

Thèse de Doctorat

Thibault Leroy

Mémoire présenté en vue de l'obtention
du grade de Docteur de l'Université d'Angers
Sous le label de l'Université Nantes Angers Le Mans

Discipline : Biologie

Spécialité : Biologie des organismes

**Laboratoire : Ecofun, IRHS, (Université d'Angers,
INRA, Agrocampus-Ouest)**

Soutenue le 13 décembre 2012

École doctorale : VENAM
Thèse N° 1261

Conséquences évolutives d'un contact secondaire chez les pathogènes: une alternative aux hypothèses écologiques?

Mise en évidence d'un contact secondaire avec flux de gènes
chez *V. inaequalis*: conséquences génétiques et épidémiologiques.

JURY

Rapporteurs :	Mme Tatiana GIRAUD , Directrice de recherche, CNRS, Orsay M. Pierre BOURSOT , Directeur de recherche, CNRS, Montpellier
Examineurs :	M. Vincent CASTRIC , Chargé de recherche, CNRS, Lille M. Didier ANDRIVON , Directeur de recherche, INRA, Rennes
Directeur de Thèse :	M Bruno LE CAM , Chargé de recherche, INRA, Angers
Co-encadrant de Thèse :	M. Christophe LEMAIRE , Maître de conférences, Université d'Angers

TABLE DES MATIERES

INTRODUCTION GENERALE	7
INTRODUCTION BIBLIOGRAPHIQUE	10
1. EVOLUTION DE L'ISOLEMENT REPRODUCTIF	11
1.1 Divergence & intensité du flux génique	11
1.2 Bases génétiques de l'isolement reproductif	11
1.2.1 Barrières pré-zygotiques	12
1.2.2 Barrières post-zygotiques	13
1.3 Zone hybrides et contact secondaire	18
1.4 Etude des barrières/Interprétations des scans génomiques	22
1.4.1 Outliers non-affectés par la sélection naturelle (Faux positifs)	22
1.4.2 Outliers affectés par la sélection naturelle	23
2. EVOLUTION DES POPULATIONS DE PATHOGENES	25
2.1 Modèles théoriques : forte adaptation à l'habitat attendue	25
2.2 Mécanismes évolutifs conduisant aux émergences	29
2.3 La remise en contact secondaire des populations de pathogènes : une hypothèse négligée	31
3. LES HYPOTHESES NON-ECOLOGIQUES : UN NOUVEAU CADRE CONCEPTUEL A TRANSPOSER CHEZ LES PATHOGENES	32
3.1 Théorie des clines	33
3.1.1 Clines géographiques	33
3.1.2 Clines génomiques	34
3.2 Détection des barrières par cartographie génétique	35
3.3 Conséquences potentielles des remises en contact secondaire pour les populations de pathogènes	35
3.3.1 Hybridation & paysage adaptatif	35
3.3.2 Importance d'intégrer ces hypothèses alternatives	37
4. PATHOSYSTEME <i>VENTURIA INAEQUALIS</i>/ POMMIER	41
4.1 Pommier domestique	41
4.1.1 Origine du pommier cultivé	41
4.1.2 Biologie et génétique du pommier	42
4.1.3 Importance économique	43
4.2 La tavelure du pommier	43
4.2.1 Importance économique	43
4.2.2 Agent pathogène : <i>V. inaequalis</i>	44
4.3 Résistance génétique du pommier à <i>V. inaequalis</i>	50
4.3.1 Résistance naturelle : large gamme de gènes de résistance	50
4.3.2 Rvi6, une résistance majeure utilisée dans les programmes de sélection du pommier	50
4.4 Questions de recherche	55

