible barriers to gene flow can emerge thereby impeding host-specific differentiation. Gene mixing may be all the more strengthened in *M. larici-populina* as sexual reproduction occurs on the alternate host with a very scattered distribution ensuring extensive genetic reshuffling among individuals with various origins (Gérard *et al.* 2006). The genetic structure revealed by our study would result from the spatial distribution and local aggregation of *M. larici-populina* lineages with three distinct evolutionary histories that came into secondary contact on the same host species.

*** Strong association among microsatellite and phenotypic markers despite extensive gene flow**

We found evidence of a highly significant relationship between genetic groups and phenotype profiles, as the ‘virulent-sexual’ group 1 was found to consist mainly in individuals

with a Vir7 profile, contrary to the two other groups, which were almost free from Vir7 individuals.

The ‘avirulent-sexual’ group 2 predominated in the Alps (Prelles and Aoste locations), which are included in the native distribution area of *M. larici-populina* (Barrès *et al.* 2008), where larches and wild poplars have evolved in sympatry. In addition, the reference population at Prelles, located in the area of poplar-larch sympatry, displayed limited evidence of demographic disequilibrium and did not depart from expectations under panmixia, which is in accordance with a long history of evolution on the wild host *P. nigra*. Despite a certain degree of population sub-structure, the clustering of these individuals into the same genetic group 2 is consistent with a common ancestral origin and a long-standing coevolution with the wild host.

Several pieces of evidence indicate that individuals from the ‘virulent-sexual’ group 1 are derived from the breakdown of resistance 7 carried by cultivated poplars such as ‘Beaupré’. First we found that group 1 was genetically less diverse than group 2. This is in accordance with the loss of heterozygosity going hand-in-hand with pathogen evolution on cultivated crops (as in Gladieux *et al.* 2008; Burdon & Thrall 2008), as a result of founder effects due to the limited number of individuals able to infect the overcome cultivar(s) at resistance breakdown (Guérin *et al.* 2007). Second, tests for mutation-drift equilibrium revealed that the virulent population experienced demographic expansion after the initial genetic bottleneck associated with the resistance breakdown. Finally, in addition to demographic changes, the test for neutral evolution revealed that some loci were clearly under positive selection, and that these loci exhibited a drop in allelic diversity in group 1. Together with the tremendous population expansion (both revealed indirectly here, and directly observed in Pinon *et al.* 1998), these results are consistent with a very drastic selection pressure exerted by the cultivated host ‘Beaupré’ on *M. larici-populina*. Both selection by plant resistance and demographic fluctuations (founder event followed by rapid population expansion) caused the observed genetic differentiation.

Some intriguing results arise nonetheless in the analysis of the demographic history of genetic groups 1 and 2: the ‘avirulent-sexual’ group 2 taken as a whole was not at mutation-drift equilibrium and also displayed extensive (although less pronounced) evidence for population expansion. This might be due to violations of the assumption of the model implemented in BOTTLENECK such as population sub-structure or the alternation of sexual/asexual reproduction in the life cycle. However, the signal of population expansion might also be due to an increase in the distribution area of the ‘avirulent-sexual’ group 2

resulting from the extensive plantation of larches as ornamental or wood-production trees since the 18th century. The fact that this demographic event is older than the one resulting from R7 breakdown might be in accordance with the larger interlocus variance in demographic parameter estimates. Using historical samples of *M. larici-populina* might allow deciphering further the demographic history of these two genetic groups.

The ‘avirulent-asexual’ group 3 was also found to consist mostly of Avr7 individuals. Such tight link between phenotypic and genetic differentiation has been thoroughly assessed in asexual species of wheat rusts such as *Puccinia triticina* (Goyeau *et al.* 2007) and *Puccinia striiformis* f. sp. *tritici* (Enjalbert *et al.* 2005). The association results from the hitchhiking of neutral marker loci on selected virulence loci thanks to the lack of recombination. Assuming that the clonal lineages of *M. larici-populina* date back to before the breakdown of R7 resistance (1994), the absence of recombination would have prevented these individuals from being introgressed by virulence 7 alleles.

*** Virulence introgression dynamics and spatial expansion of Vir7 individuals on the wild host**

The ‘virulent-sexual’ group 1 was found to be the largest, and to account for nearly 50% of sampled individuals, most of which were gathered in the northern half of France. The spread of Vir7 individuals across poplar cultivation has been thoroughly assessed by population surveys contemporaneous with breakdown of the R7 resistance (Pinon *et al.* 1998). Historically, poplar cultivars carrying the R7 resistance gene, such as the cultivar ‘Beaupré’, were planted widely on large acreages in northern France, Belgium and the Netherlands between 1980 and 1995. This agrosystem favoured the breakdown of R7 in this area in 1994 (Pinon *et al.* 1998). Virulence 7 emerged in Belgium and in northern France, and rapidly spread southwards. Within five years, Vir7 individuals occurred in all French poplar cultivation areas including those, in southern France, where the acreage of R7 cultivars was much lower or null. According to a previous population genetics study, between 36% and 75% of individuals sampled on susceptible poplars in the northern half of France in 2001 were Vir7 (Gérard *et al.* 2006). Here, we confirmed that Vir7 individuals have spread widely on susceptible plants and that the Vir7 proportion has now reached 54% to 100% of individuals in the northern half of France. This rapid invasion of Vir7 individuals is consistent with the remarkable dispersal abilities of rust spores, as already shown in *M. larici-populina* (Barrès *et al.* 2008). Accordingly, we found no isolation by distance among individuals from

the ‘virulent-sexual’ group 1 sampled across France. Nonetheless, ‘virulent-sexual’ group 1 prevails only in the northern half of France, and some southern locations (Aoste, Ramières 1, Seilh and Villenave d’Ornon) still remain free of individuals belonging to this group. Interestingly, the spatial distribution of ‘virulent-sexual’ group 1 individuals matched very well the area where R7 poplar cultivars (like ‘Beaupré’) were massively planted. This spatial correlation is in accordance with the study of Montarry *et al.* (2008) on *P. infestans* suggesting that –because of the dispersal ability of plant pathogens– patterns resulting from the adaptation to the most abundant host genotype can only be revealed at large spatial scales.

According to the theoretical work of Lenormand (2002), the combination of strong selection pressure and high gene flow would lead to the (unnecessary) spread of genes conferring adaptation to the environment most likely to be encountered. Our results are consistent with this expectation and suggest that the plantation of R7 cultivars and the subsequent breakdown of the resistance has led to “gene swamping” (Lenormand 2002) in *M. larici-populina* populations. Given the large genetic signature of selection and the demographic history of ‘virulent’ group 1, it is quite likely that the spread of virulence 7 into the wild host consisted not only of the introgression of the single virulence gene but rather resulted from an unbalanced gene mixing that led to the replacement of populations from the wild original with the observed virulent genetic group. In support of this hypothesis, we found very few hybrid individuals with mixed membership in genetic groups 1 and 2 (the only consistent exception was found in Ramières 1) and individuals from these two groups only rarely live in sympatry (and never in equal abundance). This large degree of gene mixing is fostered by the peculiarity of the poplar rust life cycle, as sexual reproduction on larch ensures extensive genetic reshuffling between the individuals from different origins that have to gather at these specific habitats.

*** Polymorphism in the reproductive mode of *M. larici-populina***

Patterns of genetic variation were consistent with genetic group 3 being composed of individuals that reproduce mainly asexually (Balloux *et al.* 2003; Halkett *et al.* 2005a). Over-represented multilocus genotypes and heterozygote excess has been used previously to find evidence and quantify the asexual reproductive mode in a wide range of organisms (e.g. fungal pathogens: Burdon & Roelfs 1985; Goyeau *et al.* 2007; aphids: Halkett *et al.* 2005b, Gilabert *et al.* 2009; trees: Stoeckel *et al.* 2006; and nematodes: Villate *et al.* 2010). The numbers of repeated multilocus genotypes and the large heterozygote excess observed in

group 3 of *M. larici-populina* were consistent with the reproductive mode being mostly clonal and including only occasionally sexual events (Balloux *et al.* 2003). Persistence of clonal lineages over several years had never been suspected so far in *M. larici-populina*, and repeated multilocus genotypes were always considered as clonal copies resulting from asexual multiplication during a single annual epidemic phase (Gérard *et al.* 2006; Barrès *et al.* 2008). However, despite being the genetic group with the smallest sample size, individuals from group 3 appeared widespread. These long lasting clonal lineages may have stayed unnoticed so far due to the smaller size and lower resolving power of the marker set formerly used (Barrès *et al.* 2006), which would not have allowed revealing genetic partition of *M. larici-populina* individuals according to their differences in reproduction mode.

Rust fungi (Pucciniales) are well known for displaying potential polymorphism in their reproductive mode (Burdon & Roelfs 1985; Samils *et al.* 2001; Goyeau *et al.* 2007; Mboup *et al.* 2009; Jin *et al.* 2010). In particular, asexuality is the only reproductive mode possible in situations where the alternate host is missing (the plant on which sexual reproduction occurs for a heteroecious rust). This is the case either after migration of the pathogen to a new geographic area where the alternate host is absent (Hovmøller *et al.* 2002; Goyeau *et al.* 2007), or after its eradication (Burdon & Roelfs 1985). Alternatively, the distribution area of host alternating species with obligate sexual phase is constrained by the occurrence of the alternate host. A well-documented example is *M. medusae* f. sp. *deltoideae* with a native distribution in poplar-larch sympatry areas in Canada and annual (recurrent) colonizations of the USA resulting from a dead-end clonal propagation on poplars (Bourassa *et al.* 2007). In species that exhibit both reproductive modes, sexual and asexual lineages generally grow in allopatry. For example, Burdon & Roelfs (1985) documented how *P. graminis* asexual populations evolved in areas where the sexual host was eradicated. In addition some particular climatic conditions can be more favourable to asexual reproduction (Groth *et al.* 1995; Barrett *et al.* 2008a). The willow rust fungus *Melampsora larici-epitea* was shown to exhibit a dual reproductive mode depending on climate, with asexual reproduction in Northern Ireland, where mild winters could favour overwintering of asexual spores (urediniospores), and sexual reproduction on larch in Sweden, where cold winters could favour sexual forms (teliospores) (Samils *et al.* 2001). In the present study, most clonal lineages were found in southern France, where milder winters may favour the asexual survival of *M. larici-populina*, and/or may be less favourable to overwintering of sexual spores. However we also observed cases of local coexistence of sexual and asexual lineages, which is more puzzling, as competition should lead to the exclusion of one of the two kinds of

lineages. Further experiments are required to document differences in the life history traits of these lineages which can define subtle spatial or temporal niches, as it was found for another host-alternating parasite species (Halkett *et al.* 2006; Gilabert *et al.* 2009).

Conclusion

This study on the genetic structure of *M. larici-populina* individuals revealed that three genetic groups coexisted on the same host plant species, the wild host *P. nigra*. Our analyses revealed that this structure may result from (i) reproductive mode polymorphism, and (ii) the strong selection pressure imposed on the pathogen by the release of resistant domesticated plants. The three genetic groups prevailed in distinct regions. Features of the agricultural landscape may explain the pattern of spatial structure observed in this plant pathogen. Further examination of the possible link between relative cultivated host density and the observed spatial genetic structure may provide more insight into the selective pressures at work. In addition to a finer grained spatial monitoring, it would be worth developing a retrospective approach to retrace the temporal evolution of the genetic signatures of this drastic selective event on the genome of *M. larici-populina*.

Acknowledgements

We would like to thank Audrey Andanson, Alain Berthelot, Gilles Bossuet, Virginie Diévert, Arnaud Dowkiw, Nicolas Feau, Olivier Forestier, Etienne Klein, and François Lefèvre for their help in collecting samples, Adeline Sipieter and Béranger Bertin for their technical help, Cyril Dutech, Jérôme Enjalbert, Frédéric Fabre and Etienne Klein for insightful discussions and Helen Rothnie for her attentive reading and careful English language checking. Three anonymous referees provided valuable comments on earlier versions of this manuscript. Constance Xhaard was supported by a fellowship from the French Ministry of Education and Research (MESR). This research program was supported by grants from INRA (projet innovant EFPA), the Agence Nationale de la Recherche, contract number ANR 07-BDIV-003 (Emerfundis project) and the French Ministry for Agriculture and Fisheries (Programme 142, sous-action 27, CTPS).

Author information box

This study forms part of the PhD research of C. Xhaard, which addresses the gene flow between fungal populations of *Melampsora larici-populina* and the genetic consequences of dispersal. F. Halkett is a population biologist interested in the population genetics and evolutionary biology of fungal plant pathogens, with a confirmed experience in population genetics of clonal and partially clonal organisms. P. Frey has a long-standing knowledge of the biology and genetics of rust fungi. B. Barrès is a population geneticist of fungal plant pathogens, who did his PhD on the inference of dispersal processes in *M larici-populina*. P. Gladieux is a population geneticist particularly interested in speciation, adaptive divergence and the evolutionary response of pathogens to the domestication of their hosts. A. Andrieux is in charge of the molecular techniques used for fungus genotyping and B. Fabre is responsible for the phenotypic characterization of rust individuals.

References

- Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology & Evolution*, **19**, 535–544.
- Balloux F, Lehmann L, de Meeus T (2003) The population genetics of clonal and partially clonal diploids. *Genetics*, **164**, 1635–1644.
- Barrès B, Dutech C, Andrieux A, Caron H, Pinon J, Frey P (2006) Isolation and characterization of 15 microsatellite loci in the poplar rust fungus, *Melampsora larici-populina*, and cross-amplification in related species. *Molecular Ecology Notes*, **6**, 60–64.
- Barrès B, Halkett F, Dutech C, Andrieux A, Pinon J, Frey P (2008) Genetic structure of the poplar rust fungus *Melampsora larici-populina*: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, **8**, 577–587.
- Barrett LG, Thrall PH, Burdon JJ, Nicotra AB, Linde CC (2008a) Population structure and diversity in sexual and asexual populations of the pathogenic fungus *Melampsora lini*. *Molecular Ecology*, **17**, 3401–3415.
- Barrett LG, Thrall PH, Burdon JJ, Linde CC (2008b) Life history determines genetic structure and evolutionary potential of host-parasite interactions. *Trends in Ecology & Evolution*, **23**, 678–685.
- Beaumont MA, Nichols RA (1996) Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proceedings of the Royal Society B-Biological Sciences*, **263**, 1619–1626.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, 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).
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188.
- Bolnick DI, Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution and Systematics*, **38**, 459–487.
- Bourassa M, Bernier L, Hamelin RC (2007) Genetic diversity in poplar leaf rust (*Melampsora medusae* f. sp. *deltoideae*) in the zones of host sympatry and allopatry. *Phytopathology*,

97, 603-610.

- Bucheli E, Gautschi B, Shykoff JA (2001) Differences in population structure of the anther smut fungus *Microbotryum violaceum* on two closely related host species, *Silene latifolia* and *S. dioica*. *Molecular Ecology*, **10**, 285-294.
- Burdon JJ, Chilvers GA (1982) Host density as a factor in plant disease ecology. *Annual Review of Phytopathology*, **20**, 143-66.
- Burdon JJ, Roelfs AP (1985) The effect of sexual and asexual reproduction on the isozyme structure of populations of *Puccinia graminis*. *Phytopathology*, **75**, 1068-1073.
- Burdon JJ (1993) The structure of pathogen populations in natural plant-communities. *Annual Review of Phytopathology*, **31**, 305-323.
- Burdon JJ, Thrall PH (2008) Pathogen evolution across the agro-ecological interface: implications for disease management. *Evolutionary Applications*, **1**, 57-65.
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001-2014.
- Couch BC, Fudal I, Lebrun MH, Tharreau D, Valent B, van Kim P, Notteghem JL, Kohn LM (2005) Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics*, **170**, 613-630.
- Daszak P, Cunningham AA, Hyatt AD (2000) Wildlife ecology - Emerging infectious diseases of wildlife - Threats to biodiversity and human health. *Science*, **287**, 443-449.
- Desprez-Loustau ML, Robin C, Buee M, Courtecuisse R, Garbaye J, Suffert F, Satche I, Rizzo DM (2007) The fungal dimension of biological invasions. *Trends in Ecology & Evolution*, **22**, 472-480.
- Diamond J (2002) Evolution, consequences and future of plant and animal domestication. *Nature*, **418**, 700-707.
- Enjalbert J, Duan X, Leconte M, Hovmøller MS, De Vallavieille-Pope C (2005) Genetic evidence of local adaptation of wheat yellow rust (*Puccinia striiformis* f. sp. *tritici*) within France. *Molecular Ecology*, **14**, 2065-2073.
- Fournier E, Giraud T (2008) Sympatric genetic differentiation of a generalist pathogenic fungus, *Botrytis cinerea*, on two different host plants, grapevine and bramble. *Journal of Evolutionary Biology*, **21**, 122-132.
- Frey P, Gérard PR, Feau N, Husson C, Pinon J (2005) Variability and population biology of *Melampsora* rusts on poplars. In *Rust diseases of Willow and Poplar* (eds. Pei MH, McCracken AR) pp. 63-72. CAB International, Wallingford, UK.

- Gérard PR, Husson C, Pinon J, Frey P (2006) Comparison of genetic and virulence diversity of *Melampsora larici-populina* populations on wild and cultivated poplar and influence of the alternate host. *Phytopathology*, **96**, 1027–1036.
- Gilbert GS (2002) Evolutionary ecology of plant diseases in natural ecosystems. *Annual Review of Phytopathology*, **40**, 13–43.
- Gilabert A, Simon JC, Mieuze L, Halkett F, Stoeckel S, Plantegenest M, Dedryver CA (2009) Climate and agricultural context shape reproductive mode variation in an aphid crop pest. *Molecular Ecology*, **18**, 3050–3061.
- Giraud T, Villareal L, Austerlitz F, Le Gac M, Lavigne C (2006) Importance of the life cycle in sympatric host race formation and speciation of pathogens. *Phytopathology*, **96**, 280–287.
- Giraud T, Gladieux P, Gavrillets S (2010) Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology & Evolution*, **25**, 387–395.
- Gladieux P, Zhang XG, Afoufa-Bastien D, Sanhueza RMV, Sbaghi M, Le Cam B (2008) On the origin and spread of the scab disease of apple: out of Central Asia. *Plos One*, **3**, e1455
- Gladieux P, Zhang XG, Roldan-Ruiz I, Caffier V, Leroy T, Devaux M, Van Glabeke S, Coart E, Le CamB (2010a) Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Molecular Ecology*, **19**, 658–674.
- Gladieux P, Caffier V, Devaux M, Le CamB (2010b) Host-specific differentiation among populations of *Venturia inaequalis* causing scab on apple, pyracantha and loquat. *Fungal Genetics and Biology*, **47**, 511–521.
- Gomez-Alpizar L, Carbone I, Ristaino JB (2007) An Andean origin of *Phytophthora infestans* inferred from mitochondrial and nuclear gene genealogies. *Proceedings of the National Academy of Sciences*, **104**, 3306–3311.
- Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.
- Goyeau H, Halkett F, Zapater MF, Carlier J, Lannou C (2007) Clonality and host selection in the wheat pathogenic fungus *Puccinia triticina*. *Fungal Genetics and Biology*, **44**, 474–483.
- Groth JV, McCain JW, Roelfs AP (1995) Virulence and isozyme diversity of sexual versus asexual collections of *Uromyces-Appendiculatus* (Bean rust fungus). *Heredity*, **75**, 234–242.

- Guérin F, Gladieux P, Le Cam B (2007) Origin and colonization history of newly virulent strains of the phytopathogenic fungus *Venturia inaequalis*. *Fungal Genetics and Biology*, **44**, 284–292.
- Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–1280.
- Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.
- Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation--the correlated allele frequencies model revisited. *Bioinformatics*, **24**, 2222–2228.
- Hansen EM (1987) Speciation in plant pathogenic fungi – the influence of agricultural practice. *Canadian Journal of Plant Pathology*, **9**, 403–410.
- Halkett F, Simon JC, Balloux F (2005a) Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, **20**, 194–201.
- Halkett F, Plantegenest M, Prunier-Leterme N, Mieuze L, Delmotte F, Simon JC (2005b) Admixed sexual and facultatively asexual aphid lineages at mating sites. *Molecular Ecology*, **14**, 325–336.
- Halkett F, Kindlmann P, Plantegenest M, Sunnucks P, Simon JC (2006) Temporal differentiation and spatial coexistence of sexual and facultative asexual lineages of an aphid species at mating sites. *Journal of Evolutionary Biology*, **19**, 809–815.
- Hovmøller MS, Justesen AF, Brown JKM (2002) Clonality and long-distance migration of *Puccinia striiformis* f. sp. *tritici* in north-west Europe. *Plant Pathology*, **51**, 24–32.
- Imbert E, Lefevre F (2003) Dispersal and gene flow of *Populus nigra* (Salicaceae) along a dynamic river system. *Journal of Ecology*, **91**, 447–456.
- Jin Y, Szabo L, Carson M (2010). Century-old mystery of *Puccinia striiformis* life history solved with the identification of *Berberis* spp. as an alternate host. *Phytopathology*, **100**, 432–435.
- Kimmel M, Chakraborty R, King JP, Bamshad M, Watkins WS, Jorde LB (1998) Signatures of Population Expansion in Microsatellite Repeat Data. *Genetics*, **148**, 1921–1930.
- King JP, Kimmel M, Chakraborty R (2000) A Power Analysis of Microsatellite-Based Statistics for Inferring Past Population Growth. *Molecular Biology and Evolution*, **17**, 1859–1868.
- Kiros-Meles A, Gomez D, McDonald BA, Yahyaoui, A, Linde CC (2011) Invasion of *Rhynchosporium commune* onto wild barley in the Middle East. *Biological Invasions*,

13, 321–330.

- Lebarbenchon C, Brown SP, Poulin R, Gauthier-Clerc M, Thomas F (2008) Evolution of pathogens in a man-made world. *Molecular Ecology*, **17**, 475–484.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.
- Lê Van A, Durel C, Le Cam B, Caffier V (2011) The threat of wild habitat to scab resistant apple cultivars. *Plant Pathology*, doi: 10.1111/j.1365-3059.2011.02437.x
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
- Loader CR (1996) Local Likelihood Density Estimation. *Annals of Statistics*, **24**, 1602–1618.
- Mboup M, Leconte M, Gautier A, Wan AM, Chen W, de Vallavieille-Pope C, Enjalbert J (2009) Evidence of genetic recombination in wheat yellow rust populations of a Chinese overwintering area. *Fungal Genetics and Biology*, **46**, 299–307.
- McDonald B (2010) How can we achieve durable disease resistance in agricultural ecosystems? *New Phytologist*, **185**, 13.
- Montarry J, Glais I, Corbière R, Andrivon D (2008) Adaptation to the most abundant host genotype in an agricultural plant–pathogen system – potato late blight. *Journal of Evolutionary Biology*, **21**, 1397–1407.
- Munkacsı AB, Stoxen S, May G (2008) *Ustilago maydis* populations tracked maize through domestication and cultivation in the Americas. *Proceedings of the Royal Society B-Biological Sciences*, **275**, 1037–1046.
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.
- Parker IM, Gilbert GS (2004) The evolutionary ecology of novel plant-pathogen interactions. *Annual Review of Ecology Evolution and Systematics*, **35**, 675–700.
- Pei MH, Bayon C, Ruiz C, Tubby I (2007) Population structure of poplar rust *Melampsora larici-populina* in the UK inferred from AFLP. *Plant Pathology*, **56**, 472–479.
- Pinon J, Frey P, Husson C, Schipfer A (1998) Poplar rust (*Melampsora larici-populina*): The development of E4 pathotypes in France since 1994. Pages 57–64 in: Proceedings of the First IUFRO Rusts of Forest Trees Conference, Saariselkä, Finland. R Jalkanen, PE Crane, JA Walla, T Aalto, eds. Finnish Forest Research Institute, Research Paper 712.
- Pinon, J, Frey, P, (2005) Interactions between poplar clones and *Melampsora larici-populina* populations and their implications for breeding for durable resistance. In Rust diseases of Willow and Poplar (eds. Pei MH, McCracken AR) pp. 139–154. CAB International,

Wallingford, UK.

- 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.
- Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature*, **457**, 843–848.
- Rizzo DM, Garbelotto M, Hansen EM (2005) *Phytophthora ramorum*: integrative research and management of an emerging pathogen in California and Oregon forests. *Annual Review of Phytopathology*, **43**, 309–335.
- Rousset F (1997) Genetic Differentiation and estimation of gene flow from *F*-Statistics under isolation by distance. *Genetics*, **145**, 1219–1228.
- Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.
- Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
- Samils B, Stepien V, Lagercrantz U, Lascoux M, Gullberg U (2001) Genetic diversity in relation to sexual and asexual reproduction in populations of *Melampsora larici-epitea*. *European Journal of Plant Pathology*, **107**, 871–881.
- Stenberg P, Lundmark M, Saura A (2003) MLGsim: a program for detecting clones using a simulation approach. *Molecular Ecology Notes*, **3**, 329–331.
- Stoeckel S, Grange J, Fernandez-Manjarres JF, Bilger I, Frascaria-Lacoste N, Mariette S (2006) Heterozygote excess in a self-incompatible and partially clonal forest tree species - *Prunus avium* L. *Molecular Ecology*, **15**, 2109–2118.
- Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9440–9445.
- Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA (2007) Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Molecular Biology and Evolution*, **24**, 398–411.
- Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, **46**, 75–100.
- Villate L, Esmenjaud D, Van Helden M, Stoeckel S, Plantard O (2010) Genetic signature of amphimixis allows for the detection and fine scale localization of sexual reproduction events in a mainly parthenogenetic nematode. *Molecular Ecology*, **19**, 856–873.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population

- structure. *Evolution*, **38**, 1358–1370.
- Wisler GC, Norris RE (2005) Interactions between weeds and cultivated plants as related to management of plant pathogens. *Weed Sciences*, **53**, 914–917.
- Xhaard C, Andrieux A, Halkett F, Frey P (2009) Characterization of 41 microsatellite loci developed from the genome sequence of the poplar rust fungus, *Melampsora larici-populina*. *Conservation Genetics Resources*, **1**, 21–25.
- Yarwood CE (1970) Man-made plant diseases. *Science*, **168**, 218–220.
- Zaffarano PL, McDonald BA, Linde CC (2008) Rapid speciation following recent host shifts in the plant pathogenic fungus *Rhynchosporium*. *Evolution*, **62**, 1418–1436.

Chapitre 2

Structure génétique spatiale de
Melampsora larici-populina dans la
zone de sympatrie peuplier-mélèze

Chapitre 2

Structure génétique spatiale de *Melampsora larici-populina* dans la zone de sympatrie peuplier-mélèze

Dans le chapitre précédent, nous avons décrit la structure génétique spatiale des populations de *M. larici-populina* en France. Nous avons montré que les populations de rouilles étaient différenciées en trois groupes présentant des signatures génétiques qui reflètent des histoires évolutives différentes. Dans la suite de ce manuscrit de thèse ces groupes génétiques seront dénommés « cultivé », « sauvage » et « asexué » (respectivement en anglais “cultivated”, “wild” and “asexual”).

Si l'on s'intéresse plus particulièrement aux groupes « cultivé » et « sauvage », nous avons montré dans le chapitre 1 qu'il existait une structuration Nord-Sud dans la répartition spatiale de ces deux entités génétiques, pouvant être expliquée par une plus forte densité d'hôtes cultivés porteurs de la résistance R7 dans le nord de la France. Dans ce second chapitre nous nous sommes intéressés à une région située dans une zone de contact entre ces deux grands groupes génétiques. La région sur laquelle nous nous sommes focalisés se situe dans les Alpes, dans une zone où les peupliers et les mélèzes évoluent en sympatrie, induisant une forte prévalence de la reproduction sexuée de *M. larici-populina*. Dans ce chapitre nous avons cherché à découvrir les raisons du maintien de la structuration génétique malgré le fort brassage génétique que l'on peut suspecter au moment de la reproduction sexuée. Plus spécifiquement nous avons testé l'hypothèse d'un effet du paysage dans cette région de la France sur la résistance à l'envahissement des individus du groupe « cultivé ».