RESULTATS	56
CHAPITRE 1	56
1. Article 1	57
Finding ecological divergence in pathogens: a needle in a haystack?	57
2. Résultats complémentaires	78
2.1 Conséquences des DMI sur cartes génétiques : une approche par simulation	78
2.1.1 Description de l'approche utilisée	78
2.1.2 Résultats et discussion	79
2.2 Génotypage des outliers en populations naturelles	81
2.2.1 Déséquilibre de liaison entre les SNPs proches des outliers	81
2.2.2 Génotypage dans de nouvelles populations	82
2.3 Barrières post-zygotiques de type DMI et gènes candidats	84
2.3.1 Hypothèses alternatives aux DMI	84
2.3.2 Caractérisations des zones génomiques et gènes candidats	84
CHAPITRE 2	90
Evolution of pathogenic traits through secondary contact with gene flow between divergent pathogen populations: an unsuspected threat for agro-ecosystems.	90
DISCUSSION	112
1. Emergences à partir du compartiment non-cultivé	113
2. Contacts secondaires et importance des barrières post-zygotiques	114
2.1 Contact secondaire avec flux génique entre des populations avrRvi6 et virRvi6	114
2.2 Barrières génétiques post-zygotiques endogènes chez les pathogènes	115
2.3 Particularité des champignons pathogènes	118
3. Asymétrie de l'introgession et conséquences épidémiologiques	119
3.1 Sens de l'introgession du fond génétique	119
3.2 Introgession des facteurs d'agressivité et conséquences épidémiologiques	120
4. Réflexions sur la durabilité des résistances	120
CONCLUSION ET PERSPECTIVES	123
1. Conclusion	124
2. Perspectives	126
2.1 A court terme	126
2.2 A plus long terme	126
BIBLIOGRAPHIE	128
ANNEXE	151

INTRODUCTION GENERALE

La classification de la biodiversité en espèces, c'est-à-dire en entités discrètes, est d'une importance majeure en biologie. Bien que cette notion soit fondamentale, le terme espèce est toutefois ambigu et regroupe un large éventail d'une vingtaine de concepts différents (Mayden 1997; Coyne & Orr 2004). Chez les organismes eucaryotes à reproduction sexuée, la délimitation des contours des espèces la plus simple est basée sur la notion d'interfertilité. Cette notion est étroitement associée au concept biologique de l'espèce définie par Mayr (1942; 1963; 1970). Sous ce concept, l'espèce est définie comme un ensemble de populations naturelles, interfécondes ou potentiellement interfécondes génétiquement isolées d'autres groupes présentant les mêmes caractéristiques d'interfécondité. Cette définition est simple et permet de mettre en évidence les processus évolutifs conduisant à la mise en place de la discrimination de ces entités, c'est-à-dire l'établissement d'un isolement reproductif entre deux taxons.

A ce titre, l'hybridation fut longtemps considérée comme un problème majeur associé au concept biologique d'espèce. En effet, l'existence de génotypes intermédiaires s'oppose directement à une classification binaire des espèces. Pourtant, la définition de l'hybridation est intimement liée à cette définition car un isolement reproductif, même presque complet, peut autoriser la production d'hybrides en faible fréquence. Ce flux génique ne se traduit toutefois pas nécessairement par une altération des génomes parentaux hors de la zone d'hybridation. En théorie, les zones d'hybridation sont le résultat d'une remise en contact suivie ou non de flux de gènes après une divergence entre les taxons. Ce phénomène est souvent considéré à tort comme un phénomène biologique relativement anecdotique. Pourtant, dans une revue relativement exhaustive réalisée par Barton & Hewitt (1985), une centaine de zones d'hybridation ont été décrites et sont réparties dans de nombreux groupes d'espèces vivants.

En théorie, c'est dans les situations d'hybridation entre taxons qu'il est souvent possible de déterminer les mécanismes responsables de l'isolement reproductif. En effet, cet isolement reproductif peut être dû à la mise en place de barrières de nature pré-zygotique et post-zygotiques. Soit elles sont endogènes et accumulées pendant la divergence, générant une contre-sélection des hybrides indépendante de l'environnement ; soit elles sont exogènes et liées aux caractéristiques de l'habitat. L'ensemble des barrières génétiques va contribuer à maintenir une différenciation durable des populations pouvant conduire à la spéciation (p. ex. Dobzhansky 1937 ; Coyne & Orr 2004). Alternativement, la recombinaison des deux pools génétiques ayant divergé peut être vue comme un formidable moteur de diversité, générant des combinaisons génétiques nouvelles n'ayant jamais été soumises à la sélection naturelle au cours des précédentes étapes de l'évolution. Exceptionnellement, certaines de ces nouvelles combinaisons peuvent conférer un avantage adaptatif aux hybrides dans des environnements intermédiaires ou même parfois être plus favorables dans les environnements parentaux (Moore 1977). Bien évidemment, cette dichotomie ne représente que des cas extrêmes où l'hybridation est clairement défavorable ou favorable au flux génique et à la mise en place de nouvelles combinaisons génétiques. La plupart des zones d'hybridation sont caractérisées par des barrières semi-perméables au flux génique avec des échanges plus ou moins efficaces en fonction des régions génomiques.