En nous plaçant à cette échelle plus fine et en utilisant des outils de génétique du paysage nous avons cherché à savoir (i) si les individus des deux groupes présentaient une structuration spatiale, et si elle correspondait à une caractéristique du paysage ; (ii) si les domaines spatiaux étaient en contact avec la présence d'hybrides entre groupes génétiques et (iii) si nous pouvions relier la structuration spatiale avec l'histoire démographique récente des populations de *M. larici-populina*.

Ce chapitre se présente sous la forme d'un article en préparation.

ARTICLE 3 : THE WEST-SIDE STORY OF A PLANT PATHOGEN: LANDSCAPE GENETICS OF THE POPLAR RUST FUNGUS IN THE FRENCH ALPS

Xhaard C., Andrieux A., Fabre B., Frey P., Halkett F.

Introduction

Landscape genetics has recently emerged as the result of the combination of landscape ecology and population genetics (Manel *et al.* 2003; Storfer *et al.* 2007). The development of landscape genetics has been fostered by the many advances in spatial genetic analyses (Guillot *et al.* 2009). The goal of this new research area is to relate landscape features directly to genetic variation across individuals, usually measured with neutral molecular markers (Manel *et al.* 2003). One of the major aims of landscape genetics is to detect natural or anthropogenic barriers to gene flow (Storfer *et al.* 2007). Barriers to gene flow are evidenced when abrupt genetic discontinuities match natural or anthropogenic features (Coulon *et al.* 2006; Fontaine *et al.* 2007; Sahlsten *et al.* 2008). To this aim a dedicated methodology is to use spatially explicit Bayesian clustering algorithms (Guillot *et al.* 2005a; 2009; Guillot 2008; Chen *et al.* 2007, Durand *et al.* 2009), which permit to delineate spatial distribution of genetic entities from posterior clustering of individuals according to their genetic proximity.

Hitherto the majority of landscape genetics studies focused on wildlife management and until recently this approach has seldom been applied to pathogens (Rieux *et al.* 2011). When applied to diseases, this research area is located at the interface of many fields that are landscape epidemiology, disease ecology and population genetics (Archie *et al.* 2009; Biek & Real 2010), with most studies dealing with host populations' distribution in association with disease incidence (Antolin *et al.* 2006; Blanchong *et al.* 2008). Historically, epidemiology and population genetic studies focused on different spatial and temporal scales. In epidemiology, the study area is generally restricted to the field, whereas in population genetics, the scale of the study is usually larger, for example at the continental or intercontinental scale. In the present study, we attempted to conciliate these two approaches in a particular environmental context, in using landscape genetic approaches.

We focused here on an airborne plant pathogen, the poplar rust fungus *Melampsora larici-populina*, which causes severe damages and economic losses in poplar cultivation (Frey *et al.* 2005). Poplar rust has become a model pathosystem for perennial plants since the

availability of the genome sequences of both the host (Tuskan *et al.* 2006) and the pathogen (Duplessis *et al.* 2011). As other airborne pathogens, *M. larici-populina* exhibits high dispersal capacities (Brown & Hovmøller 2002; Barrès *et al.* 2008), even at very long distances (Nagarajan & Singh 1990; Brown & Hovmøller 2002). Thus, finding some spatial genetic structure for wind-dispersed microorganisms can be challenging, because homogeneous genetic structure is expected for such species due to their high dispersal capacities (Finlay 2002). Accordingly, only a weak Isolation By Distance signal has been evidenced at the continental scale in the poplar rust (Barrès *et al.* 2008) indicating the lack of spatial structure. However when increasing the number of genetic markers, *M. larici-populina* individuals exhibited genetic structure at a country scale (Xhaard *et al.* 2011).

The genetic groups detected in this previous study (chapter 1) showed some signs of spatial structure with a north-south delimitation. The spatial distributions were explained by the specific epidemiological history of *M. larici-populina*, following the breakdown of the R7 resistance gene carried by some poplar cultivars widely planted in northern France. This major selective event exerted on *M. larici-populina* populations gave rise to the most abundant genetic group, denoted “cultivated” group because individuals from this group displayed phenotypic and genetic characteristics consistent with a history of selection by cultivated poplars. On the opposite, *M. larici-populina* individuals from the second genetic group were not able to infect these cultivated poplars and exhibited signs of “wild” origin (Xhaard *et al.* 2011, chapter 1). This “wild” group was more or less restricted to southern France.

The area studied here, located in the French Alps, was suggested as one of the contact zones between spatial domains of the two genetic groups (chapter 1). Deeply embanked valleys and high mountain ranges characterize this area. In this study, we interested in the spatial genetic structure of a plant pathogenic fungus in this alpine landscape and aimed to assess whether the variation in allele frequencies matches some features of this area. More specifically, we tested the hypothesis of a lack of gene flow due to a high mountain range, resulting in the geographic distribution of *M. larici-populina* genetic groups. To this aim we developed an intensive sampling effort, collecting *M. larici-populina* individuals at many sites and investigating valleys located on both sides of a mountain range. Then we analysed the samples in using landscape genetics tools.

Here we compared several Bayesian clustering algorithms with different options. We tested them with and without the information of the geographical origin of individuals sampled. We tested also algorithms with and without admixture. These diverse approaches

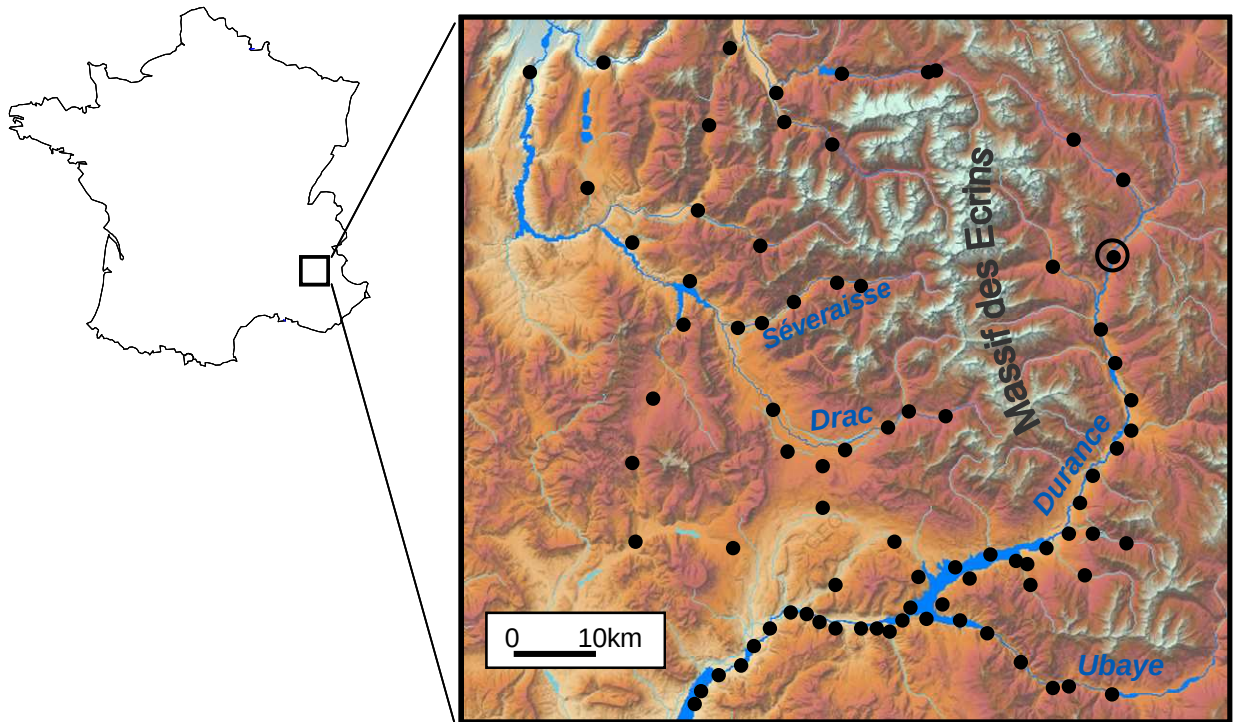


Figure 1: Sampled site distribution and geographic map of the study area (about 80 x 80 km) located in the French Alps. Lakes and main rivers are figured in blue (small rivers in pale blue). Other colours depict the relief. Names of the main rivers and that of the mountain chain are written down in blue and grey respectively. Sampled locations are pointed as black dots. The reference site of Prelles (upstream location along the Durance valley) is highlighted by a black circle.

should be used in parallel to distinguish between sharp and gradual variation in allele frequencies (Guillot & Santos 2009) and to quantify hybrid individuals in a contact zone. Lastly, spatial interpolations were used to test the relative influence of landscape features.

Materials and methods

* Study area

The study area is located in the Alps, in south-eastern France, where a sharp discontinuity between two spatially segregated genetic groups has been observed (Xhaard *et al.* 2011, chapter 1). The study area is a square of side 80 km. The spatial domain sampled is characterized by high reliefs and encompasses a large mountain range, the Ecrins range, of about 35 km long, culminating at 4100 m elevation. This high relief is likely to act as a barrier to gene flow as the *M. larici-populina* individuals exhibiting the large genetic differentiation were found on each side of the Ecrins range (Xhaard *et al.* 2011, chapter 1). This mountain range has a north-west to south-east direction, and splits the northern part of the spatial domain into two catchment basins. Two main rivers spread across this area: the Drac River, joined by the Severaisse River on the west side and the Durance River on the east side, which is joined by the Ubaye River in the Serre-Ponçon artificial lake (Fig. 1). All rivers are deeply embanked in different valleys. In this natural area, European larch (*Larix decidua*) and wild black poplar (*Populus nigra*) —the two successive hosts of *M. larici-populina*— grow in sympatry allowing an easy completion of the life cycle of this fungus. Briefly, *M. larici-populina* performs an annual sexual reproduction stage on larch needles in spring and then switches to several rounds of clonal multiplication on poplar leaves during summer and autumn (Xhaard *et al.* 2011). The disease incidence is benign on larches, but rust attacks causing the annual epidemic on wild poplars can be severe and lead to early leaf fall.

Natural larch stands are found on the slopes of the mountains and along the rivers, between 1000 and 2400 m elevation. Wild poplars —on which sampling was done— mostly grow in riparian stands facing south below 1500 m elevation, except for some old trees planted by man along the roads or fields. It is noteworthy that the studied region is almost free from cultivated poplar stands that may exert peculiar selection pressures on *M. larici-populina* populations (Gérard *et al.* 2006, Xhaard *et al.* 2011). However, there are many poplar plantations, with some recent rust-resistant cultivars, in the Isère River valley, about 50 km northwest of the studied region, downstream the Drac River.

*** Sampling strategy**

Melampsora larici-populina sampling was carried out on wild poplars, in October 2008 at the end of the annual epidemic phase. Eighty one sites were chosen to cover homogeneously the spatial distribution of wild poplars. At each sampling site, between 18 and 50 infected leaves were collected from a few closely spaced poplar trees, and GPS coordinates were recorded with a precision of 10 m. Whenever possible, sampling was performed in wild poplar riparian stands; otherwise (at higher altitudes) infected leaves were collected on adult trees along the roads. Between 16 and 48 *M. larici-populina* individuals were isolated per sampling site.

*** Individual isolation**

One single sporulating lesion of *M. larici-populina* per leaf was selected randomly in order to avoid excessive clonal replicates, and grown on fresh leaf discs of *P. x euramericana* ‘Robusta’ as described by Barrès *et al.* (2008). Mono-uredinial isolation ensures that a single genotype per sampled leaf is isolated and multiplied, so that both virulence phenotype assessment and molecular analysis can be performed on the same individual. Sporulating leaf discs were stored at –20°C until DNA extraction.

*** Virulence phenotypic assessment**

A subset of 872 individuals from 70 locations across the study area was phenotyped on their ability to overcome (or not) the R7 qualitative resistance gene carried by *P. x interamericana* ‘Beaupré’. The majority of poplar cultivars were selected for their resistance to rust attack using qualitative resistance. The most damaging resistance breakdown occurred for the cultivar ‘Beaupré’ and the corresponding virulent *M. larici-populina* individuals, initially found in 1994 in Belgium and northern France, have rapidly spread all over France (Pinon *et al.* 1998; Pinon & Frey 2005). Sampled *M. larici-populina* individuals were inoculated on two poplar cultivars, one susceptible (*P. x euramericana* ‘Robusta’) and the other carrying the R7 rust resistance gene (*P. x interamericana* ‘Beaupré’), as described by Barrès *et al.* (2008). Individuals were denoted virulent 7 (Vir7) if both resistant and susceptible leaf discs were infected; and avirulent 7 (Avr7) if only the susceptible (control) leaf disc was infected. Phenotypic analyses were conducted computing the proportion of Vir7 individuals for each sampled location. Then we interpolated these proportions with the geographical coordinates of the locations with the R software (R-Development-Core-Team,

2008) using the `mba.surf` function from MBA library in order to visualize the distribution of the proportion of *VirZ* individuals across the study area.

*** Microsatellite analysis**

DNA was extracted from infected leaf discs using the BioSprint 96 DNA plant kit used in combination with the BioSprint 96 automated workstation (Qiagen) following the BS96-DNA-plant protocol. Microsatellite analyses were performed at 25 polymorphic loci: MLP12 (Barrès *et al.* 2006); MLP49, MLP50, MLP54, MLP55, MLP56, MLP57, MLP58, MLP59, MLP66, MLP68, MLP71, MLP73, MLP77, MLP82, MLP83, MLP87, MLP91, MLP92, MLP93, MLP94, MLP95, MLP96, MLP97 (Xhaard *et al.* 2009) and MLP100 (Xhaard *et al.* 2011). Microsatellite loci were amplified in a GeneAmp PCR system 9700 thermocycler (Applied Biosystems) under the following conditions: an initial denaturing step of 5 min at 95°C; 32 cycles including 30 s of denaturation at 95°C, 90 s annealing at 60°C and 60 s extension at 72°C; and a final 30 min extension step at 60°C. PCR was carried out in a 10 µL final reaction volume containing 5 µL of Qiagen® Multiplex PCR kit, 2 µL of DNA, 0.2 µM of each reverse and forward primers. Forward primers were labelled with a fluorescent tag (FAM, NED, VIC or PET). Three multiplex PCR were run, comprising 9, 7, and 9 loci, respectively (Table 2). PCR products from the three multiplex reactions (3 µL PCR1, 4 µL PCR2 and 5 µL PCR3) were pooled and loaded on a ABI 3730 Genetic Analyzer (Applied Biosystems). Fragments were sized with a LIZ-1200 size standard, and alleles were scored using GENEMAPPER 4.0 (Applied Biosystems). Individuals for which more than height loci failed to amplify were omitted from further analyses.

*** Data analyses**

Identification of repeated multilocus genotypes

When pooling all multilocus genotypes from all locations, we identified individuals carrying repeated multilocus genotypes. These individuals were considered as clonal copies resulting from local asexual multiplication and were therefore excluded from further analyses.

Clustering analyses and assessment of spatial genetic structure

We investigated population subdivision in performing Bayesian genetic assignment tests using four different programs. First we used STRUCTURE (Pritchard *et al.*, 2000) which has been recognised as the most popular used clustering algorithm (Evanno *et al.*, 2005; Kauffer *et al.*, 2007). We used admixture model with correlated allele frequency and default

settings. Iteration parameters were set to a burning-in period of 30 000 iterations followed by 1,200 000 iterations. The likelihood of the assignments ($\ln P(D)$) was evaluated for K varying from 1 to 10. Three independent simulations were performed for each value of K to test for the consistency of the results. To assess the number of clusters K that best explain the partition of the genetic data, we computed the ΔK estimator as described by Evanno *et al.* (2005).

Second, we used GENELAND (version 3.1.4) (Guillot 2008; Guillot *et al.* 2005b). GENELAND offers two types of analyses, taking –or not– information on the geographical origin of individuals into account to assign individuals to genetic groups. In the spatial approach, clusters correspond to spatially and genetically continuous units separated by small discontinuities that occur where genetic barriers are crossed. Assignment tests were conducted using both spatial and non spatial methods, which can be more or less sensitive to the within-site genetic admixture. Another strength of GENELAND is that it infers the best number of genetic groups, K , during a single run using the specific Reverse Jump MCMC (Markov chain Monte Carlo) algorithm, which enables the Markov chain to explore the whole range of putative cluster numbers at each step of the simulation run. The number of genetic groups that best explain the data is given by the value of K in which the Markov chain spent most iterations (Guillot *et al.* 2005a,b). The value of K was allowed to vary from 1 to 30 and initialised at 12. Three independent runs were performed to check for convergence of Markov chains. This first batch of runs with varying K enabled the best putative number of genetic groups to be assessed. Once the most likely K value was determined, three additional runs were performed, with the corresponding value of K fixed, to better infer the assignment level of each individual to the K genetic groups (Guillot *et al.* 2005b). For the three runs with varying K , we used the uncorrelated allele frequency model, and for those with fixed K we used the correlated allele frequency model according to Guillot (2008) with default settings, except for the number of nuclei, which was increased to 1500 (in the non-spatial mode), given the number of individuals. Run length was set to 500 000 iterations with a burning-in period of 50 000 iterations.

Third, we used TESS 2.3.1 (Chen *et al.* 2007), which –as GENELAND– includes spatially explicit prior distributions. Compared to GENELAND, TESS also implements an admixture model (as STRUCTURE does). In order to compare the outputs across clustering programs, we used TESS under four crossed modalities: considering or not spatial information and using or not the admixture model. When considering spatial information, prior for admixture was defined according to the BYM model with quadratic trend surface (if the real spatial

coordinates were used) otherwise admixture followed a constant trend surface over the spatial domain (Durand *et al.* 2009). For each modality, run length was set to a burn-in period of 50 000 iteration followed by 500 000 iterations. We increased the number of clusters from $K = 2$ to $K = 8$ (8 replicates for each value of K). It is noteworthy that it is not possible to run TESS for $K = 1$ (as a test of no population structure). The best value of K was identified using the DIC (Deviation Index Criterion) (Durand *et al.* 2009). The value of K that best explain the data is likely to be given by smallest value of the DIC.

Finally we used NEWHYBRIDS (1.1) that implements a Bayesian genetic assignment approach dedicated to the quantification of hybridization events (Anderson & Thompson 2002). The detection of backcrosses requires much more loci (almost 50) than the 25 used here (Vaha & Primmer 2006). We therefore focused on the first generation of hybridization between parental classes. The Bayesian method allows either loading prior knowledge on the allele frequencies of the parental classes or estimating them from the offspring genetic data set. Here we used the program without any prior information. We used the following parameters: a short burn-in period of 1000 iterations, following by more than 1 000 000 iterations. Three independent runs were performed to check for the consistency of the convergence of Markov chains.

After identifying the best value of K for each Bayesian clustering programs, the matrices of posterior individual membership (q coefficients) to each genetic group was averaged across runs' outputs (taking into account label switching). The mean values of q coefficients to each genetic group were interpolated along the geographic coordinates of sampled locations using the `mba.surf` function of the R software to map the distribution of genetic groups (and putative admixture or hybrid zone) over the spatial domain considered.

Population genetics analyses

We computed summary statistics of genetic variability on all the dataset and for each genetic group identified after Bayesian clustering methods. Expected and observed heterozygosities, values of F_{ST} (differentiation coefficient between populations) and F_{IS} (deviation from random mating within population) which were estimated according to Weir and Cockerham (1984) and allelic richness (A_r) that implements a rarefaction procedure to account for differences in sample size were estimated using FSTAT version 2.9.3 (Goudet 1995). Linkage disequilibrium between loci and deviations from HW Equilibrium expectation (heterozygote excess and deficit) were calculated using GENEPOP on the web (<http://genepop.curtin.edu.au>) (Rousset 2008). In order to adjust the P value for multiple tests

Table 1: Inference of the number of cluster(s) (K) that best explain the genetic structure among sample. For STRUCTURE runs we reported the mean value of log likelihood across simulations and the ΔK index computed according to Evanno *et al.* (2005). For GENELAND runs we reported the percentage of iterations spend by the RJMCMC across simulation (after the burning period). For TESS we reported the values of DIC (Deviation Index Criterion) with the standard error across runs in brackets.

K	STRUCTURE		GENELAND		TESS			
	Log-likelihood	ΔK	Spatial	Non-spatial	Spatial Non-admixed	Non-spatial Non-admixed	Spatial Admixed	Non-spatial Admixed
1	-70691	-	64.13	0	-	-	-	-
2	-69750	722	17.76	100	207737 (0.71)	208296 (0.75)	138670 (34.04)	138658 (49.52)
3	-69531	133	10.53	0	207737 (0.99)	208296 (1.11)	137952 (990.95)	138207 (125.94)
4	-69446	107	4.77	0	207737 (0.87)	208298 (2.24)	137156 (658.74)	136663 (1594.67)
5	-69468	55	1.88	0	207737 (1.30)	208296 (3.60)	134789 (1345.53)	133979 (1096.78)
6	-69435	55	0.62	0	207738 (1.00)	208296 (2.93)	132755 (960.48)	132087 (477.49)
7	-69457	104	0.22	0	207736 (1.11)	208295 (3.06)	132637 (1184.36)	131697 (1233.36)
8	-69375	39	0.09	0	207737 (0.71)	208297 (1.73)	131854 (793.49)	130888 (1286.50)
9	-69332	7	0.01	0	-	-	-	-
10	-69297	-	0.00	0	-	-	-	-

for linkage disequilibrium, we used the false discovery rate (FDR) procedure (Benjamini & Yekutieli 2001), which controls the proportion of significant results ('false discovery rate') instead of controlling the chance of making even a single error. The resulting adjusted P values are called Q values. This procedure is implemented in the R package QVALUE (Storey & Tibshirani 2003).

Isolation by distance testing

Isolation by distance tests were performed using the continuous method (Rousset 2000) of the "ISOLDE" procedure implemented in GENEPOP (Rousset 2008, option "Analysis of isolation by distance between groups"). The strength of this method is to minimize standard deviation of the denominator of the estimates in evaluating intra-group diversity among all locations, thus much fewer individuals can be analysed per sampling site. The significance of Mantel tests between pairwise genetic distances a_r –analogue of $F_{ST}/(1-F_{ST})$ ratios– and geographical distances were performed using the ISOLDE program (10,000 permutations). For geographical distance we considered Euclidian distances between pairs of sampling sites. We subsequently extracted from this distance matrix the values of pairwise genetic distances of all locations with Preles as reference site located along the Durance River (Xhaard *et al.* 2011) to draw a map of genetic differentiation across the sampling domain (according to the same interpolation procedure described above). These values were also used to investigate patterns of genetic differentiation along Rivers located on west vs. east side of the Ecrins range.

Results

* Whole data set genetic characteristics

On the 1344 individuals genotyped with a set of 25 microsatellite markers, we retained 1258 individuals for further analysis. The 86 (6%) excluded individuals were removed because they failed to amplify or had too much missing data. On the remaining 1258 individuals, 1225 distinct multilocus genotypes were found. The few repeated multilocus genotypes were excluded from further analysis. The final dataset exhibited 4.5% missing data on average. All 25 microsatellite markers were highly polymorphic, with the number of alleles ranging from 5 to 25. The mean allele number across loci was 11.7. Global population genetics analyses showed a high genetic diversity when pooling individuals from all locations ($H_E = 0.57$) and Hardy-Weinberg equilibrium testing indicated an overall significant deficit in

Table 2: Characteristics of genetic groups estimated using four Bayesian clustering programs (STRUCTURE, GENELAND without spatial prior, TESS under the four modalities and NEWHYBRIDS). Individuals were assigned to a cluster when their membership was superior to 0.8. All F_{ST} values (level of differentiation between clusters 1 and 2) were significant ($P < 0.001$).

	STRUCTURE	GENELAND	TESS				NEWHYBRIDS
			Spatial Non-admixed	Non-spatial Non-admixed	Spatial Admixed	Non-spatial Admixed	
Cluster 1	0.12	0.55	0.55	0.59	0.69	0.82	0.65
Cluster 2	0.15	0.29	0.17	0.13	0.23	0.10	0.12
Unassigned	0.73	0.16	0.28	0.29	0.08	0.08	0.22
Hybrid cluster	-	-	-	-	-	-	0.01
F_{ST}	0.092	0.034	0.035	0.051	0.046	0.053	0.059

heterozygotes ($P < 0.05$). Moreover, linkage disequilibrium was detected for 50 out of the 300 pairs of loci ($Q < 0.05$). These two cues indicated some genetic structure may exist across the samples.

* Clustering results

All Bayesian clustering programs employed consistently partitioned genetic data into two groups (Table 1). Posterior log-likelihood of STRUCTURE runs showed a regular increase with increasing K , but the ΔK index computed according to Evanno *et al.* (2005) showed a clear mode for $K = 2$. GENELAND runs under the spatial model failed to consistently assess the number of genetic clusters, but using the non-spatial model all Markov Chains converged to $K = 2$ (Table 1). Last the four TESS modalities showed nearly equal value of DIC when increasing K (Table 1). A parsimonious reasoning would lead to consider $K = 2$ as the best number of clusters. In addition we observed that beyond $K = 2$, individual assignments amount to membership to only two clusters, which result from an intrinsic feature of the TESS algorithm to eliminate spurious (additional) clusters (Durand *et al.* 2009). Given the consistency across programs to partition genetic data into two groups we used NEWHYBRIDS in a last step to investigate further whether the unassigned/admixed individuals revealed by the first batch of clustering analyses result from hybridization events between individuals from both groups or not.

Membership profiles were overall similar across clustering algorithms (Fig. 2, Pearson rank test for significant difference, $P < 0.001$). However all individual membership values to the two genetic groups vary across models, and some individuals switch membership from one group to the other. When fixing a threshold of the membership coefficient $q > 0.8$ to assign individuals to a genetic group, the percentage of assigned individuals varied from 27 % to 92 % considering the program used. The NEWHYBRIDS analysis assigned 65 % of individuals but most of them were assigned to parental genetic groups and very few to the hybrid cluster (Table 2).

Genetic differentiation was always significant between the two genetic groups but with notable variation in F_{ST} values (from $F_{ST} = 0.092$ for STRUCTURE to $F_{ST} = 0.034$ for GENELAND), which may reflect stringent differences in membership estimates.

When applying hierarchical clustering, membership profiles gathered into three clusters (Fig 2). STRUCTURE and GENELAND grouped together and stood out from the two other programs' results. TESS outputs were very similar across three models (in particular between the two modalities under the admixture model), but the non-spatial/non-admixed

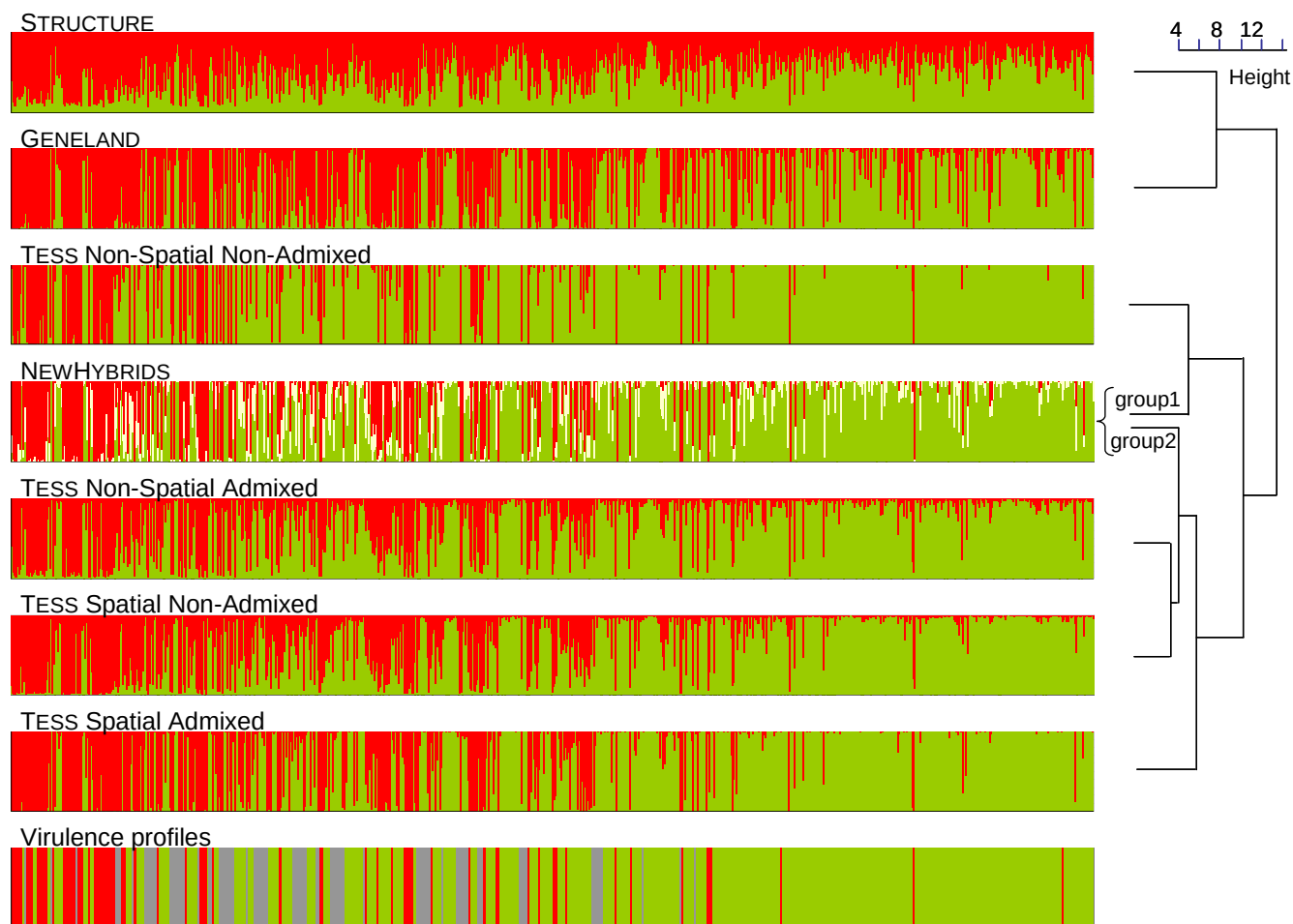


Figure 2: Comparison of the membership profiles resulting from the different Bayesian clustering models. On the left, each coloured figure depicts individual membership results according to a given model (which name is written above). Each thin vertical bar represents an individual whose genome is partitioned into up to 2 or 3 coloured segments (two genetic groups plus the hybrid genetic background detected by the NEWHYBRIDS algorithm). Segment lengths are proportional to the individual's membership coefficients (mean across five runs) in each genetic group (red: 'virulent' group 1; green: 'avirulent' group 2; pale yellow: hybrid background). Underneath is the histogram for individual virulence profile (red: Vir7 individual; green: Avr7 individual; grey: non-phenotyped individual). Individuals are ranked in the same order for all histograms (pooled by location and the locations distributed along a west-east gradient). On the right, tree diagram resulting from the hierarchical clustering procedure performed across all Bayesian assignment profiles (Euclidian distances between pairwise membership vectors). Two membership vectors were considered for NEWHYBRIDS (membership to group 1 and 1-membership to group 2), which branched at different positions in the tree.

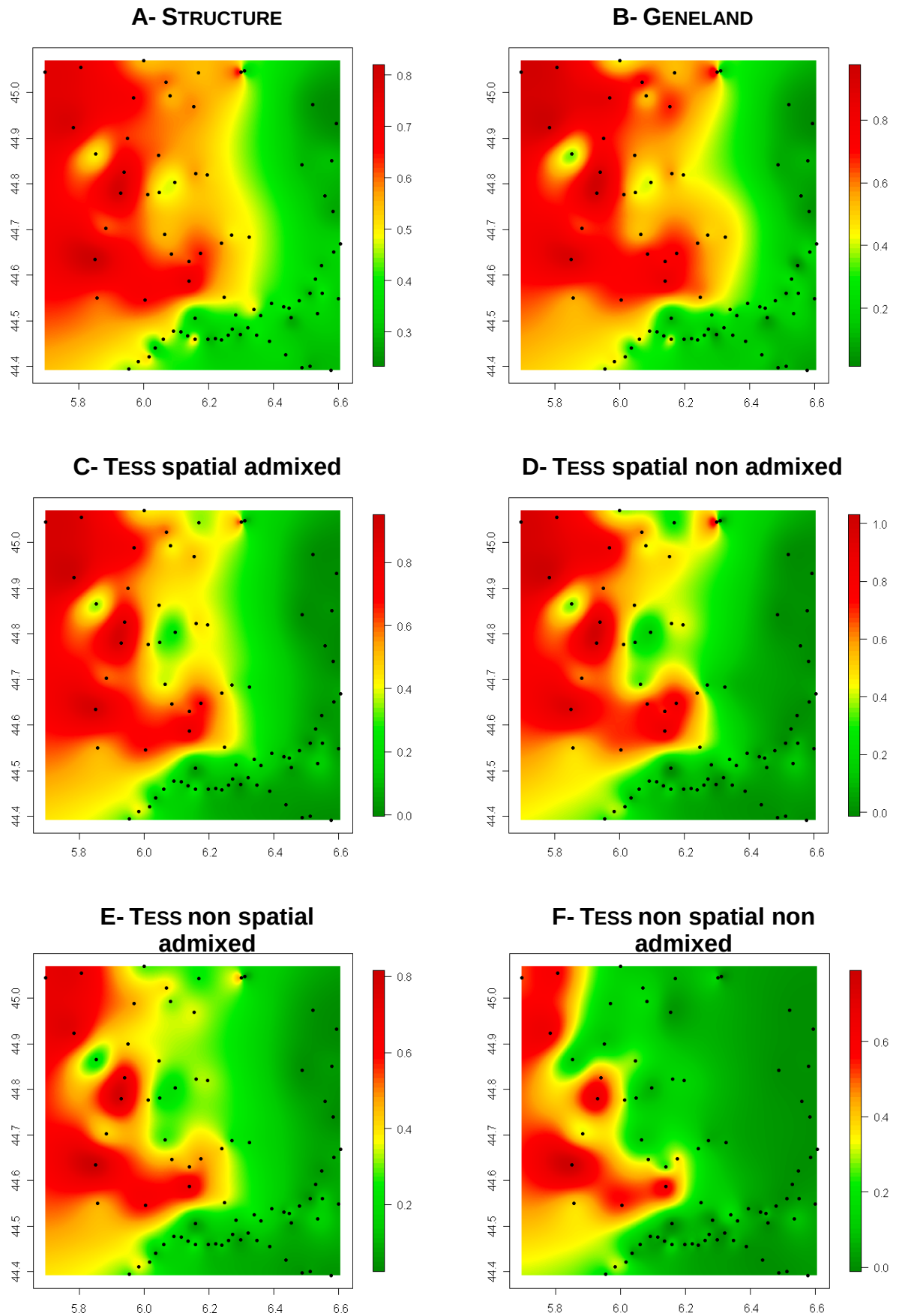
model of TESS resulted in quite different membership profiles and branched elsewhere. Interestingly, the membership profiles to the NEWHYBRIDS parental groups split, each being grouped to one of the two clusters formed by the TESS profiles. The results of NEWHYBRIDS seemed to be consensual over TESS disparities, maybe because the additional hybrid group absorbed the dissimilarities among other methods. In support of this hypothesis we observed a significant correlation ($P < 0.0001$) between individual hybrid membership in NEWHYBRIDS and the variance in group 1 membership across all other Bayesian clustering methods.

*** Spatial distribution of genetic groups**

Whatever the Bayesian clustering method employed, spatial analyses demonstrated that the genetic clusters were not randomly distributed over the sampling domain (Supplementary Figures A-F and Fig 3). Two algorithms (GENELAND and STRUCTURE) most clearly delineate genetic groups 1 and 2 and the boundary between the two groups outstandingly fit the relief barrier of the Ecrins range (Supplementary Figures A and B). The differences in the assignment profiles (Fig. 2) however translate into slight variations in the maps of local memberships to genetic groups 1 and 2. The most consistent pattern among spatial analyses is that individuals sampled on the east side of the Ecrins range (upstream of the Durance River and along the Ubaye River) always belonged to genetic group 2. Conversely, individuals sampled on the west side of the Ecrins range showed varying membership levels to genetic group 1 depending on the clustering method employed. As for the description of the assignment profiles, NEWHYBRIDS results offer a consensual view of the spatial distribution of genetic group 1 and 2 (Fig. 3). The disparity between the maps of local membership to genetic group 1 (mostly restricted to the western part of the domain) and local membership to genetic group 2 (that prevail along the Durance and Ubaye Rivers) results from an asymmetric distribution of hybrid signatures, which were only detected on the west side of the Ecrins range.

*** Phenotypic results and characteristics of the clusters**

The spatial distribution of VirZ individuals was non random (Fig 4A) and mirrored the distribution map of genetic group 1 (Fig. 3A). The within site frequency of VirZ individuals ranged from 0 to 1 with a mean proportion of 0.15. Locations from the west side of the Ecrins range exhibited the highest proportion of VirZ individuals but only two locations consisted exclusively of VirZ individuals. Locations sampled in the east side of the Ecrins range were



Supplementary Figure: Maps of local membership of the two genetic groups 1 and 2 after the different Bayesian clustering analyses: A, STRUCTURE; B, GENELAND; C, TESS spatial admixed model; D, TESS spatial non-admixed model; E, TESS non-spatial admixed model; F, TESS non-spatial non-admixed model. For each figure, local membership to the genetic group 1 is represented by green to red graded scale figuring low to high group 1 membership. Black dots represent sampled sites.

very poor in Vir7 individuals with 28 locations free from Vir7 individuals. A highly significant correlation was found between individual membership to group 1 assessed by NEWHYBRIDS and virulence profile (logistic regression, $P < 0.001$). Vir7 individuals represented 69% of individuals assigned to group 1, whereas almost all the individuals assigned to group 2 were Avr7 (Table 3). These two genetics groups will thus be referred to as the “cultivated” and “wild” groups.

The two genetic groups did not reveal major genetic differences and both have high values of F_{IS} but higher for the “wild” genetic group (Table 3). Linkage disequilibrium was found for only one population from the “wild” group and more populations did not respect HW equilibrium among this group.

*** Isolation by distance tests**

The test for isolation by distance (IBD) was significant when considering the whole dataset ($P < 0.0001$). However this significant isolation by distance does not result from a gradual and isotropic reduced dispersal. As highlighted by Fig. 4B a clear west-east gradient in genetic differentiation can be observed, in accordance with the spatial distribution of the two genetic clusters. When performing IBD tests within each cluster (defined after NEWHYBRIDS), the IBD signal disappeared within both “wild” and “cultivated” genetic groups. In between, on the western slopes of the Ecrins range, a contact zone with intermediate levels of genetic differentiation was outlined. This genetic gradient can be further illustrated when looking at the genetic differentiation along the main four rivers on both sides of the Ecrins range (Fig. 5). No significant genetic differences were found along the Durance and Ubaye Rivers (on the east side of the Ecrins range), whereas the genetic differentiation drastically increased along the Drac and Severaisse Rivers on the west side of the mountain range, which reveals the increase in admixture with individuals from the “cultivated” group. No significant difference was found among upstream locations across Rivers (on both sides of the Ecrins range), but the slight positive values -relative to the reference location of Prelles (Fig. 5)- may indicate some sub-structure, due to the geographic partitioning into different valleys.

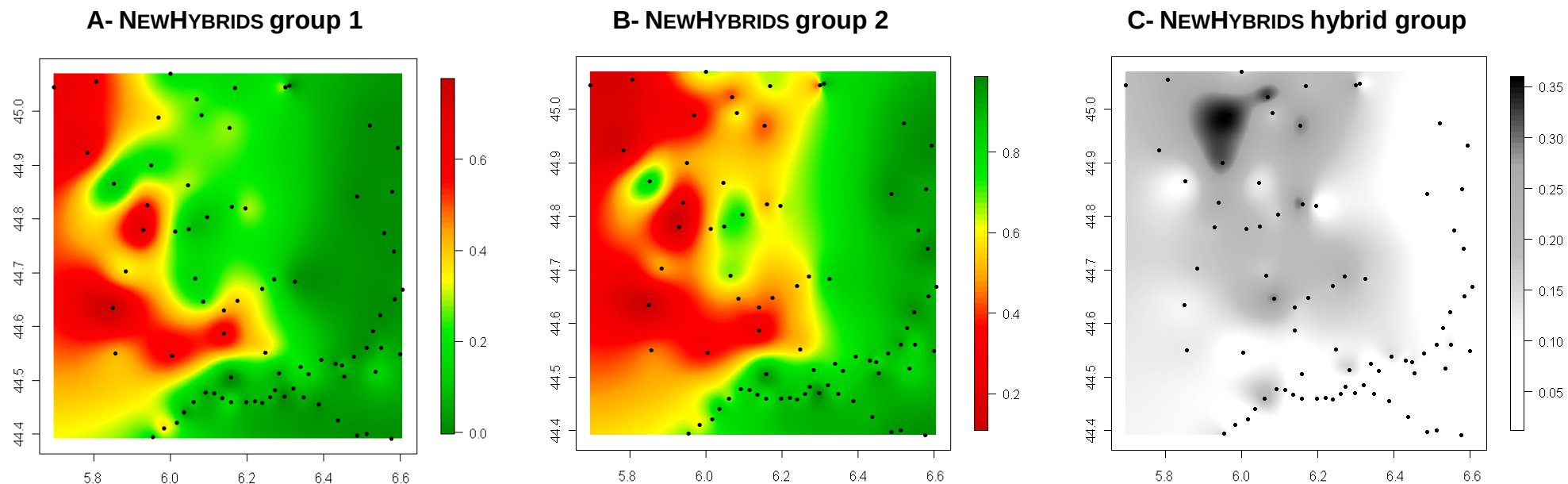


Figure 3: Maps of local membership after NEWHYBRIDS Bayesian clustering analysis interpolated over the spatial domain: A, membership to the genetic group 1 (green to red graded scale figuring low to high group 1 membership); B, membership to the genetic group 2 (red to green graded scale figuring low to high group 2 membership); C, level of hybrid background (white to dark grey graded scale figuring low to moderate hybrid membership). Black dots represent sampled sites.

Discussion

In the present study we investigated the spatial genetic structure of *M. larici-populina* individuals collected in several valleys on both sides of a mountain range. We provided evidence that this landscape feature acts as a barrier to gene flow since this high relief imposed a spatial segregation among *M. larici-populina* individuals. Two genetic groups were identified when using several different Bayesian clustering methods. Spatial interpolations of assignment results showed that individuals from the two genetic groups were spatially distributed on each side of the mountain ridge of the Ecrins range, which crossed the study area.

Most often landscape genetics studies focused on macro-organisms, such as mammals, and landscape genetics has seldom been used for microorganisms especially for those with long distance aerial dispersal. A recent landscape genetics study on the banana pathogen *Mycosphaerella fijiensis* evidenced a spatial distribution of two genetic groups, but spatial domains were not delineated by any obvious landscape feature (Rieux *et al.* 2011). Although in the present study, spatial domains of genetic groups matched with the mountain ridge of the Ecrins range.

The two genetic groups detected in this study have the same characteristics than the two main genetic groups detected in a previous study focusing on genetic composition of *M. larici-populina* at a larger scale (Xhaard *et al.* 2011). One group was mainly composed of *VirZ* individuals and the other mostly consisted of *AvrZ* individuals with a slightly higher genetic diversity. This previous study suggested that individuals from both “cultivated” and “wild” genetic groups seemed to be in contact in the French Alps (Xhaard *et al.* 2011).

In this study, we found that the spatial domains of each genetic group were split by the Ecrins range in the northern part of the study area. We used four Bayesian clustering programs with different options in order to get consistent results as recommended by François & Durand (2010). Bayesian clustering analyses were run with or without admixture and taking into account spatial information or not. All tests performed detected two genetic groups and a consistent spatial pattern: the “wild” group clearly predominated on the east side of the Ecrins range and the other group was restricted to the west side of this mountain chain. However the boundaries of spatial domains detected were not the same according to the program used due to differences in programs’ algorithms. For the “cultivated” group, the spatial domain either covered entirely the western slopes of the Ecrins range or was restricted

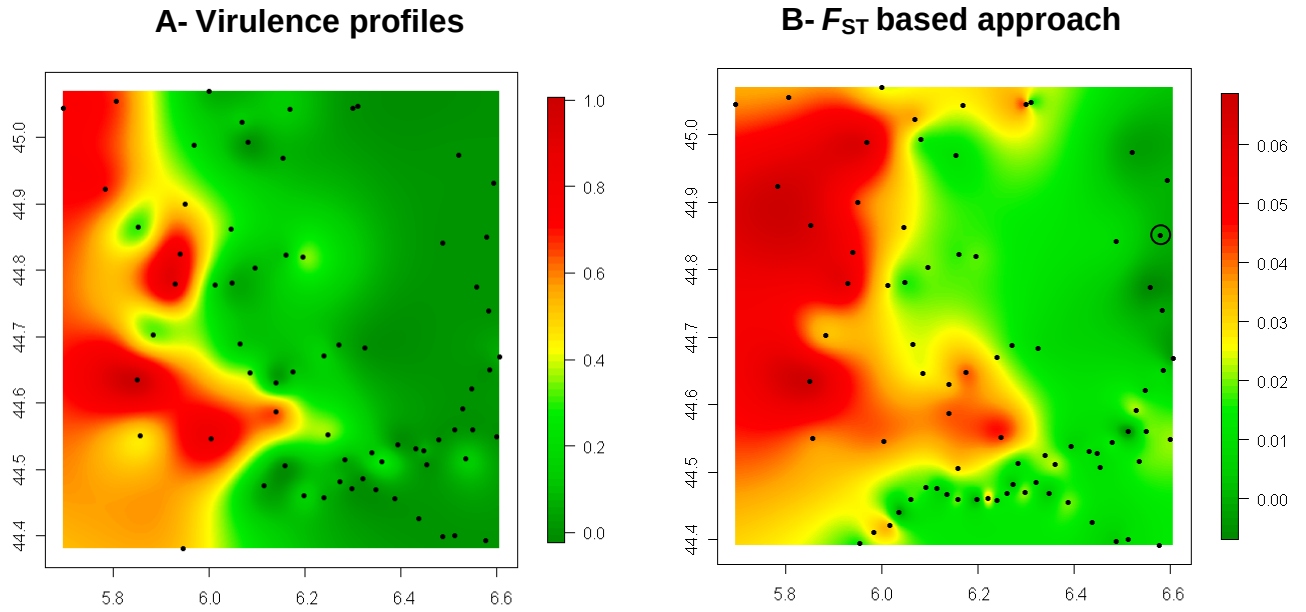


Figure 4: Interpolation maps of: A, Vir7 frequency (green to red graded scale figuring low to high local Vir7 frequency) and B, Map of the genetic differentiation ($a_r = F/(1-F)$ parameter) to Prelles location, highlighted by a black circle, interpolated over the spatial domain. Values are represented by a green to red graded scale figuring low to high local genetic differentiation. Black dots represent sampled sites.

to a smaller portion located western in the study area, downstream the Drac River valley. This area variation of the spatial domain, combined with the significant IBD signal detected from upstream to downstream the valleys in the western part of the Ecrins range indicated a gradual genetic variation consistent with a contact zone on the western slopes of the Ecrins range. There, according to NewHybrids analyses, individuals from both groups were admixed and we detected gene exchanges evidenced by hybrid background.

The rapid spread of *VirZ* individuals all over France in poplar plantations was already documented (Pinon *et al.* 1998). This invasion and population replacement by the “cultivated” group was recently confirmed in the wild poplar pool in the northern half of France by genetic analyses and Bayesian clustering methods (Xhaard *et al.* 2011). Although some *VirZ* individuals were detected in the Drac River valley as early as 1999 (Pinon & Frey, unpublished data), only a few *VirZ* individuals were detected in 2001 in some remote valleys in the French Alps (Gérard *et al.* 2006). In this study *VirZ* individuals were found mostly in the west side of the Ecrins range. The other genetic group, the “wild” group, seemed to be protected by the relief barrier from the invasion of *VirZ* individuals.

Interestingly, the maps of local membership obtained and the genetic gradient observed along the Drac River and the Séveraisse River on the west side of the Ecrins range strongly suggest a progressive eastward invasion dynamics of the “cultivated” group. This slow invasion dynamics contrasts with the rapid progression of *VirZ* individuals observed in non-mountainous regions all over France (Pinon *et al.* 1998). One hypothesis could be that the native “wild” group in the Alps represent a much bigger effective population size, due to the large poplar-larch sympatry area, compared to other regions of France. Indeed, most of the 100,000 ha of larch stands in France are concentrated in the alpine “Mélézin”, including the Ecrins range (Riou-Nivert 2001). This large sexual population of native *M. larici-populina* individuals could exert a high resilience to invasion by the “cultivated” group. Such resilience to invasion has already been shown in plants (Jager & Kowarik 2010) and animals (Hardouin *et al.* 2010).

In landscape genetics studies, different kinds of barriers can be evidenced, being classified as either natural (mountain range, river...) or anthropogenic (motorway) (Storfer *et al.* 2010). Here we proposed a new role of this mountain range, acting not as a classical barrier to gene flow, which generate spatial genetic structure through vicariance effect, but rather like a protection from genetic homogenization in impeding the spread of individuals during biological invasion. Further work is needed to test this hypothesis; in particular temporal sampling will be useful to assess the dynamic of *VirZ* individuals’ invasion. Another

Table 3: Characteristics of the whole data set and of the two genetic groups (using NEWHYBRIDS clustering results). N_{gen} , number of individuals genotyped; N_{path} , number of individuals phenotyped; $\text{Vir } 7$, proportion of Vir 7 individuals; A_r , allelic richness estimated for a minimal of sample size of 143; H_E , unbiased estimate of gene diversity (Nei 1978); H_O , observed heterozygosity; HW , number of sites that were at HW equilibrium; LD , mean number of pairs of significant linkage disequilibrium after correction for multiple tests; F_{IS} and F_{ST} values were calculated according to Weir and Cockerham (1984) taking into account the sampling location (within genetic groups).

	N_{gen}	N_{path}	$\text{Vir } 7$	A_r	H_E	H_O	HW	LD	F_{IS}	F_{ST}
Whole dataset	1225	1083	0.15	7.17	0.5692	0.4915	-	27/210	0.124	0.021
"Avirulent" (64 sites)	793	699	0.03	7.18	0.554	0.506	41/64	1/210	0.080	0.020
"Virulent" (10 sites)	143	127	0.69	5.91	0.494	0.462	9/10	0/210	0.054	0.034

approach will be to test for isolation by resistance, which takes into account the spatial heterogeneity, unlike the isolation by distance test (McRae 2006). To test for isolation by resistance different scenarios will be tested (as in Cushman *et al.* 2006), including the dynamic of the *Vir7* individuals expansion, the role of barrier of the Ecrins range and the heterogeneity of the study area.

REFERENCES

- Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217-1229.
- Antolin MF, Savage LT, Eisen RJ (2006) Landscape features influence genetic structure of black-tailed prairie dogs (*Cynomys ludovicianus*). *Landscape Ecology*, **21**, 867-875.
- Archie EA, Luikart G, Ezenwa VO (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends in Ecology & Evolution*, **24**, 21-30.
- Barrès B, Halkett F, Dutech C, Andrieux A, Pinon J, Frey P (2008) Genetic structure of the poplar rust fungus *Melampsora larici-populina*: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, **8**, 577-587.
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165-1188.
- Biek R, Real LA (2010) The landscape genetics of infectious disease emergence and spread. *Molecular Ecology*, **19**, 3515-3531.
- Blanchong JA, Samuel MD, Scribner KT, Weckworth BV, Langenberg JA, Filcek KB (2008) Landscape genetics and the spatial distribution of chronic wasting disease. *Biology Letters*, **4**, 130-133.
- Brown JKM, Hovmøller MS (2002) Aerial Dispersal of Pathogens on the Global and Continental Scales and Its Impact on Plant Disease. *Science*, **297**, 537-541.
- Chen C, Durand E, Forbes F, Francois O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747-756.

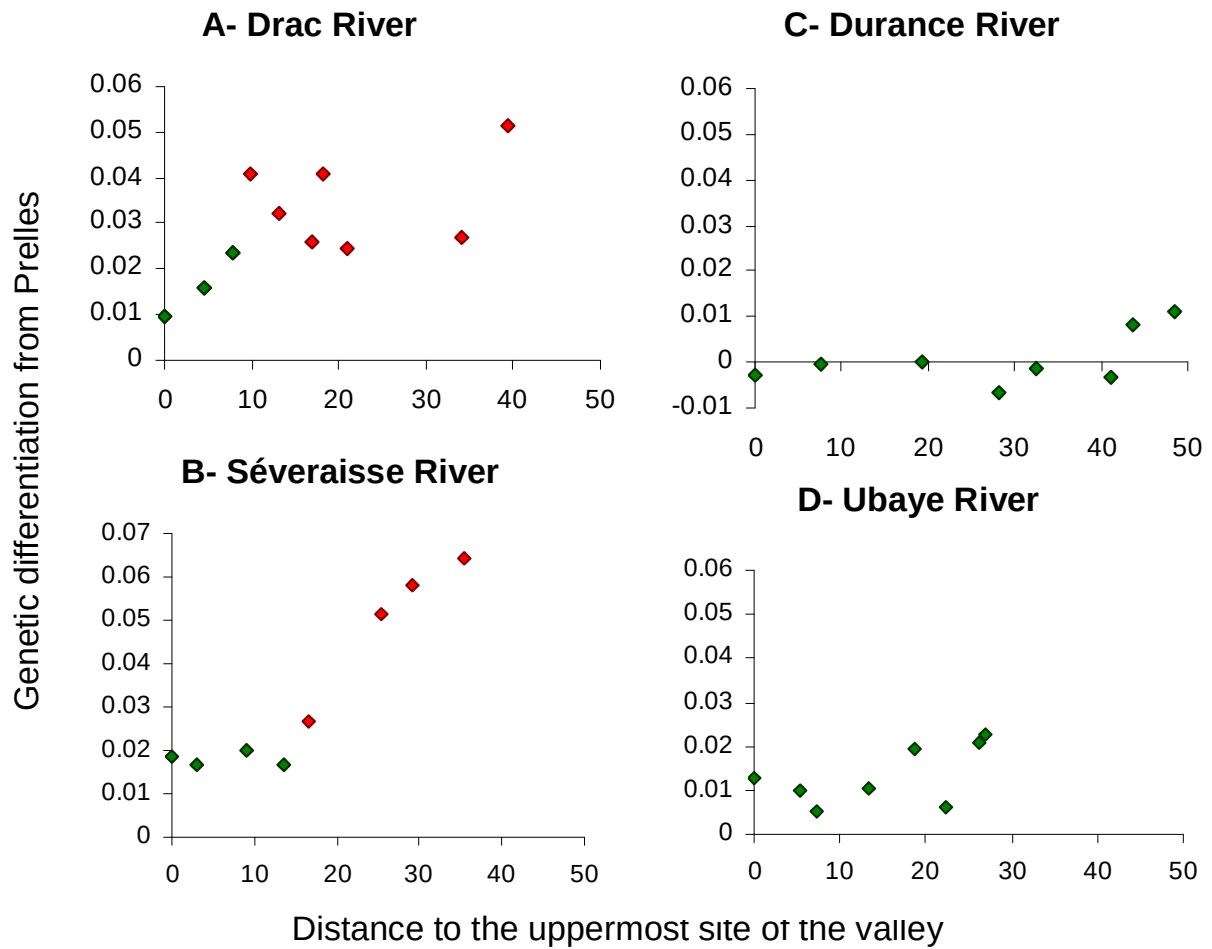


Figure 5: Genetic differentiation along the four main rivers: A, Drac River and B, Severaisse River (west side of the Ecrins range) and C, Durance River and D, Ubaye River (east side of the Ecrins range). For each figure, pairwise genetic differentiation estimates a_r between Prelles (reference location) and the sampled locations were plotted against the path distances (in kilometres) along the river from the uppermost location.