Les conséquences évolutives de ces remises en contact secondaire n'ont reçu qu'un intérêt mesuré en phytopathologie où les hypothèses de coadaptation hôtes-pathogènes sont les plus souvent invoquées pour expliquer la diversification des pathogènes (Thrall & Burdon 2003; Jones & Dangl 2006). Si quelques exemples de contact secondaire concernent l'impact des zones hybrides de l'hôte sur leur pathogène associé [nématodes infectant la souris (Mouliat *et al.* 1993), ou helminthes des grenouilles vertes (Planade *et al.* 2009)], les zones d'hybridation des pathogènes n'ont pas fait l'objet d'études approfondies. Pourtant les conséquences de ces hybridations peuvent être particulièrement néfastes pour les

agrosystèmes si des allèles contribuant à l'augmentation de la valeur sélective (virulence¹, facteurs d'agressivité, gènes de résistance aux xénobiotiques) sont introgressés entre populations de pathogènes risquant d'ouvrir la voie à de nouvelles émergences.

C'est dans ce cadre conceptuel que s'est déroulé mon travail de recherche. L'objectif général est de caractériser les barrières génétiques et leur intensité dans la rupture du flux génique entre deux génomes divergents de *Venturia inaequalis*, l'agent responsable de la tavelure du pommier (Guerin & Le Cam 2004 ; Gladieux *et al.* 2011). L'objectif de cette thèse est d'une part de renseigner les mécanismes d'isolement reproductif, c'est-à-dire les barrières génétiques responsables du maintien de la différenciation génétique entre ces deux populations divergentes et d'autre part d'évaluer les conséquences épidémiologiques d'un flux génique entre ces populations divergentes pour le pommier domestique *Malus x domestica*.

L'étude bibliographique se présente en quatre grandes parties. La première partie s'intéressera aux différentes barrières génétiques et aux contacts secondaires, la seconde décrira les hypothèses courantes de l'évolution des populations des pathogènes. La troisième partie présentera les hypothèses alternatives non-écologiques chez les pathogènes, la dernière décrira le pathosystème étudié : *V. inaequalis* – *Malus* sp.

¹ Notons qu'au cours de cette thèse, nous utiliserons la définition qualitative de la virulence dans son sens habituel en phytopathologie, c'est-à-dire la capacité ou non d'une souche à infecter un hôte donné. Cette définition sera à opposer à la mesure quantitative de cette infection par la suite dénommée agressivité.

INTRODUCTION BIBLIOGRAPHIQUE

1. EVOLUTION DE L'ISOLEMENT REPRODUCTIF

1.1 Divergence & intensité du flux génique

L'une des questions essentielles de la biologie évolutive est de savoir comment apparaît l'isolement reproductif entre des populations qui divergent dans l'espace et le temps jusqu'à former des taxons distincts (Coyne & Orr 2004). Un paramètre déterminant dans la mise en place de l'isolement reproductif est l'intensité du flux génique car celui-ci tend à homogénéiser les populations et ainsi s'oppose directement à la fixation de différences nucléotidiques, par dérive génétique ou par sélection naturelle (Mayr 1963; Slatkin 1987).

Mayr (1942; 1963) a classifié les modes de divergence en tenant essentiellement compte de l'intensité du flux génique entre deux pools génétiques. Mayr a proposé une échelle biogéographique basée sur le degré de chevauchement des aires de répartition des populations. Trois grandes situations sont possibles : (i) soit les populations ont des aires de répartition différentes (allopatrie), (ii) soit elles présentent des aires partiellement recoupées (parapatricie), (iii) soit elles vivent sur une seule et même aire de répartition (sympatricie). Dans les modèles théoriques, ces trois situations sont exprimées en fonction de l'intensité du flux génique m variant de manière continue entre 0 et 0,5. Les valeurs extrêmes de 0 et 0,5 correspondent respectivement à une absence totale de migrants (allopatricie) et un libre échange de migrants (sympatricie) entre les deux populations. Les valeurs intermédiaires sont considérées comme autant de cas de divergence parapatricie. Bien que la plupart des modèles ont été construits en prenant en compte cette distinction, l'idée selon laquelle les espèces puissent être totalement allopatricies ou totalement sympatricies est aujourd'hui contestée (Butlin *et al.* 2008). En effet, des événements de migration entre deux populations géographiquement isolées peuvent quasiment toujours être attendus, même à des fréquences rares. Inversement, pour la plupart des espèces, la dispersion limitée des espèces et la répartition généralement mosaïque de leurs habitats (Endler 1977) sont peu compatibles avec une complète sympatricie. Selon Butlin *et al.* (2008), cette classification ne peut donc conduire qu'à des attendus de différenciation parapatricie. Or, intrinsèquement ces derniers attendus sont aussi les moins bien formulés car (i) ils représentent une très large gamme d'intensité du flux génique, de quasi-nul à quasi-complet et (ii) les modèles théoriques intègrent différemment la parapatricie en la rapprochant des modèles d'allopatricie (Turelli *et al.* 2001; Barton 2010) ou de sympatricie (Kirkpatrick & Ravigné 2002; Barton & de Cara 2009).

La classification biogéographique semble de plus en plus délaissée au profit d'une meilleure intégration des processus autorisant la divergence et l'apparition de l'isolement reproductif. En théorie, si la migration efficace est faible ($2Nm < 1$), l'isolement reproductif peut être mis en place par le seul rôle de la dérive génétique. Ce processus, bien qu'inévitable, peut être particulièrement long surtout si la taille des populations est grande. La sélection naturelle peut toutefois fortement accélérer ce processus (Gavrilets 1997b; Turelli *et al.* 2001; Gavrilets 2003). *A contrario*, si la migration est supérieure à la dérive génétique ($2Nm > 1$), l'accumulation de barrières génétiques ne peut résulter que d'un processus déterministe nécessitant une forte contribution de la sélection naturelle (Gavrilets 1997b).

1.2 Bases génétiques de l'isolement reproductif

On appellera barrière génétique au flux génique, tous les locus impliqués dans les mécanismes d'isolement reproductif. Les différents mécanismes d'isolement reproductif sont généralement regroupés selon le stade du cycle de vie où elles interviennent (Figure 1), permettant ainsi de distinguer les barrières au flux génique agissant avant la formation du

zygote (pré-zygotiques) ou après (post-zygotiques). Nous verrons par la suite comment ces deux types de barrière contribuent à la réduction du flux de gènes. Précisons, d'ores et déjà, que ces deux grands types de barrière ne sont pas mutuellement exclusives et participent souvent toutes les deux au maintien de la différenciation des pools génétiques (Szymura & Barton 1986; Mallet *et al.* 1990; Bert & Arnold 1995). Les mécanismes d'isolement reproductif peuvent être aussi distingués en fonction de leur dépendance à l'environnement (barrières dites exogènes ou écologiques) ou de leur indépendance (barrières dites endogènes). Cette dernière classification est un point extrêmement important car la contribution de l'écologie dans l'isolement reproductif des populations est encore largement débattue (cf § 1.4).