- Coulon A, Guillot G, Cosson JF, Angibault JMA, Aulagnier S, Cargnelutti B, Galan M, Hewison AJM (2006) Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology*, **15**, 1669-1679.
- Cushman SA, McKelvey KS, Hayden J, Schwartz MK (2006) Gene flow in complex landscapes: Testing multiple hypotheses with causal modeling. *American Naturalist*, **168**, 486-499.
- Duplessis S, Cuomo CA, Lin Y-C, *et al.* (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences*.
- Durand E, Jay F, Gaggiotti OE, Francois O (2009) Spatial Inference of Admixture Proportions and Secondary Contact Zones. *Molecular Biology and Evolution*, **26**, 1963-1973.
- 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.
- Finlay BJ (2002) Global dispersal of free-living microbial eukaryote species. *Science*, **296**, 1061-1063.
- Fontaine MC, Baird SJE, Piry S, Ray N, Tolley KA, Duke S, Birkun A, Ferreira M, Jauniaux T, Llavona A, Ozturk B, Ozturk AA, Ridoux V, Rogan E, Sequeira M, Siebert U, Vikingsson GA, Bouquegneau JM, Michaux JR (2007) Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in Old World waters. *Bmc Biology*, **5**, 16.
- François O, Durand E (2010) THE STATE OF THE FIELD: Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, **10**, 773-784.
- Frey P, Gérard PR, Feau N, Husson C, Pinon J (2005) Variability and population biology of *Melampsora* rusts on poplars. In Rust diseases of Willow and Poplar (eds. Pei MH, McCracken AR) pp. 63-72. CAB International, Wallingford, UK.
- Gérard PR, Husson C, Pinon J, Frey P (2006) Comparison of Genetic and Virulence Diversity of *Melampsora larici-populina* Populations on Wild and Cultivated Poplar and Influence of the Alternate Host. *Phytopathology*, **96**, 1027-1036.
- Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485-486.
- Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation--the correlated allele frequencies model revisited. *Bioinformatics*, **24**, 2222-2228.

- Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261-1280.
- Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712-715.
- Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology*, **18**, 4734-4756.
- Guillot G, Santos F (2009) A computer program to simulate multilocus genotype data with spatially autocorrelated allele frequencies. *Molecular Ecology Resources*, **9**, 1112-1120.
- Hardouin EA, Chapuis JL, Stevens MI, van Vuuren JB, Quillfeldt P, Scavetta RJ, Teschke M, Tautz D (2010) House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evolutionary Biology*, **10**, 325.
- Jager H, Kowarik I (2010) Resilience of Native Plant Community Following Manual Control of Invasive *Cinchona pubescens* in Galapagos. *Restoration Ecology*, **18**, 103-112.
- Kaeuffer R, Reale D, Coltman DW, Pontier D (2007) Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. *Heredity*, **99**, 374-380.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189-197.
- McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551-1561.
- Nagarajan S, Singh DV (1990) Long-distance dispersion of rust pathogens. *Annual Review of Phytopathology*, **28**, 139-153.
- Pinon J, Frey P, Husson C, Schipfer A (1998) Poplar rust (*Melampsora larici-populina*): The development of E4 pathotypes in France since 1994. Pages 57–64 in: Proceedings of the First IUFRO Rusts of Forest Trees Conference, Saariselkä, Finland. R Jalkanen, PE Crane, JA Walla, T Aalto, eds. Finnish Forest Research Institute, Research Paper 712.
- Pinon, J, Frey, P, (2005) Interactions between poplar clones and *Melampsora larici-populina* populations and their implications for breeding for durable resistance. In Rust diseases of Willow and Poplar (eds. Pei MH, McCracken AR) pp. 139–154. CAB International, Wallingford, UK.

- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.
- Rieux A, Halkett F, de Bellaire LD, Zapater MF, Rousset F, Ravigne V, Carlier J (2011) Inferences on pathogenic fungus population structures from microsatellite data: new insights from spatial genetics approaches. *Molecular Ecology*, **20**, 1661-1674.
- Riou-Nivert P (2001) Le mélèze. Institut pour le Développement Forestier, 144 pp.
- Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58-62.
- Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103-106.
- Sahlsten J, Thorngren H, Hoglund J (2008) Inference of hazel grouse population structure using multilocus data: a landscape genetic approach. *Heredity*, **101**, 475-482.
- Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9440-9445.
- Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP (2007) Putting the 'landscape' in landscape genetics. *Heredity*, **98**, 128-142.
- Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496-3514.
- Tuskan GA, DiFazio S, Jansson S, *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **313**, 1596-1604.
- Vaha JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63-72.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.
- Xhaard C, Andrieux A, Halkett F, Frey P (2009) Characterization of 41 microsatellite loci developed from the genome sequence of the poplar rust fungus, *Melampsora larici-populina*. *Conservation Genetics Resources*, **1**, 21-25.
- Xhaard C, Fabre B, Andrieux A, Gladieux P, Barrès B, Frey P, Halkett F (2011) The genetic structure of the plant pathogenic fungus *Melampsora larici-populina* on its wild host is extensively impacted by host domestication. *Molecular Ecology*, **20**, 000-000 (in press)

Chapitre 3

Mise en évidence de deux épidémies
simultanées de rouille du peuplier dans
la vallée de la Durance

Chapitre 3

Mise en évidence de deux épidémies simultanées de rouille du peuplier (*Melampsora larici-populina*) dans la vallée de la Durance

Dans le chapitre précédent, l'étude d'individus de *M. larici-populina* collectés dans différentes vallées de la zone de sympatrie peuplier - mélèze, incluant la partie amont de la vallée de la Durance, a permis de différencier deux groupes génétiquement distincts. Ces deux groupes d'individus présentent des répartitions spatiales différentes, séparées par la barrière des Ecrins. Le groupe composé d'individus au profil génétique de type 'sauvage' prédomine nettement à l'est de la barrière des Ecrins. A noter qu'en 2008, lors de cet échantillonnage, la partie amont de la vallée de la Durance ainsi que son affluent principal (l'Ubaye) étaient colonisées uniquement par des individus de type 'sauvage'. Les individus présentant un profil génétique de type 'cultivé' ou une appartenance mixte (avec un fond hybride) sont, à ces latitudes, tous cantonnés à l'ouest de la barrière des Ecrins (vallées du Drac et de la Séveraisse notamment). Cette délimitation claire entre les deux groupes génétiques dans la partie amont de la vallée de Durance se brouille plus au sud où plus aucune barrière naturelle ne limiterait la dispersion. En aval du lac de Serre-Ponçon, les individus 'cultivés' et 'sauvages' se retrouvent en mélange localement mais peu d'hybrides sont détectés, certainement à cause d'un éloignement plus grand des mélèzes.

La partie la plus au sud de la zone de sympatrie peuplier - mélèze débouche sur la vallée de la Durance, qui se caractérise par une ripisylve quasi-continue de peupliers noirs, propices à l'établissement de *M. larici-populina*, sur près de 200 km. Cette vallée, relativement encaissée dans sa partie la plus en amont, agit comme un corridor naturel, qui canalise la dispersion des spores de *M. larici-populina*.

Les objectifs de cette étude ont été de déterminer l'origine génétique d'individus collectés le long de la vallée à la fin de l'épidémie 2004. Nous avons cherché à savoir si les individus se mélangeaient et s'hybridaient entre eux ou si au contraire les groupes génétiques 'sauvages' et 'cultivés' se maintenaient même après une phase de dispersion épidémique.

Cette étude se présente sous la forme d'un article en préparation.

ARTICLE 4 : DISENTANGLING PATHOGEN ORIGINS FROM WILD AND CULTIVATED STANDS AT THE END OF AN EPIDEMIC

Authors: C. Xhaard¹, B. Barrès², A. Andrieux¹, F. Halkett^{1*} and P. Frey^{1*}

* These authors co-directed this work

Affiliations:

1- INRA, Nancy-Université, UMR 1136 Interactions Arbres - Microorganismes, F-54280 Champenoux, France.

2- INRA, UMR Biodiversité Gènes et Communautés, Domaine de Pierroton, F-33662 Cestas, France

Corresponding author:

Pascal Frey

Phone: 33 (0) 3 83 39 40 56

Email: pascal.freyp@nancy.inra.fr

Introduction

Tracking pathogen movement, identifying source populations and understanding environmental factors (including human activities) which influence pathogen spread are central tasks in disease ecology (Milgroom & Peever 2003; Ostfeld *et al.* 2005; Archie *et al.* 2009; Estoup & Guillemaud 2010). The advent of molecular biology has raised new opportunities in epidemiology. In particular, molecular markers greatly improved our ability to infer dispersal routes of pathogens, at least corroborating previous classical surveys but most often providing additional clues on epidemiological history of disease progression (Milgroom & Peever 2003; De Meeûs *et al.* 2007; Archie *et al.* 2009). Molecular epidemiology combines epidemiology and population genetics approaches, which is relatively easy to do because these two disciplines have similar research domains; for example both focus on discovering the inoculum source or detecting the population structure according to the vocabulary used (Milgroom 2001; Milgroom & Peever 2003).

Studies on historical spread of pathogens are often represented by phylogeography, which consists in reconstructing the historical biogeography and identifying major genetic

subdivision within species (Criscione *et al.* 2005). For pathogen species, one major aim is to study the co-evolution between parasites and their hosts (Criscione *et al.* 2005). Numerous studies have dealt with phylogeography of plant pathogens (Gomez Alpizar *et al.* 2007; Kauserud *et al.* 2007; Gladieux *et al.* 2008; Hovmøller *et al.* 2008; Llewellyn *et al.* 2009; Zaffarano *et al.* 2009) and generally used coalescent-based methods.

However, these studies provided little information on the spread of pathogens in the time course of an annual epidemic, as the phylogeographic dimension interested in very large temporal and spatial scales to decipher historical migration pathways. Few studies focused on more local scales, for which coalescent methods are not suited (Archie *et al.* 2009). At more restricted spatial scales other methods have to be used, for example Bayesian assignment methods, which are well suited for inferring the source population of parasites and quantifying gene flow between genetic groups (Archie *et al.* 2009). Assignment tests became tremendously popular (Kaeuffer *et al.* 2007), especially with the increasing availability of molecular markers such as microsatellite markers and high quality genetic data sets.

Here we interested in host-alternating (heteroecious) pathogens, which life cycle enables the starting point of epidemics to be more easily identified. These pathogens switch host species to complete their life cycle, performing their annual sexual reproduction stage on one host and then multiplying asexually on the second during the summer months. Asexual multiplication on the main host causes the epidemics. In such peculiar life cycle, the starting point of an annual epidemic is given by the location of the alternate hosts where sexual reproduction stage takes place, provided the geographical distribution of the alternate hosts is known and restricted (hosts clumped together). This is the case for rust fungi (Pucciniales) among which many heteroecious species exist (Zogg 1949, Bourassa *et al.* 2007). The poplar rust fungus *Melampsora larici-populina* infects two host types: poplar (*Populus* sp.), the main (or telial) host and larch (*Larix* sp.), the alternate (or aecial) host. Consequently the source location of each poplar rust epidemic is determined by the presence of larches, on which *M. larici-populina* reproduces sexually in spring. After its sexual reproduction, *M. larici-populina* spores disperse and infect poplar leaves, on which several stages of asexual multiplication occur, leading to rust epidemics. Then the fungus differentiates an overwintering form during winter and the life cycle restarts with infection on larches (Xhaard *et al.* 2011).

The native range of European larch (*Larix decidua*) encompasses the mountains of Central Europe (Alps, Carpathians, Sudetes, Tatras) and lowlands in Northern Poland. In France, the native range of larch is restricted to the “Mélézin” in the central Alps (Riou-Nivert

2001). In this area, wild poplar (*Populus nigra*) is also present in riparian stands in sympatry with larch. This study area where larches are not scattered allows a finest localization of the poplar rust inoculum source. Another advantage of this area is the features of the Durance River, which flows southwards from the poplar-larch sympatry zone. All along this river, an almost continuous riparian forest of *P. nigra* allows the dispersal of *M. larici-populina* from north to south. Hence the Durance River valley constitutes a one-dimensional ecological corridor where the epidemics of *M. larici-populina* is constrained, thus facilitating the monitoring of the course of the epidemics.

However, this simple ecological design is complicated by the presence of some cultivated poplar stands in both the sympatry and the allopatry areas. Many poplar cultivars carry qualitative resistance genes selected for their resistance to rust attack. *Melampsora larici-populina* broke down all the quantitative resistances deployed so far in poplar cultivation (Pinon & Frey 2005). Cultivated poplars exert a strong selective pressure on *M. larici-populina* individuals and only those carrying the virulence corresponding to a given resistance gene can infect this host (Gérard *et al.* 2006), according to the gene-for-gene relationship (Flor 1971). On the other hand, qualitative resistance has not been detected so far in *P. nigra* (Légionnet *et al.* 1999, Pinon & Frey 2005).

In a previous study we found that *M. larici-populina* populations were genetically structured in France (Xhaard *et al.* 2011, chapter 1). Three distinct genetic groups were found: two of them were represented by sexually reproducing individuals. The first group was comprised of individuals with a "cultivated" genetic profile and the second with "wild" genetic and phenotypic profiles. The third group was comprised of individuals which reproduce strictly asexually, and displayed very specific genetic signatures.

The "wild" and "cultivated" genetic groups were non-randomly spatially distributed. The distribution of the "cultivated" group matched the major poplar cultivation areas, mainly located in the northern half of France. Conversely, individuals from the "wild" group predominated in the southern half of France. We subsequently demonstrated that the Alps are a contact zone between the two major genetic groups (chapter 2). Individuals from both "wild" and "cultivated" groups were found in the northern part of the studied area (the poplar larch sympatry area), but the two groups are spatially segregated on both sides of a high mountain chain (the Ecrins range). We can thus suspect that two foci exist for the epidemic that spread each year along the Durance River. However no genetic study has been performed on the *M. larici-populina* individuals that colonize the riparian stands downstream the poplar larch sympatry area. Are they spatially segregated on cultivated vs wild stands or do they

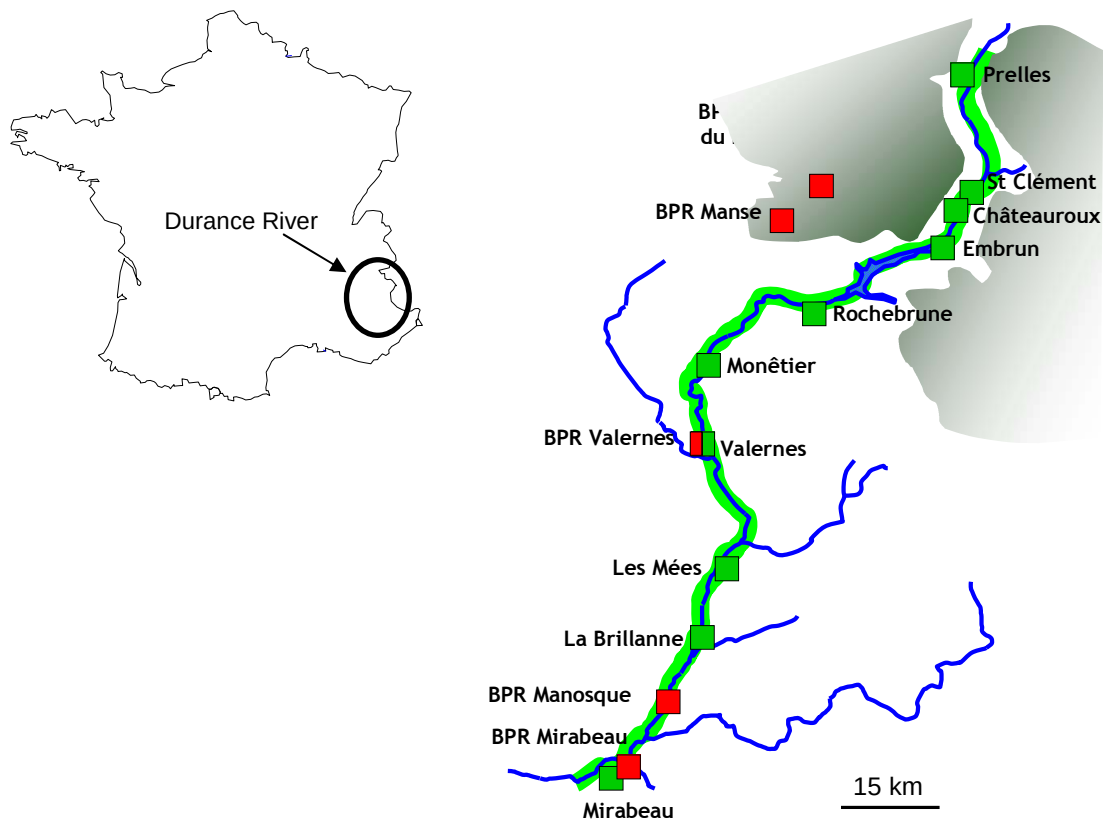


Figure 1: Schematic representation of the Durance River valley. The shadows in grey on the top right of the figure represents the poplar-larch sympatry area where the sexual reproduction of *M. larici-populina* takes place. Green and red boxes indicate the location of the wild and the cultivated stands sampled, respectively.

coexist locally? Is it possible to determine the genetic origin of the epidemic that developps along the Durance River?

Here we investigated the genetic profiles of *M. larici-populina* individuals collected along the Durance River valley at the end of the annual epidemic. Our aim was to assess the interactions between *M. larici-populina* individuals from wild and cultivated groups during an epidemic. We tested the hypothesis that two source populations contribute to the epidemic and assessed their level of admixture along the Durance River. Two extreme scenarii can be put forward. Either large gene flow occurs between wild and cultivated rust individuals, thus no genetic differentiation is expected among *M. larici-populina* individuals at the end of the epidemic. At the opposite, individuals from wild and cultivated hosts have two distinct epidemic dynamics and do not introgress each other. To disentangle these scenarii we used both selected (virulence) and neutral markers (26 microsatellites markers), and tested for a genetic partition of *M. larici-populina* individuals using several Bayesian assignment tests.

Materials and Methods

* The studied area: the Durance River valley

The sampling area was a ca. 200-km-long corridor in the Durance River valley in the French Alps (Fig. 1). The Durance River flows north to south and is bordered by a nearly continuous riparian forest of wild *P. nigra* from 0.5 to 2 km wide. We sampled 15 sites along or nearby the Durance River: 10 wild stands of *P. nigra* and 5 cultivated stands containing the hybrid poplar *P. x interamericana* ‘Beaupré’ on which individuals were sampled. Distances between contiguous sampled sites ranged from 5 to 32 km. Two sampled cultivated stands and four wild stands were located within the poplar-larch sympatry area, in the upstream part of the river (Fig. 1). Two other cultivated stands were located in the vicinity of sampled wild stands (Valernes and Mirabeau) in the allopatry area (without larch).

* Site monitoring and sampling details

The date of appearance of the first disease symptoms (rust pustules) was monitored for both wild and cultivated sites every three weeks between July and October 2004. Hence the date of first appearance of rust symptoms was known for each site. Since the incidence of rust is generally low in wild stands, individuals from the wild sites were all collected at the same time in October, at the end of the epidemic. The incidence of rust is much higher in cultivated

stands and can lead to early defoliation. Therefore, the five cultivated sites were sampled between July and October, depending on the local disease progress. Each individual was isolated from a distinct leaf collected on different trees and twigs to avoid re-sampling of clonal replicates. Infected leaves were collected on *P. nigra* in the wild sites and on *P. x interamericana* 'Beaupré' in the cultivated stands. Whenever possible, nearly 100 leaves were collected randomly for each wild site on up to 50 trees. The very focal infection at Valernes imposed to collect 30 leaves from only 5 trees. The infection level of *P. nigra* trees at Saint-Clément, Châteauroux, La Brillanne and Mirabeau was also very low, even at the end of the epidemic, and only 10, 18, 22 and 5 infected leaves were collected in these sites, respectively. In cultivated stands, about 30 infected leaves were collected from 20 to 30 trees. The high infection level in cultivated stands (which are never sprayed against rust) ensures an easy collection of infected leaves; except in Manosque where only 8 infected leaves were collected.

*** Individual isolation**

One single uredinium (sporulating lesion of *M. larici-populina* producing urediniospores) per leaf was selected randomly and grown on fresh leaf discs of *P. x euramericana* 'Robusta' as described by Barrès *et al.* (2008). Mono-uredinial isolation ensures that a single genotype per sampled leaf is isolated and multiplied, so that both virulence phenotype assessment and molecular characterisation can be performed on the same individual genotype.

*** Virulence profile assessment**

All *M. larici-populina* individuals were subjected to a second inoculation on two poplar cultivars, one susceptible (*P. x euramericana* 'Robusta') and the second carrying the R7 rust resistance gene (*P. x interamericana* 'Beaupré'), as described by Barrès *et al.* (2008). The success of the differential infection revealed the virulence phenotype, denoted Vir7 if both resistant and susceptible leaf discs were infected, and Avr7 if only susceptible (control) leaf discs were infected. After virulence identification, leaf discs of 'Robusta' were stored at -20°C until DNA extraction.

*** Microsatellite analysis**

When more than 50 individuals per sites were available, only a subset of 48 individuals was selected randomly for DNA analyses. Otherwise, all individuals were extracted for DNA (Table 1). DNA was extracted using DNeasy® 96 Plant kit (Qiagen) as described in Barrès *et al.* (2006). Microsatellite analyses were performed at 26 polymorphic loci: μ MLP 09, μ MLP12, μ MLP22, μ MLP27, μ MLP28, μ MLP30 and μ MLP31 (Barrès *et al.* 2006); MLP49, MLP50, MLP54, MLP55, MLP56, MLP57, MLP58, MLP61, MLP66, MLP68, MLP74, MLP77, MLP81, MLP82, MLP87, MLP91, MLP92, MLP96 and MLP97 (Xhaard *et al.* 2009). We followed the same PCR protocol as described in Xhaard *et al.* (2009) except that we used Qiagen® Multiplex PCR kit and the forward primers were labelled with three different dyes: D2, D3 and D4 (Proligo), according to the Beckman tagging technology. Three multiplex PCR were run, the first with 12 loci (D2: MLP49, MLP54, MLP56, MLP66 and MLP77; D3: MLP57, MLP61, MLP68 and MLP87; D4: MLP50, MLP55 and MLP97), the second with 6 loci (D2: μ MLP22 and μ MLP27; D3: μ MLP28 and μ MLP30; D4: μ MLP09 and μ MLP12) and the third with 8 loci (D2: MLP58, MLP74 and MLP82; D3: μ MLP31, MLP81 and MLP92; D4: MLP91 and MLP96). PCR products were separated, sized and analysed on a CEQ8000 Genetic Analysis System (Beckman Coulter).

*** Genetic data analysis**

Identification of repeated multilocus genotypes

Data were checked for misprint and scoring errors using MICRO-CHECKER software (Van Oosterhout *et al.* 2004). The identical genotypes were identified and we further tested if individuals with identical multilocus genotype were indeed true clones using MLGsim software (Stenberg *et al.* 2003). First the probability of observing a multilocus genotype n times in a population was computed. Second, using a Monte Carlo simulation method, the program determines the significance threshold for the probability values, indicating the multiple copies of the same multilocus genotype that did not occur by chance (true clone). Calculations took into account population size and allele frequencies.

Population structure and Bayesian assignment analysis

We first illustrated the genetic differentiation between individuals with a correspondence analysis performed using GENETIX (4.05) (Belkhir *et al.* 2004). We further tested for a genetic partition of the data set in performing Bayesian genetic assignment tests. We used three different softwares to check the consistency of the apportionment and the assignment results. First, we used STRUCTURE version 2.2 (Falush *et al.* 2003; Pritchard *et al.* 2000), a reference software for such analyses (Evanno *et al.* 2005). We used the admixture model with correlated allele frequency and default settings. Iteration parameters were set to a burning-in period of 30 000 iterations followed by 1,200 000 iterations. The likelihood of the assignments ($\ln P(D)$) was evaluated for K varying from 1 to 10. Three independent simulations were performed for each value of K to test for the consistency of the results. To assess the number of clusters K that best explain the partition of the genetic data, we computed the ΔK estimator as described by Evanno *et al.* (2005).

Second, we used GENELAND (version 3.1.4) (Guillot 2008; Guillot *et al.* 2005b), without including spatial information, to cross-validate our assignment procedure. The strength of this second method is that the best number of clusters, K , is inferred during a single run (not as STRUCTURE does) because of the specific Reverse Jump MCMC (Markov chain Monte Carlo) algorithm implemented, which enable the Markov chain to explore the whole range of putative cluster number at each step of the simulation. Hence, the number of clusters that best explain the data is given by the dimension value (here the value of K) in which the Markov chain spent most iterations (Guillot *et al.* 2005a, b). The value of K was allowed to vary from 1 to 10, and initialised at 8. Once the most likely K value was determined, additional runs were performed, with the corresponding value of K fixed, to better infer the assignment level of each individual to the K clusters, as advised in Guillot *et al.* (2005b). For the runs with varying K , we used the uncorrelated allele frequency model, and for those with fixed K we used the correlated allele frequency model according to Guillot (2008) with default settings, except for the number of nuclei, which was increased to 500 (in the non-spatial mode), given the number of individuals. Iteration parameters were set to 550 000 iterations with a burning-in period of 50 000 iterations.

Finally, we used NEWHYBRIDS (1.1) that implements a Bayesian genetic assignment approach dedicated to the quantification of hybridization events (Anderson & Thompson 2002). The detection of backcrosses requires much more loci (almost 50) than the 26 used here (Vaha & Primmer 2006). We therefore focused on the first generation hybrids between parental classes. The Bayesian method allows either loading prior knowledge on the allele

frequencies of the parental classes or estimating them from the offspring genetic data set. Here we used the program without any prior information. We used the following parameters: a short burn-in period of 1 000 iterations, following by more than 1 000 000 iterations. With the three assignment methods, three independent runs were performed to check for the consistency in the convergence of Markov chains.

Standard population genetics analysis

After assignment tests, standard population genetics analyses were carried out on each genetic group to further assess their characteristics. Expected and observed heterozygosities were calculated with GENETIX (4.05) (Belkhir *et al.* 2004). As the observed number of alleles in a sample is highly dependent on sample size, we calculated instead allelic richness (A_r) using FSTAT version 2.9.3 (Goudet, 1995). F -statistics were computed taking into account sample location (within each genetic groups) using FSTAT according to Weir & Cockerham (1984). Linkage disequilibrium between loci and deviations from Hardy-Weinberg (HW) expectations (heterozygote excess and deficit) were calculated using GENEPOP on the web (<http://genepop.curtin.edu.au/>) (Rousset 2008). In order to adjust the P values for multiple tests for linkage disequilibrium, we used the false discovery rate (FDR) procedure (Benjamini & Yekutieli 2001), which controls the proportion of significant results ('false discovery rate') instead of controlling the chance of making even a single error. The resulting adjusted P values are called Q values. This procedure is implemented in the R package QVALUE (Storey & Tibshirani 2003). At last we tested the correlation between genotypic assignment results and phenotypic information through a logistic regression between individual memberships and virulence profiles (Boolean data, 1 for Vir7 and 0 for Avr7 individuals).

Analysis of isolation by distance

Isolation by distance (IBD) was assessed within genetic groups (considering only locations with at least five assigned individuals) using the continuous method (Rousset 2000) of the 'Isolde' procedure implemented in GENEPOP (Rousset 2008, option "Analysis of isolation by distance between groups"). The significance of Mantel tests between pairwise genetic distances [analogue of $F_{ST}/(1-F_{ST})$ ratios] and geographical distances were performed using the ISOLDE program (10,000 permutations). For geographical distances we considered Euclidian distances between pairs of sampling sites. It is noteworthy that our aim here was not to infer dispersal distance from the slope of the regression line, as our biological system does not meet the requirements of the IBD model (Rousset *et al.* 1997). Instead, we just wanted to

Table 1: Characteristics of the *M. larici-populina* individuals analyzed, including the site of collection; the geographical coordinates, the host plant, the first date of infection (1 rust detected and 0 rust not detected) the number of individuals phenotyped and the number of individuals genotyped.

Collection site	Latitude	longitude	Host plant	Disease progression						No of individuals	
				7 Jul.	28 Jul.	18 Aug.	8 Sep.	29 Sep.	13 Oct.	phenotyped	genotyped
Prelles	44°51' N	6°34' E	<i>P. nigra</i>	1	1	1	1	1	1	103	48
Saint-Clément	44°39' N	6°35' E	<i>P. nigra</i>	0	0	0	1	1	1	10	10
Châteauroux	44°37' N	6°32' E	<i>P. nigra</i>	0	1	1	1	1	1	16	16
Embrun	44°33' N	6°30' E	<i>P. nigra</i>	0	0	1	1	1	1	102	48
Rochebrune	44°27' N	6°11' E	<i>P. nigra</i>	1	0	1	1	1	1	84	46
Monétier	44°22' N	5°56' E	<i>P. nigra</i>	0	0	1	1	1	1	97	47
Valernes	44°14' N	5°55' E	<i>P. nigra</i>	0	0	0	0	1	1	108	48
Les Mées	44°02' N	5°57' E	<i>P. nigra</i>	1	1	1	1	1	1	108	48
La Brillanne	43°55' N	5°54' E	<i>P. nigra</i>	0	0	0	0	1	1	24	24
Mirabeau	43°41' N	5°40' E	<i>P. nigra</i>	0	0	0	0	0	1	5	5
BPR-Pont-du-Fossé	44°40' N	6°14' E	Beaupré	1	1	1	1	1	1	19	30
BPR-Manse	44°37' N	6°08' E	Beaupré	0	0	1	1	1	1	27	32
BPR-Valernes	44°15' N	5°55' E	Beaupré	0	0	0	1	1	1	28	29
BPR-Manosque	43°48' N	5°49' E	Beaupré	0	0	0	1	1	1	8	9
BPR-Mirabeau	43°42' N	5°43' E	Beaupré	0	0	0	1	1	1	30	30

decipher whether the observed variation in pairwise genetic distance is gradual (according to geographical distances) or random.

Results

*** Timing of disease progression along the Durance River valley**

Poplar leaf infections were first detected -at the beginning of summer- in the upstream part of the valley where larches grow in sympatry with poplar trees (Fig.1 & Table 1). Then the other sites became infected according to their location downstream the valley; indicating that the epidemic begins in the upstream part of the valley and then spreads out downstream the river. This is particularly obvious for the wild sites. Nonetheless there were two exceptions to this regular spread, in Les Mées where the infection began very early, at the beginning of the survey period (7 July), and conversely we observed a late infection in Saint-Clément (8 Sep.) although it is located in the poplar-larch sympatry area. In the cultivated poplar stands, the epidemic spread also downstream the valley. Nonetheless the location of cultivated stands (two in the poplar-larch sympatry area in the north and the three remaining in the very south part of the valley) did not allow a proper quantification of disease progression (compared to the wild sites). Remarkably, there was a discrepancy in the relative timing of infection between wild and cultivated stands as the latter became infected earlier in the downstream part of the valley (Table 1).

*** Site by site virulence profiles**

All individuals collected from 'Beaupré' exhibited a Vir7 phenotype, which makes sense because otherwise they could not infect this resistant cultivar. For individuals sampled on *P. nigra*, the frequency of Vir7 ranged from 0% (Saint-Clément and Châteauroux) to 40% at Mirabeau (the southernmost site). Samples from the wild stands located in the upstream part of the valley (Prelles, Saint-Clément, Châteauroux and Embrun) exhibited a very low percentage (0 – 3%) of Vir7 individuals. In contrast, the proportion of Vir7 individuals was much greater in all wild sites downstream the Durance River valley, except in Les Mées.

*** Overall genetic and genotypic diversity**

Of the 478 individuals genotyped, only 8 were excluded from further analyses because they failed to amplify. The remaining 470 individuals exhibited only 0.6 % missing data on

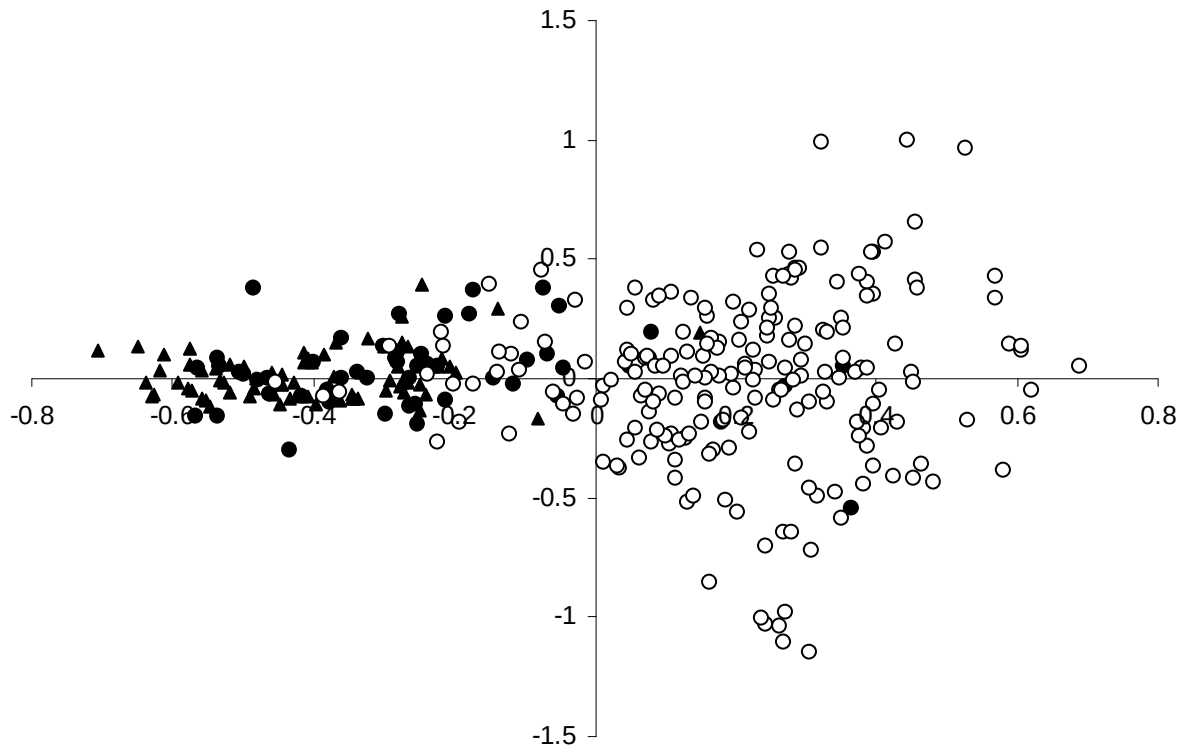


Figure 2: Correspondence analysis performed with GENETIX on the 331 unicopy genotypes of *M. larici-populina*. ▲ cultivated sites (Vir7), ○ wild sites (avr7) ● wild sites (Vir7).

Table 2: Inference of the number of cluster(s) (K) that best explain the genetic structure among samples (tests performed with a single copy per repeated genotype). For STRUCTURE runs we reported the mean value of log likelihood across simulations and the ΔK index computed according to Evanno *et al.* (2005). For GENELAND runs we reported the proportion of iterations spend by the RJMCMC across simulation (after the burning period).

K	STRUCTURE		GENELAND
	Log-likelihood	ΔK	
1	-16582		~0
2	-15856	382	~1
3	-15512	176	~0
4	-15343	8	~0
5	-15182	31	~0
6	-14990	43	~0
7	-14840	113	~0
8	-14803	174	~0
9	-14592	225	~0
10	-14607	14	~0

average. All 26 microsatellite markers were highly polymorphic (as in Xhaard *et al.* 2011), with the number of alleles ranging from 4 to 19 (mean number across loci: 7.11).

Of the 470 individuals analyzed, 331 distinct multilocus genotypes were found. All the repeated multilocus genotypes consisted of clonal copies according to the analysis performed using MLGsim (Stenberg *et al.* 2003). Repeated genotypes were found in both wild and cultivated stands. Three repeated genotypes were largely represented with 34, 30, and 20 clonal copies. The remaining 29 repeated multilocus genotypes occurred in two to ten clonal copies. Almost all repeated genotypes were restricted to a single site, except six repeated genotypes that were found at four distinct sites, with a maximal re-sampling distance of 126 km.

*** Genetic structure and assignment results**

Individual genotypes were largely separated according to their Vir7 or Avr7 profile in the correspondence analysis (Fig. 2). Most Vir7 individuals clustered, whatever their sampling origin (wild or cultivated stands). While the first axis differentiated individuals according to their Vir7/Avr7 profile, the second axis differentiated mostly among individuals collected on the wild sites. Thus, individuals from the cultivated stands were gathered, in contrast to those from wild stands, which were much more scattered. However, the first two axes only supported 5.8% of the total inertia. We therefore subsequently explored the relevance of such a genetic structure among sampled individuals through Bayesian assignment tests.

All Bayesian assignments tests were conducted including only one copy of repeated multilocus genotypes in order to minimize departure from the HW expectations on which assignment procedures were based. Whatever the method employed, Markov Chains converged rapidly and the three replicate runs gave consistent results. Both STRUCTURE and GENELAND algorithms agreed to sort individuals into two clusters (Table 2). However GENELAND seemed to perform better and gave clearer results: (i) Markov Chain for GENELAND runs rapidly stuck in $K = 2$, whereas likelihood analysis of STRUCTURE runs showed bimodal values for the best number of clusters and (ii) assignment profile was more clear cut for GENELAND (Fig. 3). Applying an assignment coefficient $q > 0.9$ lead to the exclusion of 93 (28%) and 22 (7%) individuals for STRUCTURE and GENELAND, respectively.

In a second step, we investigated further whether the non-assigned individuals revealed by STRUCTURE and GENELAND could result from hybridization events between

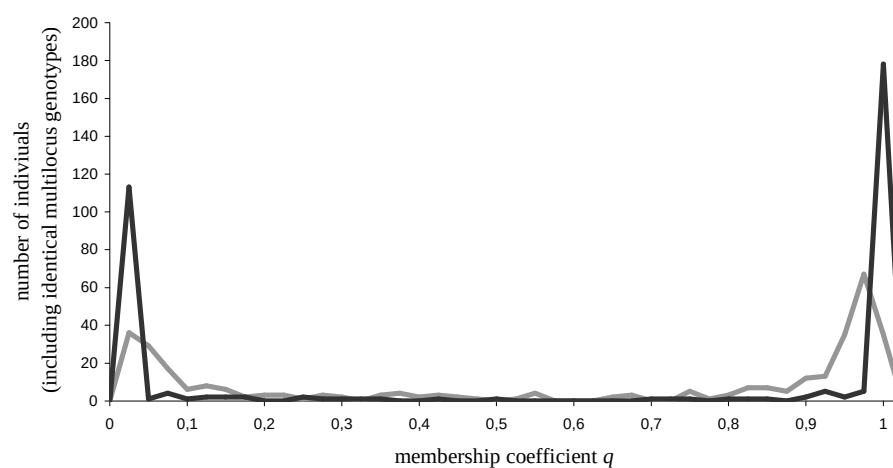


Figure 3: Distribution of the mean value of the membership coefficient (q) calculated over three simulation runs performed with GENELAND (black curve) and with STRUCTURE (grey curve).

individuals of the two clusters. Interestingly, the assignment profile obtained with NEWHYBRIDS was very close to the one obtained with GENELAND (data not shown). In particular, similar membership values were observed for individuals previously assigned (membership coefficient $q > 0.9$) to the two clusters. Among the previously non-assigned individuals, a single multilocus genotype (occurring twice in the sample) was undisputedly a hybrid when considering the same threshold of $q > 0.9$ (Fig. 4). The others non-assigned individuals showed various proportions of their genome attributed to the hybrid and parental clusters (Fig. 4).

In the following, we focussed on NEWHYBRIDS results to further unravel the biological property of the two genetic entities. Interestingly the distinction made according to the virulence phenotype roughly matched a segregation in individual memberships: the Vir7 individuals being mostly assigned to the first cluster and the Avr7 individuals to the second cluster (Fig. 4). Moreover, a highly significant correlation was found between individuals belonging to the first cluster and their virulence profile (logistic regression, $P < 0.001$).

Hence, nearly all individuals from the cultivated stands belonged to cluster 1, except one individual from BPR-Manse, which was assigned to cluster 2. Conversely the rust individuals sampled from wild stands were more admixed, with most individuals belonging to cluster 2 (those with a Avr7 phenotype profile), but a large proportion (mostly Vir7 individuals) was assigned to cluster 1, in particular in the sites located downstream the poplar – larch sympatry area (Fig. 4). A few individuals from the wild stands did not respect the correspondence between virulence phenotype and membership coefficient. These exceptions included the two Vir7 individuals from Embrun, which were assigned to cluster 2 and some Avr7 individuals: two individuals from Prelles and three individuals from Monétier, which were assigned to cluster 1. In addition, most unassigned individuals (including the hybrids) originated from the wild stands (which can reflect the more scattered distribution of genotypes in the correspondence analysis diagram (Fig. 2)).

*** Genetic characteristics and spatial repartition of individuals from the two genetic groups**

Given the consistency in the assignment results, between the virulence profiles (which denote the ability to infect cultivated poplars or not), and the wild vs. cultivated character of the sampled sites, cluster 1 and 2 were hereafter referred to as the ‘cultivated’ and the ‘wild’ genetic groups, respectively.

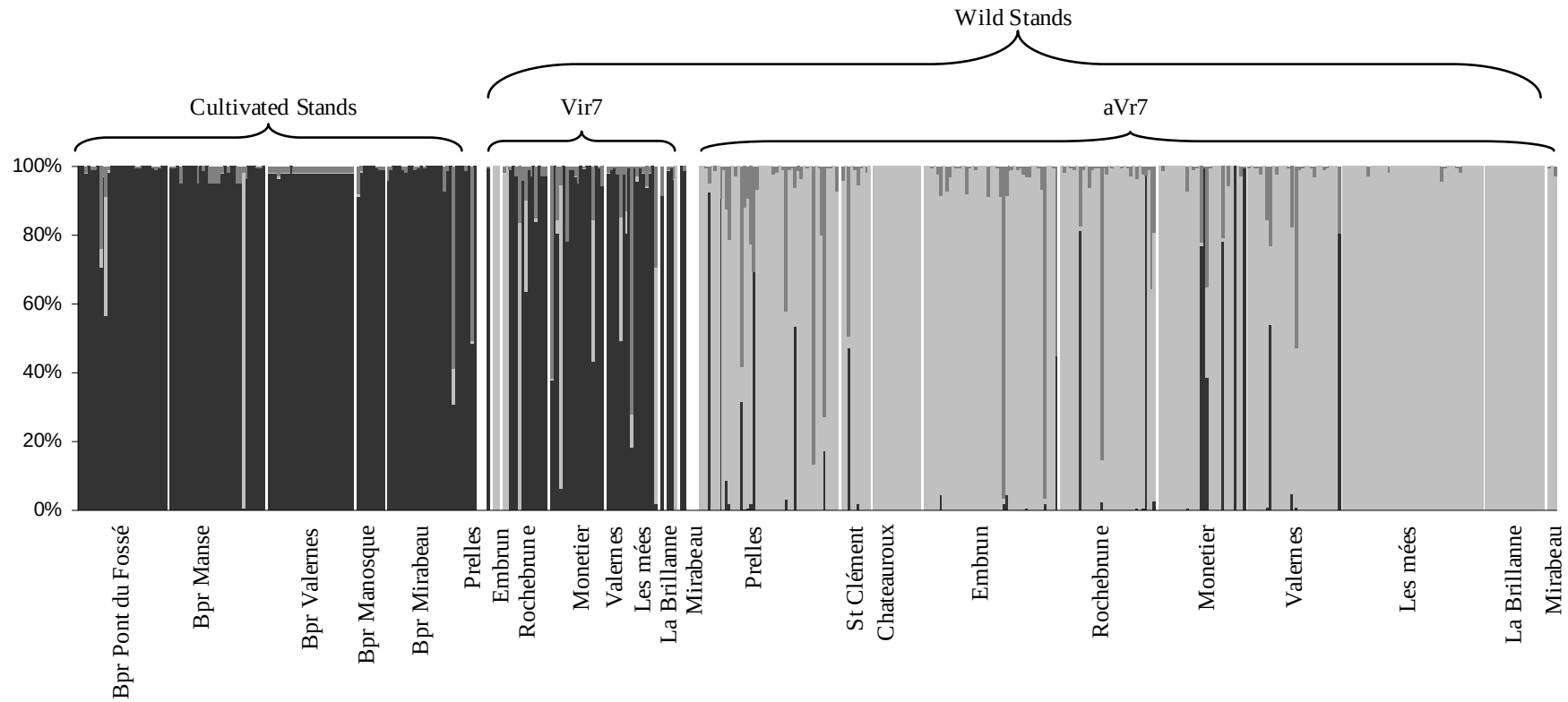


Figure 4: Assignment profile of *M. larici-populina* individuals using NEWHYBRIDS. The histogram includes all copies of repeated genotypes. Individuals were sorted according to their sampling sites and virulence profile. The genome proportion belonging to clusters 1, 2, and of hybrid origin is figured in black, light grey and dark grey, respectively.

Re-including clonal copies and applying an assignment coefficient threshold of $q < 0.9$, more than 93% of sampled individuals could be assigned, 251 belonging to the 'wild' group and 174 to the 'cultivated' group. There was a moderate, albeit significant, genetic differentiation level between the two groups ($F_{ST} = 0.094$, $P < 0.05$).

In accordance with previous results, a site by site examination of individual membership to the wild and cultivated genetic groups mirrors the spatial distribution of *Vir7* / *Avr7* individuals presented above (Fig. 5). In particular, it is obvious that individuals from the two genetic groups can be locally admixed in the wild stands –without detecting hybrids–, whereas the cultivated stands consisted almost exclusively in individuals from the 'cultivated' group. In addition, 'cultivated' individuals are not evenly distributed among the wild sites: sites located in the northern part of the valley (Prelles, Saint-Clément, Châteauroux, Embrun) were almost free from 'cultivated' individuals (just as they mostly consisted in *Avr7* individuals), but the frequency of 'cultivated' (and thus *Vir7*) individuals abruptly increased when considering sites located downstream the poplar-larch sympatry area (downstream of Rochebrune), reaching up to 40% of sampled individuals in the southernmost site (Mirabeau) (Fig. 5). The sole exception in this pattern concerned the site of Les Mées where no 'cultivated' individuals have been sampled.

Comparison of genetic characteristics of individuals from 'wild' and 'cultivated' groups revealed that –at several instances– individuals from the 'cultivated' group were less genetically diverse than those from the 'wild' group. First, values of all genetic diversity indices (observed and expected heterozygosity and allelic richness) were lower for individuals from the 'cultivated' group (Table 3). In addition rare alleles only occurred among 'wild' genotypes.

Second the 'cultivated' genetic group is more likely to form a panmictic unit (most sites are in HW equilibrium and only rare pairs of loci in linkage disequilibrium Table 3). Conversely, the 'wild' genetic group displayed more evidence for sub-structuring, undetected using Bayesian clustering methods, (larger global F_{ST} value, much more pair of loci in linkage disequilibrium, non negligible linkage disequilibrium as illustrated in Table 3). In addition the variance in genetic and genotypic diversity indices is much greater across location for the 'wild' group (Table 3).

Tests for isolation by distance conducted on 'wild' and on 'cultivated' groups were in accordance with these results. For individuals from the 'cultivated' group no isolation by distance was detected ($P = 0.85$). However a significant positive correlation ($P = 0.02$) was detected between the $F_{ST} / (1 - F_{ST})$ values and geographical distances for individuals from the

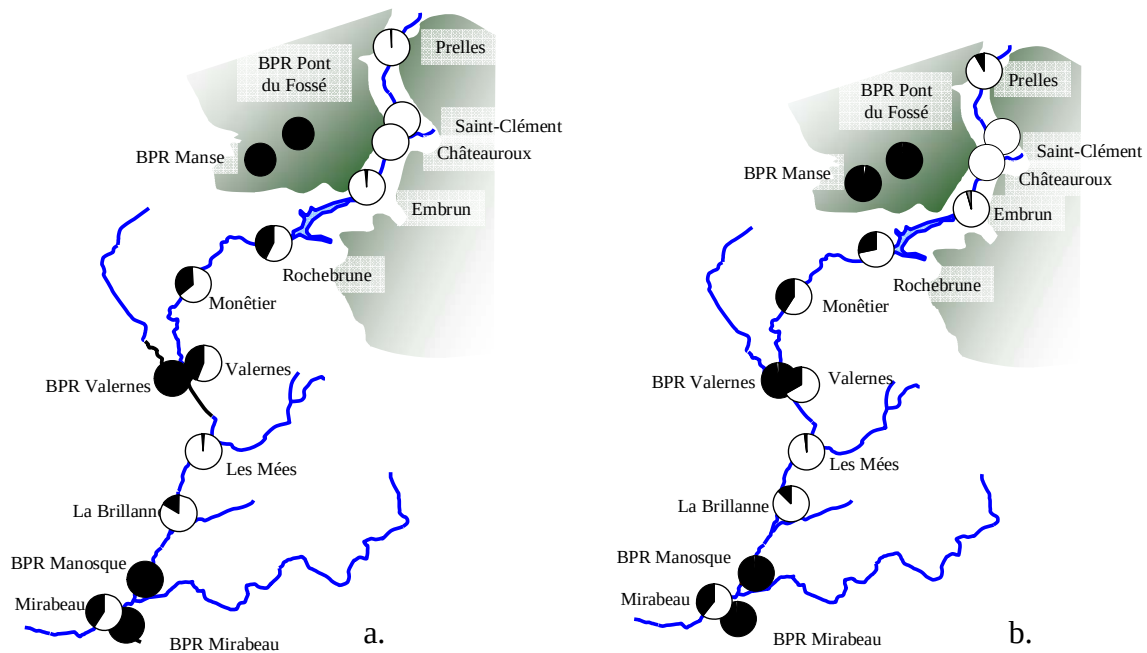


Figure 5: a; Geographical distribution of phenotypic profiles showing proportion of Avr7 (white) and Vir7 (black) *M. larici-populina* individuals at each site. b; Geographical distribution of membership of the genetic groups, showing proportion of individuals assigned to the wild cluster (white), the cultivated cluster (black) and the hybrid cluster (grey) at each site, following the result of NEWHYBRID.

Table 3: Characteristics of the two genetic groups expressed as the means and ranges (in brackets) of several indices across sampling sites. Only sites with at least 9 distinct genotypes were retained. N_{tot} , number of individuals assigned to each genetic groups; G/N , genotypic diversity index; H_E , expected heterozygosity; H_O , observed heterozygosity; HW; number of sites at Hardy-Weinberg (HW) equilibrium; A_r , allelic richness; LD, number of pairs of significant linkage disequilibrium after correction for multiple tests; F -statistics (F_{ST} and F_{IS}) calculated according to Weir and Cockerham (1984).

	N_{tot} (range)	G/N (range)	A_r (range)	H_E (range)	H_O (range)	HW	LD (range)	F_{IS}	F_{ST}
'Wild' (8 sites)	28.25 (9-47)	0.83 (0.23-1)	3.02 (2.17-3.43)	0.45 (0.36-0.49)	0.46 (0.38-0.50)	3	97.75/325 (5-305)	0.069	0.047
'Cultivated' (7 sites)	19.29 (9-31)	0.78 (0.45-1)	2.64 (2.46-2.80)	0.41 (0.38-0.43)	0.38 (0.35-0.40)	5	39.7/325 (2-72)	0.054	0.035

'wild' group, a result which was consistent with the sub-structure evidenced above. In addition, 6 out of 21 comparisons of pairwise F_{ST} between 'cultivated' sites were significant ($P < 0.05$) whereas for 'wild' sites 21 out of 28 comparisons were significant ($P < 0.05$).

Discussion

In this study we demonstrated that an epidemic of poplar rust occurred along the Durance River valley, which was initiated by two genetically different groups of *M. larici-populina* individuals that admixed locally downstream the valley, but did not hybridize.

The first result of this study was the demonstration of the spread of the rust epidemic along the Durance River valley. Monitoring the spatio-temporal occurrence of *M. larici-populina*, we showed that the epidemic began in the upstream part of the valley where larches and poplars grow in sympatry and then spread downstream to the allopatry area. This regular southward spread of the disease was assessed through count data recorded in the time course of the epidemic. In a similar manner, Zogg (1949) described the spatio-temporal dynamics of an annual maize rust epidemic caused by *Puccinia sorghi* in the higher Rhine River valley in Switzerland, over an 80 km-long transect. The secondary (aecial) host of this heteroecious rust fungus, *Oxalis stricta*, was restricted to a small area in the downstream (northern) part of the studied transect, near the Constance Lake. In 1945-1947, the annual maize rust epidemics were shown to begin in this Maize – *Oxalis* sympatry area and then to spread regularly southwards to the upstream part of the valley. Thus, in some heteroecious rust fungi, the precise location of the sympatry area between the primary (telial) and secondary (aecial) hosts is of great interest because it allows the knowledge of the starting point of the annual epidemics. This is also true at larger geographic scales. In a continental study of the genetic diversity of the North American poplar rust fungus *Melampsora medusae* f. sp. *deltoidae*, Bourassa *et al.* (2007) have shown that the genetic diversity was higher in the poplar – larch sympatry area (Great Lakes Region, Ontario and Quebec) than in the area with poplars only (most of eastern United States). The genetic features of the poplar rust individuals collected in the allopatry area suggested that they resulted from annual asexual spread of individuals originating from sexual populations in the poplar – larch sympatry area (Bourassa *et al.* 2007).

In this study, the population genetic analysis of individuals provided additional insights onto the genetic origin of this epidemic. We found that *M. larici-populina* individuals –collected after disease spread– clustered into two genetically distinct groups. This genetic partition amounts to the two sexual groups examined at a wider scale in chapter 1 (Xhaard *et*

al., 2011) and in the poplar-larch sympatry area, upstream the Durance River, in chapter 2. The contrasted phenotypic and genotypic characteristics of these two groups are consistent across studies: a majority of individuals with a Vir7 phenotypic profile and less genetic diversity in the "cultivated" group *versus* almost only Avr7 individuals and a higher genetic diversity in the "wild" group.

Here we observed a quasi perfect match between phenotypic profiles and assignments to the genetic groups. Exceptions to this rule can include some typing errors during phenotype assessment. However it is much unlikely that these mismatches would result from the assignment procedures. The distinction between the two genetic groups was highly robust and supported by the two Bayesian algorithms we used (STRUCTURE and GENELAND) to assess the number of genetic clusters among the samples. Assignment profiles were very similar, although GENELAND appeared to be more efficient than the reference software STRUCTURE at assigning individuals to the genetic groups. In addition to consistently assign individuals, the use of the third assignment software (NEWHYBRIDS) evidenced some hybridization events between the "wild" and "cultivated" groups. Such evidence for effective gene flow can provide a third explanation for the mismatches between phenotypic and genotypic profiles: genetic determinants of the Vir7 phenotype may have introgressed some individuals with a "wild" genetic background and *vice versa*.

Overall, very few hybrids were found, although at many instances individuals from the two genetic groups were gathered at the same location, especially in the southernmost part of the valley. This lack of hybrids may appear puzzling, but has to be interpreted in the light of the fungal life cycle and the specific features of the Durance River valley. The southernmost sites of the valley are located far away from the poplar-larch sympatry area where the sexual reproduction takes place. As evidenced in Bourrassa *et al.* (2007), the many asexual cycles which results in the disease spread on the primary host is a dead end process in the absence of a secondary (aecial) host in the vicinity. Along the Durance River, individuals collected in the southern sites are very unlikely to contribute to the sexual reproduction upstream the valley.

In order to understand the lack of gene flow between 'cultivated' and 'wild' groups, one has to focus on the poplar-larch sympatry area. There we observe a clear split in the spatial distribution of the two genetic entities, with no individual from the "wild" group on the western part of the sympatry area and a quasi absence of individuals from the "cultivated" group on the eastern part (that is upstream the Durance River). This spatial segregation is in accordance with the results of a previous landscape genetic study (chapter 2), which showed that a mountain chain (the Ecrins range) act as a barrier protecting the "wild" individuals on

the eastern slopes from introgression by the “cultivated” individuals invading this area from the northwest. At a four years interval, analysis of samples collected upstream the Durance River confirmed the near absence of “cultivated” individuals there, which hence account for the very few hybrids detected.

The population genetic analyses performed in this study thus clearly identified that the epidemic which spread along the Durance River originates from at least two inoculum sources located on both sides of the Ecrins range. Individuals of each “cultivated” or “wild” group then disperse asexually southwards along the Durance River, giving rise to two simultaneous, but genetically distinct, poplar rust epidemics. In the allopatry area (south of the genetic barrier formed by the Ecrins range) individuals belonging to both genetic groups mixed on the wild *P. nigra* stands. Conversely, the cultivated stands, even if located in the downstream part of the valley, remained free from individuals with a “wild” origin, because almost none of them were able to infect cultivated poplars carrying the resistance R7.

Epidemiological data allowed us to document a very regular spread of the disease on both wild and cultivated stands. Nonetheless, the timing differed between the two types of stands, with an earlier record of first disease symptoms in cultivated stands for the southern locations, certainly because cultivated poplars enable a faster multiplication of the fungus (Barrès *et al.* in prep). It could be interesting to further assess with a dedicated spatio-temporal sampling scheme, whether individuals from the “cultivated” group (which can hence multiply faster) are indeed the first to colonize the adjacent wild stands. Anyway, the lower number of cultivated stands did not allow us to monitor the spatio-temporal dynamics of the disease as precisely as along the wild stands.

Focussing on the colonization of the Durance River, we denoted some irregularities in the spread of the disease. Some sites were infected too late or too early according to this dispersal pattern. For example, although Saint-Clément is located in the poplar – larch sympatry area, the first rusted poplar leaves were found quite late (8 September). This case of late infection may be explained by local microclimatic conditions unfavourable to the disease. Another exception to this regular pattern of colonization is Les Mées, but there, the first rust symptoms were recorded very early (7 July), although this site is located far in the south from the sympatry area. This case of early infection may be due to a long distance dispersal event, as poplar rust is known for its long distance dispersal capacities (Wilkinson & Spiers 1976, Barrès *et al.* 2008). But a careful examination of the genotypes found at Les Mées revealed that they almost all belonged to the same repeated multilocus genotype and that this genotype was the same as that of the individuals from the “asexual” group identified in chapter 1

(Xhaard *et al.* 2011). These individuals have not been sorted out as a third genetic group through assignment tests certainly because they were too few (as we moreover retained only one copy of each repeated genotype to perform these tests). These “asexual” individuals were assigned to the “wild” group, which may indicate some genetic kinship, but more likely result from the “wild” group being the more genetically diverse. The early infection found in Les Mées in 2004 could therefore result from asexual overwintering of individuals belonging to the “asexual” group (Xhaard *et al.* 2011). Several authors have already suggested asexual overwintering in poplar rusts (Dupias 1965, Taris 1968, Walker *et al.* 1974, Wilkinson & Spiers 1976, Pinon 1980), but the survival mechanisms still need further investigations.

In addition to deciphering the genetic origins of *M. larici-populina* individuals collected at the end of a poplar rust epidemic, this population genetic analysis also gives the opportunity to confirm the gradual dispersal process responsible for the spread of the disease along the Durance River. A significant correlation between genetic and geographical distances revealed an isolation by distance pattern among individuals from the “wild” group. This spatial structure is consistent with a gradual dispersal pattern (Rousset, 1997, Leblois *et al.* 2004) resulting from the colonization dynamics we observed (Halkett *et al.* 2010). This result highlights the genetic consequences of dispersal which distort allele frequencies along colonization waves. Since we can clearly identify the source of the populations and the path followed by the epidemic wave, the Durance River valley provides us with a unique ecological situation to examine further the genetic consequences of dispersal in combining population genetics and epidemiological data in a spatio-temporal survey.

REFERENCES

- Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217-1229.
- Archie EA, Luikart G, Ezenwa VO (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. *Trends in Ecology & Evolution*, **24**, 21-30.
- Barrès B, Dutech C, Andrieux A, Caron H, Pinon J, Frey P (2006) Isolation and characterization of 15 microsatellite loci in the poplar rust fungus, *Melampsora larici-populina*, and cross-amplification in related species. *Molecular Ecology Notes*, **6**, 60-64.

- Barrès B, Halkett F, Dutech C, Andrieux A, Pinon J, Frey P (2008) Genetic structure of the poplar rust fungus *Melampsora larici-populina*: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, **8**, 577-587.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, 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).
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165-1188.
- Bourassa M, Bernier L, Hamelin RC (2007) Genetic diversity in poplar leaf rust (*Melampsora medusae* f. *sp deltoidea*) in the zones of host sympatry and allopatry. *Phytopathology* **97**, 603-610.
- Criscione CD, Poulin R, Blouin MS (2005) Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Molecular Ecology*, **14**, 2247-2257.
- de Meeûs T, McCoy KD, Prugnolle F, Chevillon C, Durand P, Hurtrez-Bousses S, Renaud F (2007) Population genetics and molecular epidemiology or how to "debusquer la bete". *Infection Genetics and Evolution*, **7**, 308-332.
- Dupias G (1965) Les rouilles des peupliers dans les Pyrénées et le bassin sous-pyrénéen. *Bulletin de la Société Mycologique Française* **81**, 188-196.
- Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology*, **19**, 4113-4130.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**, 1567-1587.
- Flor HH (1971) Current Status of the Gene-For-Gene Concept. *Annual Review of Phytopathology*, **9**, 275-296.
- Frey P, Gérard PR, Feau N, Husson C, Pinon J (2005) Variability and population biology of *Melampsora* rusts on poplars. In *Rust diseases of Willow and Poplar* (eds. Pei MH, McCracken AR) pp. 63-72. CAB International, Wallingford, UK.

- Gérard PR, Husson C, Pinon J, Frey P (2006) Comparison of Genetic and Virulence Diversity of *Melampsora larici-populina* Populations on Wild and Cultivated Poplar and Influence of the Alternate Host. *Phytopathology*, **96**, 1027-1036.
- Gladieux P, Zhang XG, Afoufa-Bastien D, Sanhueza RMV, Sbaghi M, Le Cam B (2008) On the Origin and Spread of the Scab Disease of Apple: Out of Central Asia. *Plos One*, **3**, e1455.
- Gomez-Alpizar L, Carbone I, Ristaino JB (2007) An Andean origin of *Phytophthora infestans* inferred from mitochondrial and nuclear gene genealogies. *Proceedings of the National Academy of Sciences*, **104**, 3306-3311.
- Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485-486.
- Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation--the correlated allele frequencies model revisited. *Bioinformatics*, **24**, 2222-2228.
- Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261-1280.
- Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712-715.
- Halkett F, Coste D, Platero GGR, Zapater MF, Abadie C, Carlier J (2010) Genetic discontinuities and disequilibria in recently established populations of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Molecular Ecology*, **19**, 3909-3923.
- Hovmoller MS, Yahyaoui AH, Milus EA, Justesen AF (2008) Rapid global spread of two aggressive strains of a wheat rust fungus. *Molecular Ecology*, **17**, 3818-3826.
- Kaeuffer R, Reale D, Coltman DW, Pontier D (2007) Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. *Heredity*, **99**, 374-380.
- Kausserud H, Svegarden IB, Saetre GP, Knudsen H, Stensrud O, Schmidt O, Doi S, Sugiyama T, Hogberg N (2007) Asian origin and rapid global spread of the destructive dry rot fungus *Serpula lacrymans*. *Molecular Ecology*, **16**, 3350-3360.
- Leblois R, Rousset F, Estoup A (2004) Influence of Spatial and Temporal Heterogeneities on the Estimation of Demographic Parameters in a Continuous Population Using Individual Microsatellite Data. *Genetics*, **166**, 1081-1092.

- Legionnet A, Muranty H, Lefevre F (1999) Genetic variation of the riparian pioneer tree species *Populus nigra*. II. Variation In susceptibility to the foliar rust *Melampsora larici-populina*. *Heredity*, **82**, 318-327.
- Llewellyn MS, Lewis MD, Acosta N, Yeo M, Carrasco HJ, Segovia M, Vargas J, Torrico F, Miles MA, Gaunt MW (2009) Trypanosoma cruzi IIC: Phylogenetic and Phylogeographic Insights from Sequence and Microsatellite Analysis and Potential Impact on Emergent Chagas Disease. *Plos Neglected Tropical Diseases*, **3**.
- Milgroom MG (2001) The synthesis of genetics and epidemiology: Contribution of population biology in plant pathology. *Journal of Plant Pathology*, **83**, 57-62.
- Milgroom MG, Peever TL (2003) Population Biology of Plant Pathogens: The Synthesis of Plant Disease Epidemiology and Population Genetics. *Plant Disease*, **87**, 608-617.
- Ostfeld RS, Glass GE, Keesing F (2005) Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution*, **20**, 328-336.
- Pinon J (1980) Etude de la survie des urédospores de *Melampsora larici-populina* Kleb. et de *M. allii-populina* Kleb. *Annales des Sciences Forestières* **37**, 85-89.
- Pinon J, Frey P (1997) Structure of *Melampsora larici-populina* populations on wild and cultivated poplar. *European Journal of Plant Pathology*, **103**, 159-173.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.
- Riou-Nivert P (2001) Le mélèze. Institut pour le Développement Forestier, 144 pp.
- Rousset F (1997) Genetic Differentiation and estimation of gene flow from *F*-Statistics under isolation by distance. *Genetics*, **145**, 1219–1228.
- Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
- Stenberg P, Lundmark M, Saura A (2003) MLGsim: a program for detecting clones using a simulation approach. *Molecular Ecology Notes*, **3**, 329-331.
- Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9440-9445.
- Taris B (1968) Contribution à l'étude des rouilles des *Populus* observées en France. *Annales des Epiphyties* **19**, 5-54.
- Vaha JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63-72.

- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535-538.
- Walker J, Hartigan D, Bertus AL (1974) Poplar rusts in Australia with comments on potential conifer rusts. *European Journal of Forest Pathology* **4**, 100-118.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.
- Wilkinson AG, Spiers AG (1976) Introduction of poplar rusts *Melampsora larici-populina* and *Melampsora medusae* to New-Zealand and their subsequent distribution. *New Zealand Journal of Science* **19**, 195-198.
- Wilkinson AG, Spiers AG (1976) Introduction of the poplar rusts *Melampsora larici-populina* and *M. medusae* to New Zealand and their subsequent distribution. *New Zealand Journal of Science*, **19**, 195-198.
- Wisler GC, Norris RE (2005) Interactions between weeds and cultivated plants as related to management of plant pathogens. *Weed Science*, **53**, 914-917.
- Xhaard C, Andrieux A, Halkett F, Frey P (2009) Characterization of 41 microsatellite loci developed from the genome sequence of the poplar rust fungus, *Melampsora larici-populina*. *Conservation Genetics Resources*, **1**, 21-25.
- Xhaard C, Fabre B, Andrieux A, Gladieux P, Barrès B, Frey P, Halkett F (2011) The genetic structure of the plant pathogenic fungus *Melampsora larici-populina* on its wild host is extensively impacted by host domestication. *Molecular Ecology*, **20**, 000-000 (in press)
- Zaffarano PL, McDonald BA, Linde CC (2008) Rapid speciation following recent host shifts in the plant pathogenic fungus *Rhynchosporium*. *Evolution*, **62**, 1418–1436.
- Zogg Z (1949) Untersuchungen über die Epidemiologie des Maisrostes *Puccinia sorghi* Schw. *Phytopathologische Zeitschrift*, **15**, 143-192.

Chapitre 4

Déroulement de l'épidémie dans la vallée de la Durance

Chapitre 4

Mise en évidence du déroulement de l'épidémie dans la vallée de la Durance

1. Introduction

Des observations réalisées depuis 2001 ont permis de mettre en évidence qu'une épidémie de rouille du peuplier se répétait annuellement le long de la vallée de la Durance, depuis la zone de sympatrie peuplier - mélèze selon toute vraisemblance (Frey *et al.* 2005, Gérard *et al.* 2006). Cependant aucun relevé n'avait encore permis de connaître le déroulement temporel exact de l'épidémie, à savoir comment s'effectuait la colonisation, si elle avait lieu graduellement d'amont en aval ou s'il pouvait y avoir des sauts de plusieurs kilomètres le long de la vallée.

Une première approche pour quantifier l'épidémie avait déjà été réalisée en 2004 (Chapitre 3), mais les indices de notation choisis se sont avérés peu précis, de plus cette année-là l'épidémie s'était assez peu développée. C'est pourquoi le suivi du déroulement spatio-temporel de l'épidémie de *M. larici-populina* avec une méthode de quantification idoine de l'incidence a revêtu une grande importance dans ces travaux de thèse. Cette caractérisation fine est en outre un pré-requis pour l'analyse des caractéristiques génétiques des individus prélevés lors de cette même épidémie. A terme, ces travaux ont pour but d'associer l'épidémiologie et la génétique des populations pour analyser l'évolution des caractéristiques génétiques en regard du contexte épidémiologique qui a généré la colonisation.

Les objectifs de cette étude sont de décrire au mieux les processus démographiques sous-jacents au déroulement de l'épidémie, en particulier de confirmer qu'elle débute bien dans la zone de sympatrie peuplier - mélèze et de déterminer son patron de colonisation. Est-ce qu'il y a colonisation graduelle des sites d'amont en aval avec une progression régulière de l'épidémie dans le temps et dans l'espace, ou est-ce que l'on observe des événements de colonisation plus stochastiques et quelle est leur fréquence ? Pour atteindre ces objectifs, une campagne de notations successives dans le temps sur des sites sélectionnés le long de la vallée a été réalisée afin de transcrire au mieux de la réalité du processus.

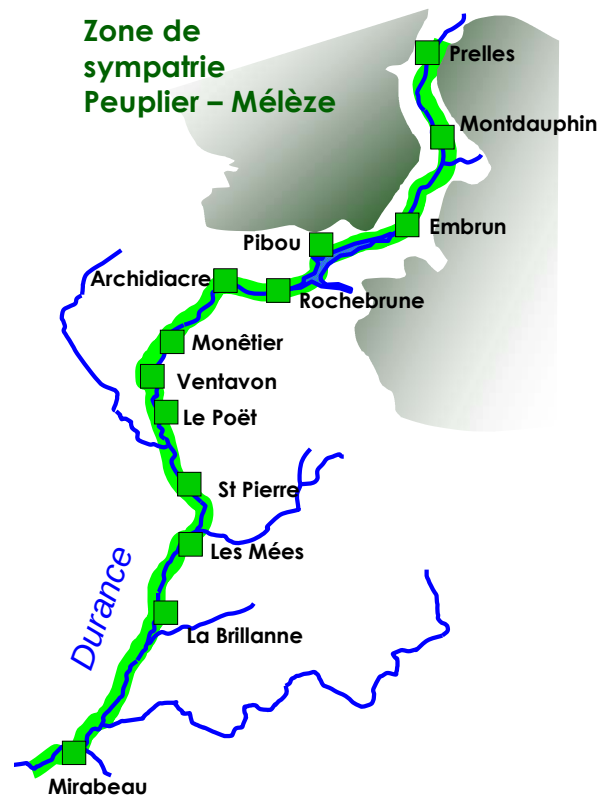


Figure 1 : Carte schématique de la vallée de la Durance avec les sites étudiés en 2008. La zone de sympatrie peuplier - mélèze est indiquée en vert foncé. La ripisylve continue le long de la rivière est indiquée en vert pâle.

2. Méthodologie pour le suivi de l'épidémie

Le déroulement de l'épidémie de rouille du peuplier dans la vallée de la Durance a été étudié en 2008 entre mi-juillet et mi-novembre. Sept campagnes de notation, espacées de trois semaines, ont été réalisées par des groupes de trois personnes. Treize sites ont été visités à chaque fois, tous situés le long de la ripisylve : quatre sites sont situés dans la zone de sympatrie peuplier - mélèze, les autres dans la zone d'allopatricie en aval du lac de Serre Ponçon. (Fig. 1). Ces treize sites étaient plus nombreux et plus régulièrement espacés le long de la vallée que ceux notés lors de l'étude de 2004 (Chapitre 3). A chaque visite, deux types de notations ont été réalisés : une notation prenant en compte le temps de collecte ou le nombre de feuille infectées trouvées dans le temps imparti et l'autre réalisée sur le nombre de rameaux infectés par *M. larici-populina*.

2.1. Première notation : temps de collecte de feuilles infectées

Dans un premier temps, les trois collecteurs disposaient d'un temps maximal de quinze minutes pour trouver et collecter au maximum quinze feuilles infectées. Si moins de quinze feuilles avaient été collectées pendant les quinze minutes imparties, le nombre de feuilles collectées était noté. Dans l'autre cas, quand moins de quinze minutes étaient nécessaires pour collecter les quinze feuilles, le temps de collecte pour chaque collecteur était noté. Le nombre d'arbres visités était également noté, en effet une seule feuille infectée par arbre était prise en compte dans la notation afin d'éviter un biais dû à la multiplication clonale (surreprésentation) locale. Ainsi pour cette première notation, les informations acquises consistent en un nombre de feuilles infectées collectées par unité de temps et ce pour chacun des trois collecteurs. Les feuilles collectées ont été conservées pour des analyses génétiques.

2.2. Deuxième notation : pourcentage de feuilles infectées au sein de rameaux

Une deuxième notation plus précise a été effectuée en cas d'incidence non négligeable de rouille du peuplier sur le site visité. Vingt rameaux étaient prélevés au hasard, le nombre de feuilles totales et le nombre de feuilles portant au minimum une urédie ont été relevés. Ce second indice nous a informés sur le pourcentage de rameaux infectés et sur le pourcentage de feuilles infectées. Cette seconde notation n'a été effectuée que dans le cas où un nombre minimum de dix feuilles infectées avait été dénombré par l'ensemble des collecteurs pendant la première notation.

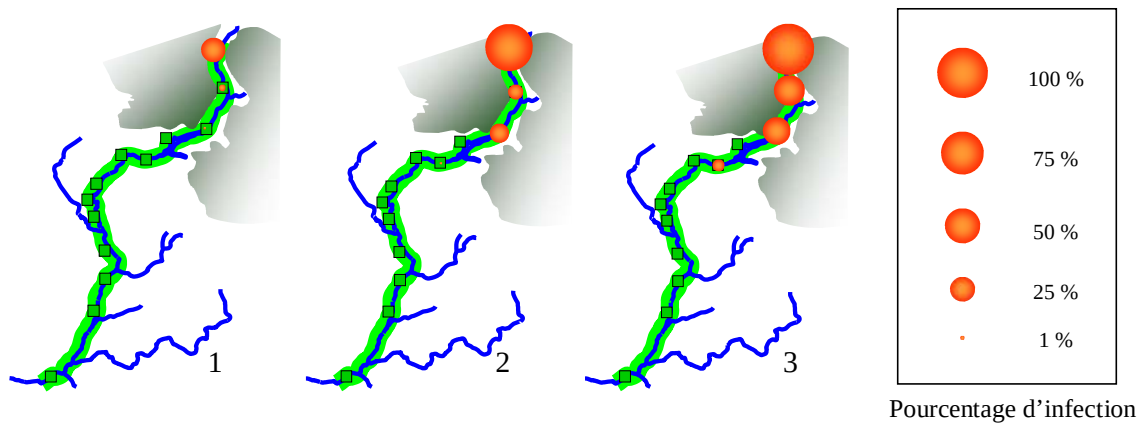


Figure 2 : Evolution des pourcentages de feuilles infectées au cours des sept campagnes de notation sur les treize sites étudiés le long de la Durance.

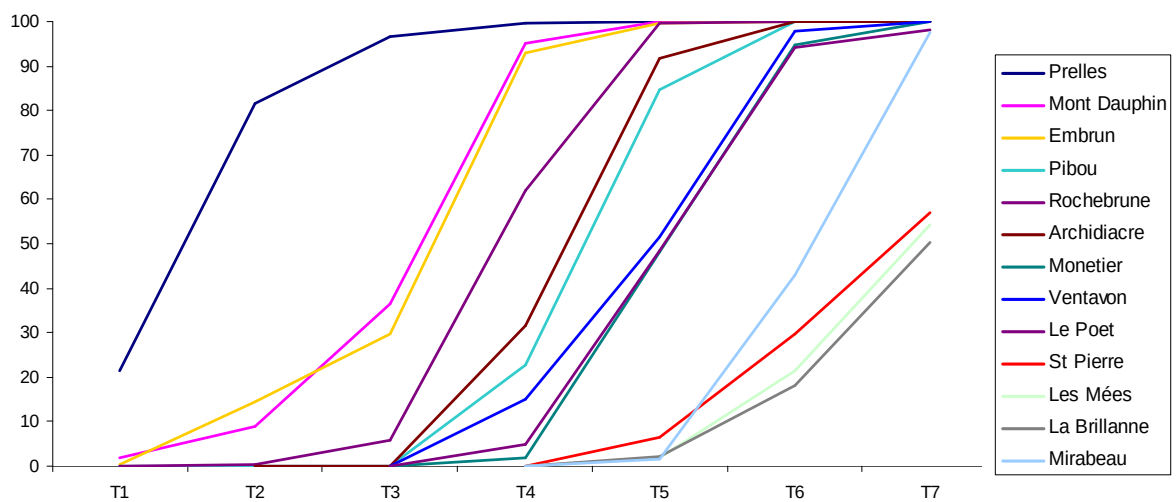


Figure 3 : Pourcentage de feuilles infectées par site et par date de notation (T) entre mi-juillet et mi-novembre.

3. Résultats

Les deux types de notations utilisées indiquent que d'une manière générale les sites ont été infectés de l'amont vers l'aval suivant leur ordre le long de la rivière. L'épidémie débute dans la zone de sympatrie peuplier - mélèze et se propage le long de la rivière au cours du temps. Les deux indices sont relativement bien corrélés ($R^2=0.71$), mais le premier indice semble assez sensible aux différences entre les notateurs, en particulier l'expérience du notateur à rechercher des feuilles infectées a joué sur les résultats.

En revanche le second indice semble à la fois plus précis et moins sensible aux différences entre les notateurs. C'est pourquoi nous avons privilégié l'analyse des résultats obtenus grâce au second indice. La première notation du temps de collecte a simplement été utilisée pour certifier l'absence de rouille en un site. Seule l'analyse des pourcentages de feuilles infectées est reportée ici (Fig. 2), l'évolution de la maladie était trop rapide à l'échelle du rameau (saturation obtenue en l'espace d'un intervalle entre deux campagnes de notation successives).

Le processus de colonisation est similaire entre les sites (Fig. 3), mais les sites sont infectés les uns après les autres selon leur position dans le transect. Ils arrivent à saturation dans l'ordre de leur emplacement le long de la rivière (Fig. 2 & 3) avec néanmoins quelques exceptions. En effet, le site le plus en aval, Mirabeau, atteint la saturation à la dernière campagne, contrairement aux trois autres sites légèrement en amont (Saint-Pierre, Les Mées et La Brillanne) qui ne sont infectés qu'à 50% environ même à la mi-novembre. Les conditions climatiques de l'année 2008 se sont avérées très favorables à l'épidémie de rouille, car l'infection est arrivée à saturation pour la plupart des sites à la fin de la saison, ce qui n'a quasiment jamais été observé entre 2002 et 2007. Il est toutefois à noter que le barème de notation ne prenait pas en compte le nombre de lésions dues à la rouille sur chaque feuille. Ainsi, bien que les sites soient indiqués comme saturés (100% de feuilles infectées), de nouvelles lésions pouvaient émerger sur des feuilles déjà infectées par *M. larici-populina*.

4. Perspectives sur la modélisation du déroulement de l'épidémie

Les perspectives de ce travail sont de modéliser l'épidémie pour en inférer les paramètres de croissance démographique et de dispersion entre sites. Ce travail de modélisation a été initié au cours d'une collaboration avec Frédéric Fabre et Samuel Soubeyrand de l'INRA d'Avignon.

Chapitre 5

Conséquences génétiques de la dispersion de la rouille du peuplier dans la vallée de la Durance

Chapitre 5

Conséquences génétiques de la dispersion de la rouille du peuplier dans la vallée de la Durance

1. Introduction

Dans le chapitre précédent nous avons précisé la nature du processus épidémique qui a lieu annuellement dans la vallée de la Durance. Nous avons tout d'abord décrit la dynamique spatio-temporelle d'infection de treize sites répartis dans la vallée. Puis nous avons montré que l'infection débutait dans la partie la plus en amont de la Durance, et que la colonisation des peupliers situés le long de la rivière s'effectuait graduellement. Les sites suivis ont atteint les uns après les autres la saturation de l'indice d'incidence en fonction de leur position dans la vallée. Cette description de l'épidémie basée uniquement sur les données épidémiologiques n'a pas pris en compte les résultats des chapitres antérieurs sur la structuration génétique des individus de *M. larici-populina* aussi bien à l'échelle de la France (chapitre 1) que régionale (chapitre 2). A l'échelle de la vallée de la Durance, la dimension épidémiologique est complexifiée par le fait que les individus présents dans la zone d'initiation de l'épidémie (zone de sympatrie peuplier – mélèze) constituent deux pools génétiques différents séparés par le Massif des Ecrins (chapitre 2). L'épidémie qui s'en suit dans la Vallée de la Durance présente donc un mélange d'individus des deux origines. Leur appartenance peut néanmoins être facilement établie à partir de leur génotype par les méthodes d'assignation. Dans la partie amont de la vallée, jusqu'au lac de Serre-Ponçon, l'épidémie est constituée quasiment uniquement d'individus appartenant au groupe génétique « sauvage ». En revanche, dans la partie plus en aval, les individus des deux groupes génétiques sont en mélange. Ceci nous a permis de conclure en chapitre 3 que deux épidémies simultanées ont lieu dans la vallée de la Durance avec des individus de deux groupes génétiques distincts.

Cette dernière étude complète la description de l'épidémie, en s'intéressant à l'évolution des caractéristiques génétiques des individus collectés le long du gradient généré par l'épidémie. Cette caractérisation ne peut se faire que sur les individus de chaque groupe considéré indépendamment. Pour cette étude, seuls les individus du groupe génétique dit « sauvage » ont été considérés. En effet, l'autre groupe, dit « cultivé », présente un nombre beaucoup plus faible d'individus, moins bien répartis sur les sites de collecte et aucun signal

d'isolement par la distance n'a été détecté, c'est-à-dire qu'il n'y a pas de signe évident de différenciation entre les individus de ce groupe quelle que soit leur position géographique le long de la vallée. En revanche, un signal d'isolement par la distance a été montré pour les individus de l'autre groupe génétique. Nous avons donc cherché dans cette étude à détailler plus finement les patrons spatiaux de diversité génétique sur le groupe « sauvage ».

Les objectifs de cette étude ont été de préciser la structure génétique spatiale des individus de *M. larici-populina* appartenant au groupe « sauvage », collectés le long de la Durance à la fin de l'épidémie 2004, afin d'évaluer les conséquences génétiques du déroulement de l'épidémie.

2. Matériels et méthodes

A partir de l'échantillonnage réalisé à l'automne 2004 dans la vallée de la Durance et en nous basant sur les résultats des assignations réalisées dans le chapitre 3, nous avons sélectionné uniquement les individus de *M. larici-populina* appartenant au groupe « sauvage ». Sur l'ensemble des individus de ce groupe, nous n'avons conservé que les sites pour lesquels un minimum de cinq individus était disponible.

Nous avons ensuite réalisé des analyses standard de génétique des populations pour chacun des sites retenus. Nous avons réalisé les analyses en prenant en compte ou en ôtant les répétitions d'un même génotype multilocus (une seule copie par clone). Les niveaux d'hétérozygotie observée et attendue ont été calculés avec GENETIX (4.05) (Belkhir *et al.* 2004). La richesse allélique, et les indices de différenciation intra-groupes (F_{IS}) et inter-groupes (F_{ST} par paire, calculés seulement entre Prelles, considéré comme la population de référence, et les autres sites) ont été calculés avec FSTAT version 2.9.3 (Goudet, 1995). Puis nous avons examiné comment variaient les valeurs des différents indices pour chaque site, en fonction de la distance de ces sites par rapport à Prelles. Tous ces indices ont été calculés en prenant ou non en compte les répétitions clonales.

Tableau 1 : Caractéristiques génétiques des individus prélevés en fin d'épidémie en 2004 et appartenant au groupe « sauvage » en fonction de leur site de collecte. Nb Ind. : nombre d'individus ; G/N : richesse génotypique ; Ar : richesse allélique calculée pour un échantillon de 9 individus (ou de 8 individus dans le cas où les copies clonales sont prises en compte) ; H_E : hétérozygotie attendue (Nei 1978) ; H_O : hétérozygotie observée (les valeurs des écart-types sur les locus sont données entre parenthèses) ; F_{IS} : différenciation intra-groupe ; F_{ST} / Prelles : différenciation par rapport aux individus du site de Prelles (* indique une $P < 0.05$).

Sites	1 ^{ère} mention de la maladie	Distance /Prelles	Nb Ind.	G/N	Sans les copies clonales					Avec les copies clonales				
					Ar	H_E	H_O	F_{IS}	F_{ST} / Prelles	Ar	H_E	H_O	F_{IS}	F_{ST} / Prelles
Prelles	07-juil	0	34	1	3.43	0.499 (0.189)	0.461 (0.212)	0.078		3.43	0.506 (0.188)	0.458 (0.212)	0.096	
Saint-Clément	08-sept	22	9	1	3.12	0.475 (0.223)	0.479 (0.288)	- 0.009	0.011	3.12	0.491 (0.230)	0.466 (0.288)	0.055	0.010
Châteauroux	28-juil	26	16	0.75	2.17	0.378 (0.243)	0.449 (0.338)	- 0.197	0.129 (*)	2.14	0.355 (0.245)	0.440 (0.338)	- 0.250	0.155 (*)
Embrun	18-aout	33	43	0.95	3.17	0.481 (0.205)	0.448 (0.236)	0.070	0.005 (*)	3.15	0.489 (0.211)	0.439 (0.236)	0.104	0.007 (*)
Rochebrune	07-juil	53	28	0.89	3.22	0.483 (0.187)	0.429 (0.216)	0.112	0.004	3.16	0.478 (0.192)	0.419 (0.216)	0.126	0.005
Monêtier	18-aout	73	23	0.96	3.37	0.503 (0.199)	0.447 (0.228)	0.115	0.014 (*)	3.38	0.510 (0.200)	0.442 (0.228)	0.137	0.013 (*)
Valernes	29-sept	85	26	0.62	2.72	0.403 (0.253)	0.366 (0.301)	0.094	0.058 (*)	2.67	0.403 (0.253)	0.373 (0.301)	0.077	0.065 (*)
Les Mées	07-juil	103	47	0.25	2.96	0.450 (0.242)	0.419 (0.249)	0.073	0.055 (*)	2.26	0.337 (0.226)	0.396 (0.249)	- 0.178	0.223 (*)
La Brillanne	29-sept	117	21	0.10	-	-	-	-	-	1.41	0.213 (0.256)	0.381 (0.484)	- 0.826	0.304 (*)

3. Résultats

3.1. Sites de collecte

Sur les dix sites de collecte étudiés précédemment, neuf sites présentant plus de cinq individus assignés au groupe « sauvage » ont été retenus. Le site le plus au sud, Mirabeau, n'a pas été conservé car il comportait seulement trois individus assignés au groupe « sauvage ». Les neufs sites de collecte sont localisés le long de la Durance sur un transect de 120 km et sont espacés de 4 à 22 km (Tableau 1). Pour le site de La Brillanne, 21 individus ont été collectés, mais représentés seulement par deux génotypes distincts.

3.2. Diversité génotypique

La richesse génotypique varie de 1 à 0,1 (Tableau 1) et diminue fortement en fonction de la distance par rapport à Prelles (Fig. 1). Pour les sites situés dans la partie amont, la diversité génotypique décroît légèrement, puis elle diminue drastiquement pour les sites situés dans la partie aval. D'ailleurs le site échantillonné le plus en aval (La Brillanne) présente une richesse génotypique très faible, avec 20 copies du même génotype répété et une seule copie d'un génotype différent. Ces résultats montrent que la proportion locale de génotypes répétés augmente avec l'éloignement du site de collecte avec la source de l'épidémie (ici représentée par le site de Prelles).

3.3. Richesse allélique et diversité génétique

D'une manière générale ces deux indices présentent les mêmes variations (Tableau 1 et Fig. 2 et 3). Ils présentent des valeurs élevées et quasiment constantes dans la partie la plus en amont, jusqu'au site de Monétier, puis une diminution régulière pour les sites en aval de Monétier. Cette diminution est surtout visible pour les calculs réalisés en prenant en compte les répétitions de génotypes. Une exception à la tendance générale apparaît pour les individus provenant du site de Châteauroux, ceux-ci présentant des valeurs inférieures à celles des sites voisins. Bien que situé très en amont de la vallée, ce site présente des valeurs plus faibles de richesse allélique et de diversité génétique. Les variations de ces indices sont néanmoins cohérentes avec une plus grande proportion de génotypes répétés dans ce site.

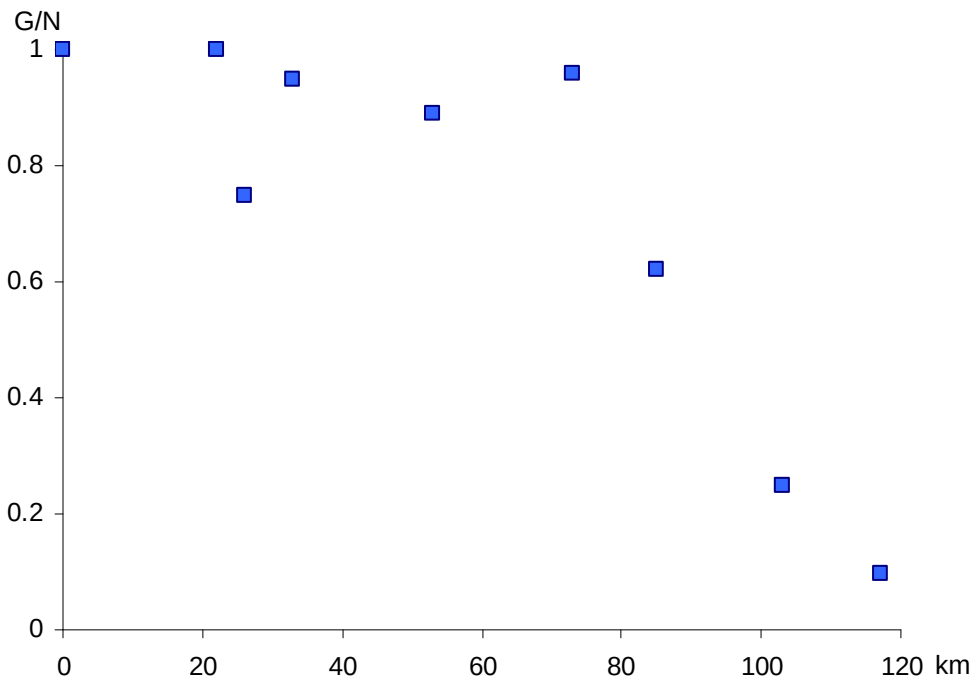


Figure 1 : Evolution de la richesse génotypique (G/N) en fonction de la distance de chaque site par rapport à Prelles.

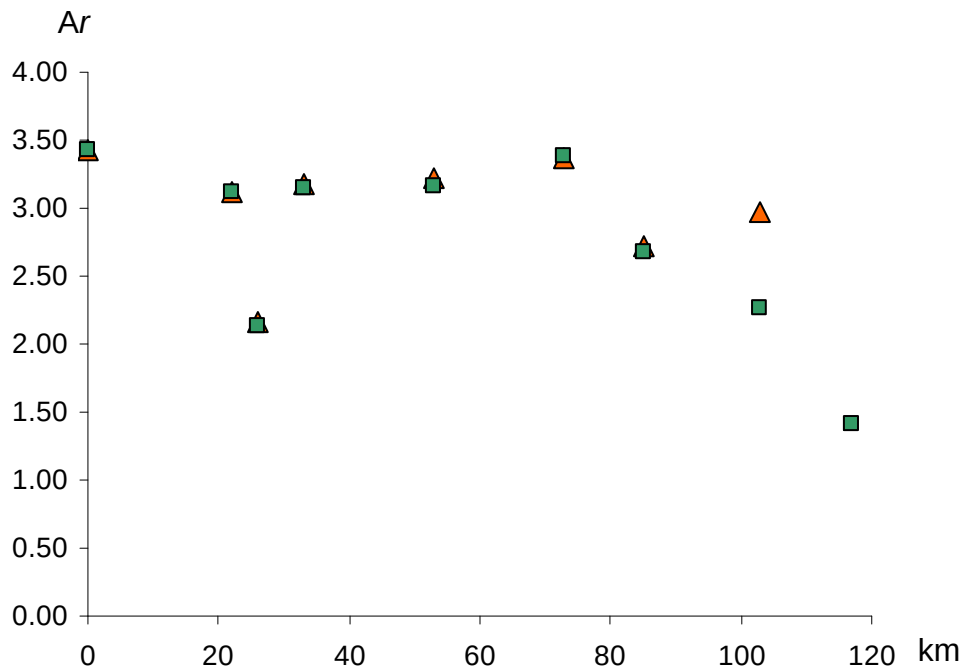


Figure 2 : Richesse allélique (Ar) en fonction de la distance de chaque site par rapport à Prelles (triangle orange : sans tenir compte des répétitions d'un même génotype, carré vert : en prenant en compte tous les individus).

3.4. Différenciation génétique

Le calcul de la différenciation a été réalisé en utilisant les individus de Prelles comme référence pour les comparer aux autres individus de *M. larici-populina* collectés le long de la Durance. Comme pour les autres indices, les différenciations présentent deux profils en fonction de la distance par rapport à Prelles (Tableau 1 et Fig. 4). Les valeurs de F_{ST} pour les individus des sites situés dans la partie amont de la vallée jusqu'à Monétier sont très faibles, indiquant un faible niveau de différenciation entre ces individus et ceux de Prelles. En revanche, dans la partie aval du transect, pour des sites plus éloignés de Prelles, les valeurs de F_{ST} augmentent régulièrement avec la distance par rapport à Prelles. Les comparaisons entre les F_{ST} sont quasiment toutes significativement différentes de Prelles, à l'exception de Saint-Clément et Embrun. Comme pour la variation des autres indices, les individus collectés à Châteauroux marquent une exception à la tendance générale, et sont fortement différenciés par rapport aux individus de Prelles.

4. Discussion

Les différents indices génétiques révèlent deux profils contrastés le long de la vallée de la Durance lorsqu'on les compare à la population de référence (Prelles). Le premier profil concerne la partie la plus en amont de la vallée et le second profil la partie la plus en aval. Dans la partie amont de la vallée, la diversité génétique et génotypique des individus est relativement stable et les individus des différents sites ne présentent pas de différenciation avec la population de référence, Prelles. Dans la partie aval de la vallée, on constate une érosion de la diversité génétique et génotypique accompagnée d'une augmentation graduelle de la différenciation génétique en fonction de la localisation des individus le long de la vallée de la Durance, toujours en se référant aux individus de Prelles. La diminution de la diversité accompagnée d'une augmentation de la différenciation révèle que la dispersion des individus de *M. larici-populina* s'effectue par des effets de fondations récurrents le long de la Durance (Hewitt 2000). L'épidémie s'effectuerait sous la forme d'une vague d'invasion progressive générée par la diffusion des individus le long du corridor formé par la Durance. Pendant la phase épidémique, les individus se multiplient uniquement de manière clonale, ainsi ce processus de ré-échantillonnages successifs expliquerait l'augmentation de la proportion de génotypes répétés le long de la vallée. Au dernier point d'échantillonnage, les individus collectés avaient quasiment tous le même génotype.

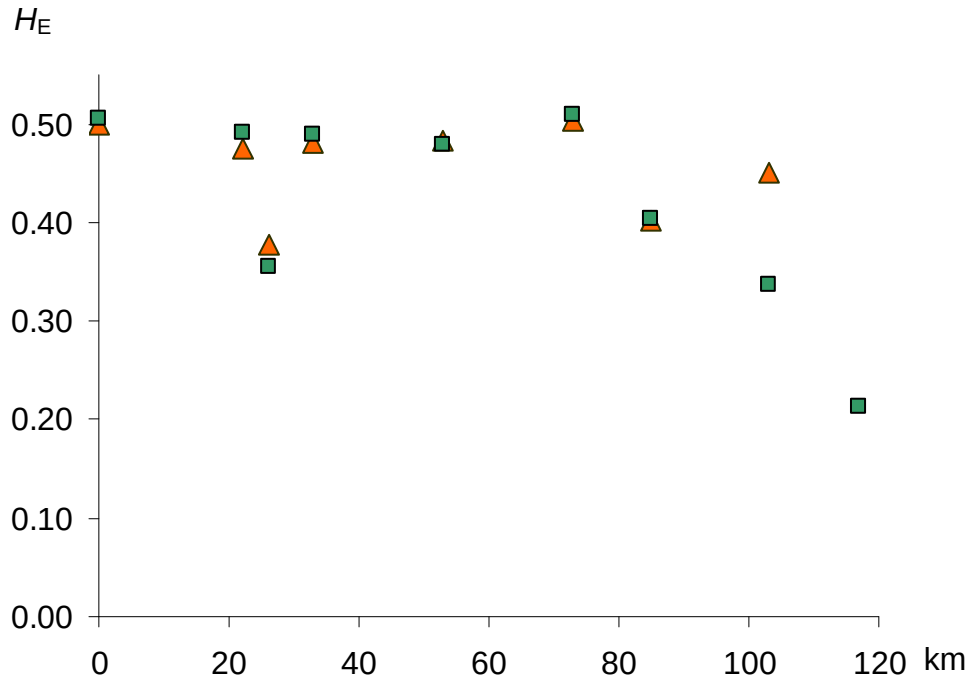


Figure 3 : Hétérozygoties attendues (H_E) en fonction de la distance des sites par rapport à Prelles (triangle orange : sans tenir compte des répétitions d'un même génotype, carré vert : en prenant en compte tous les individus).

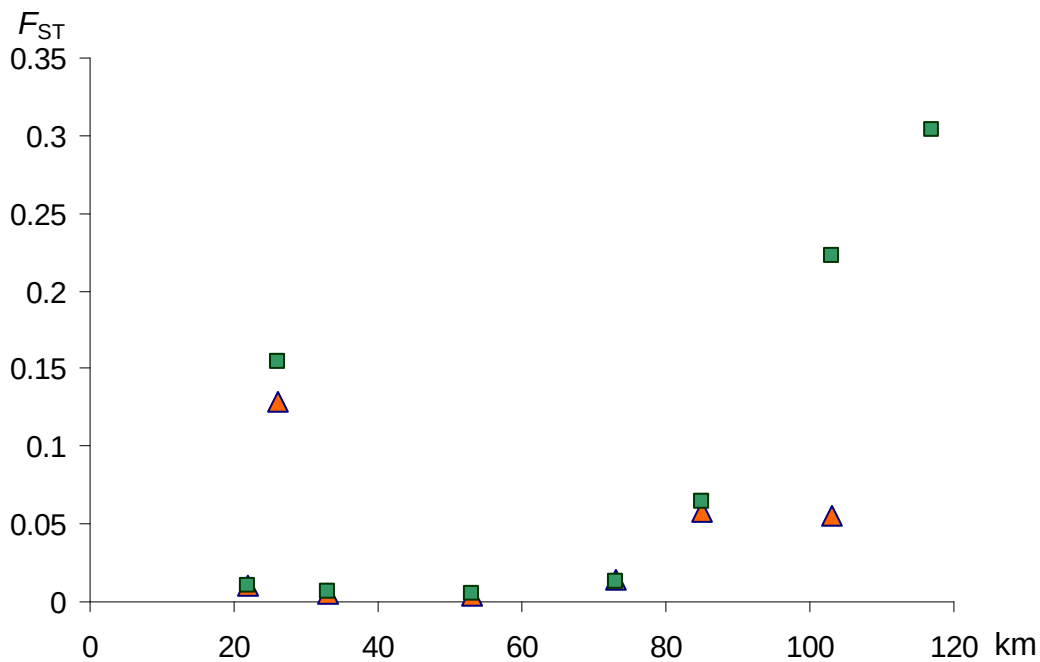


Figure 4 : Différenciations génétiques (F_{ST}) entre les individus de Prelles et les individus des autres sites indiquées en fonction de la distance par rapport à Prelles (triangle orange : sans tenir compte des répétitions d'un même génotype, carré vert : en prenant en compte tous les individus).

Cependant ce scénario ne fonctionne pas pour les individus de la partie la plus en amont de la vallée, où la différenciation n'augmente pas et la diversité ne diminue pas non plus. Deux hypothèses peuvent expliquer cette absence de signal. Une première explication réside dans le fait que cette partie de la vallée est située dans la zone de sympatrie peuplier - mélèze ou à proximité. Ainsi les individus de cette première partie de la vallée sont directement issus de la phase de reproduction sexuée et présentent donc une forte diversité génétique. La deuxième explication serait que ces sites (situés à proximité de la source d'inoculum) aient subi plusieurs vagues d'invasion au cours de l'épidémie, constituées à chaque fois d'individus différents. Les individus de ces sites seraient donc issus du mélange de plusieurs vagues d'invasion, noyant ainsi le signal d'expansion. Les individus du site de Châteauroux créent une exception par rapport à cette tendance, ils présentent à la fois une forte différenciation et une faible diversité génétique, indiquant une situation locale défavorable à l'installation de la rouille du peuplier, comme par exemple une humidité trop faible. Seuls quelques individus auraient pu se développer et s'installer en se multipliant de manière asexuée localement.

Les sites de la partie aval de la vallée, en revanche, sont beaucoup plus loin de la source et reflètent les signes d'expansion clonale de l'épidémie de rouille. L'augmentation de la différenciation ainsi que la diminution de la diversité peuvent être liés à un phénomène de « gene surfing », c'est-à-dire une multiplication locale au niveau du front de migration d'individus aux fréquences alléliques plus rares que dans la population d'origine (Hallatschek & Nelson, 2008).

Toutefois on ne peut pas exclure qu'il y ait aussi sélection d'individus aux capacités de dispersion plus élevées au cours de l'épidémie. La détection d'un phénomène de sélection au cours des différentes phases de la dispersion a fait l'objet de deux stages de Master 2. Différents traits de vie ont été comparés entre des individus issus de Prelles et des individus provenant du site le plus au sud du transect (Mirabeau), collectés en 2008 et en 2009. Les premières conclusions vont dans le sens d'une sélection vers des individus produisant un plus grand nombre de spores, mais de plus petite taille pour le site le plus au sud. Ceci favoriserait la dispersion mais diminuerait probablement les capacités compétitrices de ces individus.

5. Conclusion et Perspectives

Ce travail ne représente qu'une étude préliminaire réalisée sur les individus collectés à la fin de l'épidémie 2004. Durant la saison 2008, nous avons augmenté le nombre de sites de collecte en les choisissant mieux répartis spatialement et nous avons également réalisé un échantillonnage temporel. Ainsi les perspectives de ce travail sont de pouvoir caractériser l'évolution des caractéristiques génétiques à la fois de manière temporelle et spatiale. De plus un échantillonnage rapproché a été réalisé dans le but de mettre en évidence les variations spatiales fines des fréquences alléliques des individus de *M. larici-populina*. L'analyse de ces échantillons permettra de distinguer les différents scénarii proposés ci-dessus et de caractériser plus finement les conséquences génétiques de la dispersion de *M. larici-populina* au cours de l'épidémie. De manière plus générale, la vallée de la Durance est particulièrement bien adaptée à l'étude des invasions biologiques. En effet, chaque année une nouvelle épidémie de rouille du peuplier débute, avec une « remise à zéro » car les individus situés hors de la zone de sympatrie peuplier - mélèze ne participent pas à la reproduction sexuée.

Discussion générale

DISCUSSION GENERALE

Les analyses effectuées au cours de ce travail sur des populations de *Melampsora larici-populina* ont pour objectif commun la caractérisation de l'influence de la sélection et des processus démographiques sur la structure des populations, à des échelles spatiales plus ou moins larges. Dans un premier temps, le développement d'outils moléculaires efficaces a été nécessaire pour mettre en évidence des structures génétiques fines. Nous avons pu profiter du fait que la séquence du génome de *M. larici-populina* soit disponible (Duplessis *et al.* 2011) pour développer 41 nouveaux marqueurs microsatellites. Un sous-ensemble formant un panel de 25 marqueurs a ensuite été utilisé en routine pour l'ensemble des études de cette thèse. Les différentes analyses réalisées ensuite se divisent en deux thématiques principales : la mise en évidence de structures génétiques à la suite d'un contournement d'une résistance portée par l'hôte cultivé et le couplage d'approches d'épidémiologie et de génétique des populations pour caractériser au mieux les processus démographiques.

1. Conséquences génétiques et démographiques d'un contournement de résistance

Melampsora larici-populina n'infecte pas indifféremment des peupliers sauvages ou cultivés, car le plus souvent les peupliers cultivés sont porteurs de résistances qualitatives qui leur permettent d'éviter l'infection par certains génotypes d'un agent pathogène. Ainsi l'hôte cultivé porteur d'une résistance exerce une pression de sélection importante sur les populations de l'agent pathogène car seuls les individus qui auront contourné la résistance seront capables d'infecter l'hôte cultivé. La première thématique s'intéresse donc aux conséquences génétiques et démographiques du contournement de la résistance R7, qui a été particulièrement bien documenté (Pinon *et al.* 1998). Cette problématique a été abordée à des échelles spatiales différentes. Après une étude descriptive de la structure des populations de *M. larici-populina* à une large échelle, un zoom sur des territoires particuliers a permis d'enrichir notre compréhension de la dynamique d'interaction entre groupes génétiques présentant des histoires évolutives différentes.

1.1. A l'échelle de la France

Nous avons évalué l'impact du contournement de la résistance R7 sur la structure génétique de populations de *M. larici-populina* échantillonnées à travers la France uniquement sur l'hôte sauvage. Les tests d'assignations ont détecté la présence de trois groupes génétiques, deux majoritaires et un minoritaire.

Les deux principaux groupes ont des caractéristiques très dissemblables. D'un point de vue phénotypique, le premier groupe est composé en majorité d'individus Vir7. Les individus de ce groupe sont aussi moins diversifiés génétiquement que ceux du deuxième groupe, et des loci sous sélection ont été détectés pour ce premier groupe génétique. Enfin ce groupe présente des signes d'expansion démographique suite à la réduction drastique des effectifs au moment du contournement. Toutes ces caractéristiques démontrent l'origine cultivée des individus du premier groupe (Burdon & Thrall 2008 ; Gladieux *et al.* 2010). A l'inverse le deuxième groupe est composé d'individus Avr7 et présente une plus grande diversité génétique. Ces différentes caractéristiques nous ont permis de nommer le premier groupe « cultivé » et le deuxième « sauvage ». En outre, les deux groupes génétiques sont structurés spatialement. Les individus du groupe « cultivé » se trouvent en majorité dans la moitié nord de la France alors que les individus du groupe « sauvage » sont plutôt répartis dans la moitié sud. Cette structuration spatiale pourrait résulter de la plantation plus importante de cultivars interaméricains porteurs de la résistance R7 dans la moitié nord de la France par rapport à la moitié sud. Ces hôtes exercent une telle pression de sélection qu'elle se ressent même sur les populations de *M. larici-populina* infectant l'hôte sauvage. Des différenciations génétiques sont assez souvent mises en évidence pour des populations d'agents pathogènes infectant des hôtes sauvages ou cultivés (Gladieux *et al.* 2008 ; Frenkel *et al.* 2010). Mais dans la plupart des études il s'agit d'individus collectés sur des hôtes différents. Notre étude à la particularité d'avoir été conduite sur l'hôte sauvage uniquement. Le fait que l'on retrouve à la fois des individus des groupes « sauvage » et « cultivé » sur ces hôtes sauvages indique la forte influence des hôtes cultivés sur les populations de *M. larici-populina*.

Le troisième groupe génétique détecté au cours de cette étude présente les caractéristiques d'une lignée asexuée. Bien que suspectée par certains auteurs (Dupias 1965, Taris 1968, Walker *et al.* 1974, Wilkinson & Spiers 1976, Pinon 1980), aucun signe de reproduction asexuée durable chez *M. larici-populina* n'avait été clairement mis en évidence jusqu'alors. La mise en évidence d'individus de *M. larici-populina* présentant toutes les

caractéristiques génétiques d'une reproduction asexuée durable constitue donc une découverte aussi inattendue qu'intéressante. Au cours de cette thèse nous avons assez peu exploré les particularités de ces individus. Pourtant cette découverte ouvre tout un pan de recherches concernant l'étude de leur biologie et de leur maintien dans le temps. Sur cette question du maintien de cette lignée asexuée, nous avons pu montrer au travers des différents échantillonnages que des individus au génotype très proches avaient perduré sur une période de cinq ans (échantillonnage 2004 en chapitre 3 et échantillonnage 2009 en chapitre 1). Cependant de nombreuses questions restent ouvertes. Comment effectuent-ils leur cycle de vie ? Comment se comportent-ils en cas de compétition avec des individus à reproduction sexuée ? Quels sont les déterminants de la perte de reproduction sexuée ? Cette nouvelle thématique alimentera les activités de recherche de l'équipe d'accueil bien au-delà de cette thèse.

1.2. A l'échelle régionale : apport de l'étude de la vallée de la Durance

Nous nous sommes ensuite intéressés, à une échelle plus restreinte, à une zone géographique située à l'interface des domaines spatiaux des deux principaux groupes génétiques « cultivé » et « sauvage ». Nous savions grâce aux études précédentes que les deux groupes génétiques étaient en contact dans cette zone des Alpes au relief fortement marqué. De plus, dans cette région, peupliers et mélèzes évoluent en sympatrie entraînant une importante recombinaison pour les populations de *M. larici-populina*. Comment des groupes génétiques distincts pouvaient alors se maintenir ? L'utilisation d'outils de génétique du paysage a permis de détecter le rôle de barrière aux flux de gènes joué par le Massif des Ecrins. En effet les deux groupes génétiques étaient bien séparés spatialement de part et d'autre du massif. L'originalité de cette étude réside dans l'application de la génétique du paysage à des champignons pathogènes à forte dispersion aérienne. Etant donné que les spores de *M. larici-populina* peuvent se disperser à grande distance, l'application de méthodes de génétique du paysage a permis de mettre en évidence une barrière relativement récente et non pas historique comme dans la majorité des études de génétique du paysage (Storfer *et al.* 2010). Dans cette étude, le Massif des Ecrins agirait plutôt comme une barrière récente protégeant les individus « sauvages » de l'expansion des individus « cultivés » venant de l'ouest, comme nous l'avions déjà vu pour le reste du territoire.

En perspective de ce travail, il serait intéressant d'ajouter la dimension temporelle pour suivre la dynamique de ce système, afin de voir si l'invasion des individus du groupe

« cultivé » va se poursuivre au-delà du massif des Ecrins ou si la résistance du paysage est suffisamment forte pour les en empêcher (Cushman *et al.* 2006, McRae 2006).

Cependant nous possédons une indication sur la stabilité temporelle de ce système. Les deux pools génétiques situés de part et d'autre du Massif des Ecrins constituent le point de départ d'une épidémie annuelle qui se déroule du nord vers le sud le long de la vallée de la Durance. A la fin de la phase épidémique, il est possible de distinguer les individus présents sur un même site selon leur appartenance à l'un ou l'autre des groupes génétiques. Nous avons ainsi prouvé l'absence de flux de gènes entre les individus des deux groupes. Dans la partie amont, les individus « sauvages » et « cultivés » ne s'hybrident pas car ils sont séparés par le Massif des Ecrins ; dans la partie aval alors qu'ils se retrouvent en mélange sur les mêmes peupliers, la reproduction sexuée n'a pas lieu du fait de l'absence de méléze.

1.3. Stabilité des groupes génétiques et enseignements pour les stratégies de déploiement des cultivars résistants.

En effectuant des analyses sur des échantillons collectés à cinq ans d'intervalle (2004-2009), on constate une persistance temporelle des groupes génétiques. En particulier on retrouve toujours le groupe « cultivé » même pour des individus échantillonnés sur l'hôte sauvage, ce groupe génétique semble donc stable et pérenne. Il semblerait que même en l'absence d'hôtes R7, un retour en arrière avec la disparition de ce groupe serait peu probable. En effet, la Virulence 7 est quasiment fixée dans les populations de la moitié nord de la France, alors qu'elle est inutile à l'infection des hôtes sauvages. La persistance du groupe « cultivé » pourrait également nous renseigner sur l'existence ou non d'un coût de la Virulence 7. Un tel coût de la virulence a été mis en évidence dans de nombreux pathosystèmes évoluant sous la relation gène-pour-gène, pour lesquels l'acquisition d'une nouvelle virulence entraîne une baisse de la fitness (Leach *et al.* 2001, Thrall & Burdon 2003, Huang *et al.* 2006, Laine & Tellier 2008, Bahri *et al.* 2009). Nos résultats sur la persistance de la Virulence 7 dans des populations collectées sur l'hôte sauvage suggéreraient que le coût associé à cette virulence serait faible ou nul (Gérard *et al.* 2006). La pérennité des individus virulents illustre les conséquences néfastes d'une mauvaise gestion du déploiement d'un cultivar porteur d'une résistance qualitative. Une meilleure gestion serait de privilégier des cultivars à résistance quantitative qui exercent une pression de sélection moins forte sur les populations d'agents pathogènes (Jorge *et al.* 2005, Guérin *et al.* 2007 ; Brun *et al.* 2010).

2. Couplage de l'épidémiologie et de la génétique des populations

Ce deuxième thème principal a abordé l'épidémie de rouille du peuplier dans la vallée de la Durance sous un tout autre angle, il s'agit de la caractérisation du déroulement de l'épidémie annuelle qui a lieu le long de la Durance. Cette situation particulière nous a permis de coupler des approches d'épidémiologie et de génétique des populations pour décrire les événements démographiques ayant lieu au cours de l'épidémie, ainsi que leurs conséquences génétiques. L'épidémie se déroule graduellement du nord vers le sud, avec des phases de multiplication asexuée locale et de dispersion. Nous avons constaté lors de l'épidémie 2008 que les dynamiques d'infection étaient sensiblement les mêmes quels que soient les sites observés, seul un décalage dans le temps en fonction de leur position dans la vallée a été observé. Cette année-là, la quasi-totalité des sites ont été infectés jusqu'à saturation les uns après les autres selon leur position sur le gradient amont / aval.

Ce déroulement graduel a pour conséquence une évolution des caractéristiques génétiques des individus en fonction de leur position le long du gradient. Elles se traduisent par une augmentation de la différenciation par rapport aux individus situés dans la zone de sympatrie peuplier – mélèze et une érosion graduelle de la diversité génétique. La réduction de la diversité génétique est une conséquence souvent retrouvée lors d'invasions biologiques (Miura 2007). Au cours de l'épidémie, on constate un enrichissement local en génotypes répétés en fonction de la distance par rapport à la source de l'épidémie, qui conduit à la présence de seulement quelques génotypes répétés pour les sites les plus en aval. Ces caractéristiques correspondent à un phénomène de « gene surfing », pour lequel des allèles rares dans la population d'origine peuvent être localement multipliés pendant les phases d'expansion (Hallatschek & Nelson, 2008).

Mais le cas de l'épidémie de *M. larici-populina* dans la vallée de la Durance présente deux particularités qui le différencie des autres cas d'invasions biologiques. Premièrement, ce système est remis à zéro chaque année, car les individus formant l'épidémie n'ont pas la possibilité de terminer leur cycle biologique du fait de l'absence de mélèze dans la partie aval de la vallée de la Durance. Ainsi seuls les individus qui n'auront pas dispersé et qui seront restés dans la zone de sympatrie peuplier – mélèze participeront à la phase de reproduction sexuée et leurs descendants initieront l'épidémie l'année suivante. Deuxièmement, des événements de reproduction asexuée ont lieu pendant toute la période de dispersion, rendant plus complexe la modélisation des processus de dispersion. De plus la détection de « gene surfing » peut être artificiellement amplifiée par la multiplication clonale.

Pour finaliser la description de l'épidémie et ses conséquences en terme génétique, il serait bon de prendre en compte la dimension temporelle de l'invasion en plus de la dimension spatiale que nous avons étudiée. La connaissance de la dynamique de l'évolution des caractéristiques génétiques permettrait de dissocier encore plus finement les processus démographiques dus à la migration et la multiplication clonale locale.

Il serait également nécessaire de dissocier entre une évolution au hasard des caractéristiques génétiques au cours de la phase de dispersion par effet de la dérive et un effet de sélection favorisant certains génotypes. Une manière de tester l'éventuelle sélection est de s'intéresser à l'évolution de certains traits d'histoire de vie pour des individus collectés en différents points depuis la source de l'épidémie jusqu'au site le plus éloigné. Les premiers tests réalisés au laboratoire ont révélé, pour le site le plus éloigné, une sélection d'individus produisant un plus grand nombre de spores mais de plus petite taille, ce qui favoriserait la dispersion mais diminuerait les capacités compétitrices de ces individus.

Conclusion

CONCLUSION

Grâce aux études réalisées sur ces deux principales thématiques, nous avons souligné le lien important qui existe entre différents processus affectant les caractéristiques génétiques. D'une part, nous avons montré que l'évènement de sélection drastique qui accompagne un contournement de résistance avait des conséquences démographiques sur les populations de l'agent pathogène, qui pouvaient être révélées à l'aide d'outils moléculaires. Et d'autre part, ces mêmes outils moléculaires ont été utilisés pour mettre en évidence les conséquences génétiques d'un évènement démographique, l'épidémie qui se déroule dans la vallée de la Durance. Toutefois il est également nécessaire de tester si une sélection s'opère pendant l'épidémie.

En conclusion, on constate qu'une certaine indissociabilité existe entre les différents processus étudiés, la sélection et la démographie, mise en lumière grâce à l'association de l'épidémiologie et de la génétique des populations. Il devient donc nécessaire de combiner différentes disciplines pour prendre en compte l'ensemble des forces évolutives qui agissent sur des populations. En particulier, l'apport de la génétique des populations à l'épidémiologie permet une meilleure quantification des maladies car elle prend en compte les origines génétiques des populations d'agents pathogènes et permet de mieux inférer les processus de dispersion (Milgroom 2001 ; Milgroom & Peever 2003). Le lien entre la génétique des populations et l'épidémiologie est particulièrement intéressant pour caractériser la démographie, un autre lien à faire serait d'associer la génétique des populations avec des outils de génomique (Luikart *et al.* 2003 ; de Mita *et al.* 2007). En effet, la génomique des populations permet de générer des données à haut débit grâce une importante densification des loci. Cette approche permettrait ainsi de dissocier des ensembles de loci au comportement neutre ou sélectionnés pour (i) établir la part respective de la sélection et de l'histoire démographique et (ii) à partir de chaque ensemble de loci pris séparément inférer l'histoire démographique et déterminer les bases génétiques de l'adaptation de *M. larici-populina* à ses hôtes.

Bibliographie

BIBLIOGRAPHIE

(Hors references citées dans les articles)

- Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217-1229.
- Balloux F, Lehmann L, de Meeus T (2003) The population genetics of clonal and partially clonal diploids. *Genetics*, **164**, 1635-1644.
- Barrès B (2006) Etude de la structuration génétique de populations de *Melampsora larici-populina*, agent de la rouille du peuplier, à différentes échelles spatiales. Thèse de doctorat, Nancy 1, Université Henri Poincaré, 181pp.
- Barrès B, Dutech C, Andrieux A, Caron H, Pinon J, Frey P (2006) Isolation and characterization of 15 microsatellite loci in the poplar rust fungus, *Melampsora larici-populina*, and cross-amplification in related species. *Molecular Ecology Notes*, **6**, 60-64.
- Barrès B, Halkett F, Dutech C, Andrieux A, Pinon J, Frey P (2008) Genetic structure of the poplar rust fungus *Melampsora larici-populina*: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, **8**, 577-587.
- Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nat Rev Genet*, **5**, 251-261.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, 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).
- Biek R, Real LA (2010) The landscape genetics of infectious disease emergence and spread. *Molecular Ecology*, **19**, 3515-3531.
- Bourassa M, Bernier L, Hamelin RC (2007) Genetic diversity in poplar leaf rust (*Melampsora medusae* f. sp. *deltoidae*) in the zones of host sympatry and allopatry. *Phytopathology*, **97**, 603-610.
- Brown JKM, Hovmøller MS (2002) Aerial Dispersal of Pathogens on the Global and Continental Scales and Its Impact on Plant Disease. *Science*, **297**, 537-541.
- Burdon JJ, Roelfs AP (1985) The effect of sexual and asexual reproduction on the isozyme structure of populations of *Puccinia graminis*. *Phytopathology*, **75**, 1068-1073.

- Burdon JJ, Thrall PH (2008) Pathogen evolution across the agro-ecological interface: implications for disease management. *Evolutionary Applications*, **1**, 57–65.
- Chen C, Durand E, Forbes F, Francois O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes*, **7**, 747-756.
- Chen W, Harrington C (2006) Genetic diversity of poplar leaf rust populations in the north-central United States. *Canadian Journal of Forest Research*, **36**, 2047-2057.
- Corander J, Waldmann P, Marttinen P, Sillanpää MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics*, **20**, 2363-2369.
- de Meeûs T, McCoy KD, Prugnolle F, Chevillon C, Durand P, Hurtrez-Bousses S, Renaud F (2007) Population genetics and molecular epidemiology or how to "debusquer la bête". *Infection Genetics and Evolution*, **7**, 308-332.
- De Mita S, Ronfort J, McKhann HI, Poncet C, El Malki R, Bataillon T (2007) Investigation of the demographic and selective forces shaping the nucleotide diversity of genes involved in nod factor signaling in *Medicago truncatula*. *Genetics*, **177**, 2123-2133.
- Dean RA, Talbot NJ, EbboleDJ, *et al.* (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*, **434**, 980-986.
- Desprez-Loustau ML, Robin C, Buee M, Courtecuisse R, Garbaye J, Suffert F, Sache I, Rizzo DM (2007) The fungal dimension of biological invasions. *Trends in Ecology & Evolution*, **22**, 472–480.
- Duan X, Tellier A, Wan A, Leconte M, de Vallavielle-Pope C, Enjalbert J (2010) *Puccinia striiformis* f.sp. *tritici* presents high diversity and recombination in the over-summering zone of Gansu, China. *Mycologia*, **102**, 44-53.
- Duplessis S, Cuomo CA, Lin Y-C, *et al.* (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proceedings of the National Academy of Sciences*. Doi : 10.1073/pnas.1019315108
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-2620.
- Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, **7**, 745-758.
- Flor HH (1971) Current Status of the Gene-For-Gene Concept. *Annual Review of Phytopathology*, **9**, 275-296.
- François O, Durand E (2010) THE STATE OF THE FIELD: Spatially explicit Bayesian clustering models in population genetics. *Molecular Ecology Resources*, **10**, 773-784.

- Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493-505.
- Frey P, Gérard PR, Feau N, Husson C, Pinon J (2005) Variability and population biology of *Melampsora* rusts on poplars. In Rust diseases of Willow and Poplar (eds. Pei MH, McCracken AR) pp. 63-72. CAB International, Wallingford, UK.
- Fry WE (2008) *Phytophthora infestans*, the crop (and *R* gene) destroyer. *Molecular Plant Pathology*, **9**, 385-402.
- Gérard PR, Husson C, Pinon J, Frey P (2006) Comparison of genetic and virulence diversity of *Melampsora larici-populina* populations on wild and cultivated poplar and influence of the alternate host. *Phytopathology*, **96**, 1027–1036.
- Giraud T, Gladieux P, Gavrillets S (2010) Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology & Evolution*, **25**, 387-395.
- Gladieux P, Byrnes E, Fisher M, Aguileta G, Heitman J, Giraud T (2010) Epidemiology and evolution of fungal pathogens, in plants and animals. In Genetics and Evolution of Infectious Diseases (Ed Tibayrenc M. Elsevier).
- Gladieux P, Zhang XG, Afoufa-Bastien D, Sanhueza RMV, Sbaghi M, Le Cam B (2008) On the Origin and Spread of the Scab Disease of Apple: Out of Central Asia. *Plos One*, **3**, e1455.
- Goss EM, Larsen M, Chastagner GA, Givens DR, Grunwald NJ (2009) Population Genetic Analysis Infers Migration Pathways of *Phytophthora ramorum* in US Nurseries. *Plos Pathogens*, **5**, 12.
- Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485-486.
- Goyeau H, Halkett F, Zapater MF, Carlier J, Lannou C (2007) Clonality and host selection in the wheat pathogenic fungus *Puccinia triticina*. *Fungal Genetics and Biology*, **44**, 474-483.
- Groth JV, McCain JW, Roelfs AP (1995) Virulence and isozyme diversity of sexual versus asexual collections of *Uromyces-Appendiculatus* (Bean rust fungus). *Heredity*, **75**, 234-242.
- Guérin F, Gladieux P, Le Cam B (2007) Origin and colonization history of newly virulent strains of the phytopathogenic fungus *Venturia inaequalis*. *Fungal Genetics and Biology*, **44**, 284-292.

- Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261-1280.
- Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. *Molecular Ecology*, **18**, 4734-4756.
- Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712-715.
- Guillot G, Santos F (2009) A computer program to simulate multilocus genotype data with spatially autocorrelated allele frequencies. *Molecular Ecology Resources*, **9**, 1112-1120.
- Halkett F, Simon JC, Balloux F (2005) Tackling the population genetics of clonal and partially clonal organisms. *Trends in Ecology & Evolution*, **20**, 194-201.
- Halkett F, Coste D, Platero GGR, Zapater MF, Abadie C, Carlier J (2010) Genetic discontinuities and disequilibria in recently established populations of the plant pathogenic fungus *Mycosphaerella fijiensis*. *Molecular Ecology*, **19**, 3909-3923.
- Hallatschek O, Nelson DR (2008) Gene surfing in expanding populations. *Theoretical Population Biology*, **73**, 158-170.
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.
- Hovmøller MS, Justesen AF, Brown JKM (2002) Clonality and long-distance migration of *Puccinia striiformis* f.sp. *tritici* in north-west Europe. *Plant Pathology*, **51**, 24-32.
- Huang YJ, Li ZQ, Evans N, Rouxel T, Fitt BDL, Balesdent MH (2006) Fitness cost associated with the AvrLm4 avirulence function in *Leptosphaeria maculans* (phoma stem canker of oilseed rape). *European Journal of Plant Pathology*, **114**, 77-89.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322-1332.
- Jarne P, Lagoda PJJ (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution*, **11**, 424-429.
- Jin Y, Szabo L, Carson M (2010). Century-old mystery of *Puccinia striiformis* life history solved with the identification of *Berberis* spp. as an alternate host. *Phytopathology*, **100**, 432-435.
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.
- Jorge V, Dowkiw A, Faivre-Rampant P, Bastien C (2005) Genetic architecture of qualitative

- and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytologist* **167**, 113-127.
- Kaeuffer R, Reale D, Coltman DW, Pontier D (2007) Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. *Heredity*, **99**, 374-380.
- Laine AL, Tellier A (2008) Heterogeneous selection promotes maintenance of polymorphism in host-parasite interactions. *Oikos*, **117**, 1281–1288.
- Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295-302.
- Leach JE, Cruz CMV, Bai JF, Leung H (2001) Pathogen fitness penalty as a predictor of durability of disease resistance genes. *Annual Review of Phytopathology*, **39**, 187–224.
- Leblois R, Rousset F, Estoup A (2004) Influence of Spatial and Temporal Heterogeneities on the Estimation of Demographic Parameters in a Continuous Population Using Individual Microsatellite Data. *Genetics*, **166**, 1081-1092.
- Legionnet A, Muranty H, Lefevre F (1999) Genetic variation of the riparian pioneer tree species *Populus nigra*. II. Variation in susceptibility to the foliar rust in *Melampsora larici-populina*. *Heredity*, **82**, 318–327.
- Lepoivre P (2003) Phytopathologie. DeBoeck & Larcier, Bruxelles. 427 p.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981-994.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189-197.
- Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, **20**, 136-142.
- Mboup M, Leconte M, Gautier A, Wan AM, Chen W, de Vallavieille-Pope C, Enjalbert J (2009) Evidence of genetic recombination in wheat yellow rust populations of a Chinese oversummering area. *Fungal Genetics and Biology*, **46**, 299-307.
- McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, **40**, 349-+.
- Milgroom MG (2001) The synthesis of genetics and epidemiology: Contribution of population biology in plant pathology. *Journal of Plant Pathology*, **83**, 57-62.

- Milgroom MG, Peever TL (2003) Population Biology of Plant Pathogens: The Synthesis of Plant Disease Epidemiology and Population Genetics. *Plant Disease*, **87**, 608-617.
- Montarry J, Corbiere R, Lesueur S, Glais I, Andrivon D (2006) Does selection by resistant hosts trigger local adaptation in plant-pathogen systems? *Journal of Evolutionary Biology*, **19**, 522-531.
- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *Plos Genetics*, **2**, 2074-2093.
- Pinon J (1980) Etude de la survie des urédospores de *Melampsora larici-populina* Kleb. et de *M. allii-populina* Kleb. *Annales des Sciences Forestières*, **37**, 85-89.
- Pinon J, Frey P (1997) Structure of *Melampsora larici-populina* populations on wild and cultivated poplar. *European Journal of Plant Pathology*, **103**, 159-173.
- Pinon J, Frey P, Husson C, Schipfer A (1998) Poplar rust (*Melampsora larici-populina*): The development of E4 pathotypes in France since 1994. Pages 57–64 in: Proceedings of the First IUFRO Rusts of Forest Trees Conference, Saariselkä, Finland. R Jalkanen, PE Crane, JA Walla, T Aalto, eds. Finnish Forest Research Institute, Research Paper 712.
- Pinon, J, Frey, P, (2005) Interactions between poplar clones and *Melampsora larici-populina* populations and their implications for breeding for durable resistance. In Rust diseases of Willow and Poplar (eds. Pei MH, McCracken AR) pp. 139–154. CAB International, Wallingford, UK.
- Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536-539.
- Prakash CS, Thielges BA (1987) Pathogenic variation in *Melampsora medusae* leaf rust of poplars. *Euphytica* **36**, 563-570.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.
- Prospero S, Lung-Escarmant B, Dutech C (2008) Genetic structure of an expanding *Armillaria* root rot fungus (*Armillaria ostoyae*) population in a managed pine forest in southwestern France. *Molecular Ecology*, **17**, 3366-3378.
- Riou-Nivert P (2001) Le mélèze. Institut pour le Développement Forestier, 144 pp.
- Rivas G-G, Zapater M-F, Abadie C, Carlier J (2004) Founder effects and stochastic dispersal at the continental scale of the fungal pathogen of bananas *Mycosphaerella fijiensis*. *Molecular Ecology*, **13**, 471-482.

- Rousset F (1997) Genetic Differentiation and Estimation of Gene Flow from F-Statistics Under Isolation by Distance. *Genetics*, **145**, 1219-1228.
- Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58-62.
- Rowe G, Beebee TJC (2007) Defining population boundaries: use of three Bayesian approaches with microsatellite data from British natterjack toads (*Bufo calamita*). *Molecular Ecology*, **16**, 785-796.
- Sache I (2003) L'épidémiologie. In *Phytopathologie* (eds. Lepoivre P) pp. 193-213. DeBoeck & Larcier, Bruxelles.
- Samils B, Stepien V, Lagercrantz U, Lascoux M, Gullberg U (2001) Genetic diversity in relation to sexual and asexual reproduction in populations of *Melampsora larici-epitea*. *European Journal of Plant Pathology*, **107**, 871-881.
- Selkoe KA, Toonen RJ (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615-629.
- Shoemaker JS, Painter IS, Weir BS (1999) Bayesian statistics in genetics - a guide for the uninitiated. *Trends in Genetics*, **15**, 354-358.
- Stokstad E (2007) Deadly Wheat Fungus Threatens World's Breadbaskets. *Science*, **315**, 1786-1787.
- Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP (2007) Putting the 'landscape' in landscape genetics. *Heredity*, **98**, 128-142.
- Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology*, **19**, 3496-3514.
- Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, **46**, 75-100.
- Stukenbrock EH, McDonald BA (2009) Population Genetics of Fungal and Oomycete Effectors Involved in Gene-for-Gene Interactions. *Molecular Plant-Microbe Interactions*, **22**, 371-380.
- Sunnucks P (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution*, **15**, 199-203.
- Taris B (1968) Contribution à l'étude des rouilles des *Populus* observées en France. *Annales des Epiphyties*, **19**, 5-54.
- Taylor J, Jacobson D, Fisher M (1999) The evolution of asexual fungi: Reproduction, Speciation and Classification. *Annual Review of Phytopathology*, **37**, 197-246.

Bibliographie

- Vaha JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63-72.
- Vialle A, Frey P, Bernier L, Hamelin RC Historical review of poplar rust systematics and refinement of *Melampsora* species delineation. *Soumis à Fungal Diversity*.
- Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, **34**, 275-305.
- Walker J, Hartigan D, Bertus AL (1974) Poplar rusts in Australia with comments on potential conifer rusts. *European Journal of Forest Pathology*, **4**, 100-118.
- Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419-1439.
- Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity*, **82**, 117-125.
- Wilkinson AG, Spiers AG (1976) Introduction of poplar rusts *Melampsora larici-populina* and *Melampsora medusae* to New-Zealand and their subsequent distribution. *New Zealand Journal of Science*, **19**, 195-198.
- Wright S (1943) Isolation by distance. *Genetics*, **28**, 114-138.

Influence des processus démographiques sur la structure et les caractéristiques génétiques des champignons phytopathogènes, cas de l'agent de la rouille du peuplier *Melampsora larici-populina*

Résumé

La structure génétique et la dynamique des populations des champignons phytopathogènes peuvent être influencées par de nombreux facteurs, ce que nous avons illustré à différentes échelles chez *Melampsora larici-populina*, l'agent de la rouille du peuplier. A l'échelle de la France, deux principaux groupes génétiques se différencient. Le premier, inféodé aux hôtes sauvages, est le produit de l'évolution des populations en conditions naturelles. Le second, formé suite au contournement de la résistance R7 portée par le cultivar 'Beaupré', se caractérise par une forte proportion d'individus virulents 7 et présente des signatures de sélection et d'expansion démographique, témoins de l'invasion de ce groupe sur tous les peupliers (y compris sauvages) de la moitié nord de la France. A une échelle plus restreinte, nous avons examiné une zone de contact située dans les Alpes, où la délimitation entre les groupes « cultivé » et « sauvage » était la plus marquée. En testant l'effet du paysage, nous avons montré que le massif des Ecrins protégeait le groupe « sauvage » situé à l'est, en amont de la vallée de la Durance, de l'invasion des individus au profil « cultivé » venant du nord-ouest. Il n'en est pas de même dans la partie aval de cette vallée, colonisée annuellement par une vague épidémique, issue de ces deux groupes génétiques présents en amont de part et d'autre du Massif des Ecrins. En dernier lieu nous avons examiné les conséquences génétiques des événements de colonisation, qui se traduisent par une augmentation de la différenciation par rapport à la source de l'épidémie ainsi qu'une érosion de la diversité génétique. Ce travail original a permis de souligner l'importance de combiner les approches d'épidémiologie et de génétique des populations pour caractériser au mieux les processus démographiques et leurs conséquences génétiques.

Mots clés : *Melampsora larici-populina*, champignon phytopathogène, génétique des populations, épidémiologie, démographie, structure génétique spatiale.

Impact of demographic processes on population structure and genetic characteristics of fungal plant pathogens: a case study with the poplar rust fungus *Melampsora larici-populina*

Abstract:

Many factors can impact the genetic structures and population dynamics of fungal plant pathogens. Here we illustrated some of them at different spatial scales for the poplar rust fungus *Melampsora larici-populina*. At the scale of France, two main genetic groups were found. The first one infects only wild hosts and results from natural evolution of rust populations. The second one was formed after the R7 resistance breakdown, which is carried by the cultivar 'Beaupré'. This group exhibited a high proportion of virulent 7 individuals and presented signs of selection and demographic expansion; these signs indicate the recent invasion of individuals from this group on both wild and cultivated poplars in the northern half of France. At a regional scale, we focused on a contact zone between the 'cultivated' and the 'wild' genetic groups, located in the Alps. Upstream the Durance River valley, the influence of landscape has been highlighted by the effect of the Ecrins range which protects the 'wild' group located on the east side from the invasion of individuals from the 'cultivated' group, which arise from the northwest. Downstream the valley the annual epidemic wave was shown to be composed of admixed individuals from 'wild' and 'cultivated' groups, originating from both sides of the Ecrins range. Lastly we assessed the genetic consequences of the colonization wave. We evidenced a gradual increase of genetic differentiation with the epidemic source and a loss of genetic diversity. This work highlights the need of combining population genetics and epidemiology to characterize demographic processes and their genetic consequences.

Keywords: *Melampsora larici-populina*, fungal plant pathogen, population genetics, epidemiology, demography, spatial genetic structure