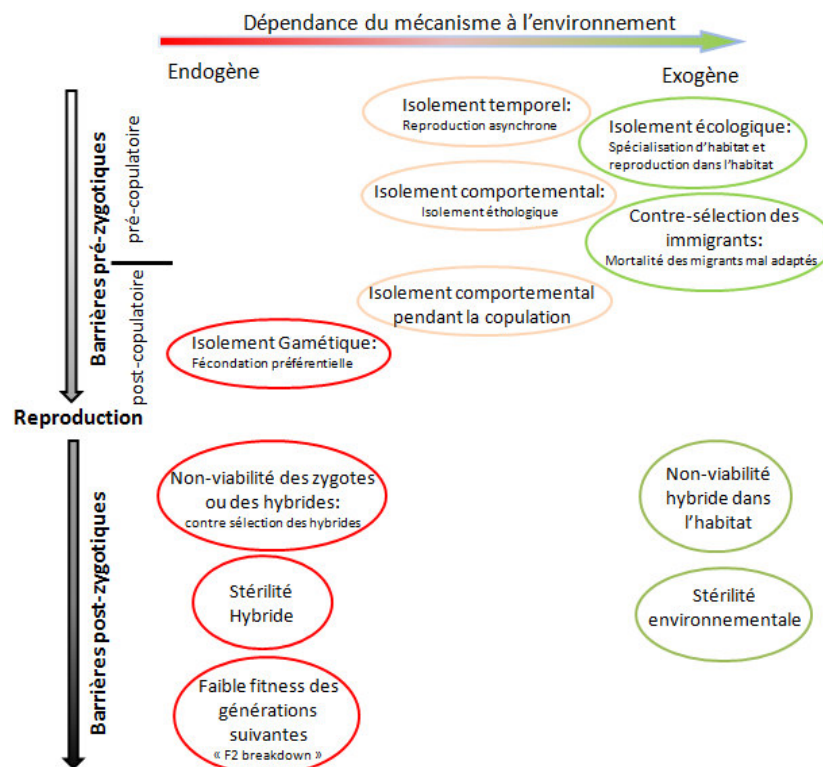


Figure 1 : classification des barrières d'isolement reproductif potentielles. Modifié d'après Bierne (2001), Coyne & Orr (2004) et Nosil *et al.* (2005).

1.2.1 Barrières pré-zygotiques

L'adaptation à des habitats différents n'est généralement possible que si les combinaisons génétiques favorables sont différentes et dépendantes des habitats. Les allèles favorables à un habitat sont donc défavorables dans l'autre habitat. La grande difficulté de ce modèle est alors de déterminer comment le couplage entre les locus d'adaptation et d'accouplement préférentiel est possible et empêche l'homogénéisation des deux génomes par la recombinaison.

Les travaux théoriques sur la spéciation sympatrique de Maynard Smith (1966) et Felsenstein (1981) suggèrent que les conditions nécessaires à l'évolution de l'isolement reproductif en présence de flux géniques sont particulièrement restrictives. Felsenstein (1981) a proposé deux modèles pour expliquer le choix actif de partenaires : un modèle à un allèle et un modèle à deux allèles. Dans le modèle à un allèle, un nouvel allèle se fixe dans l'ensemble

des populations et permet aux individus de se reconnaître et de se croiser préférentiellement avec les génotypes adaptés. La préférence sexuelle pour les individus proches est alors sous la dépendance d'un seul allèle qui réduit par exemple la période de reproduction. Si les populations diffèrent dans leur période de reproduction, le nouvel allèle peut alors augmenter l'isolement pré-zygotique. Dans le modèle à deux allèles, les individus porteurs d'un allèle (disons A1) se croisent préférentiellement entre eux et de la même façon pour les individus porteurs d'un second allèle (A2). Sous ce modèle, la divergence est possible si les allèles sont associés de manière non aléatoire à la préférence d'habitat. Par conséquent, le modèle à deux allèles nécessite soit un fort déséquilibre de liaison (par exemple, physique ou *via* des réarrangements chromosomiques) soit une pléiotropie entre les allèles conférant le choix du partenaire et l'adaptation locale (Rice 1984). Ce second modèle est souvent considéré comme le modèle le plus probable (Felsenstein 1981) mais quelques exemples semblent plutôt en accord avec le modèle à un allèle (Kirkpatrick & Ravigné 2002). En effet, dans ce modèle, la recombinaison ne joue pas un rôle homogénéisant. Le renforcement de l'isolement reproductif évolue donc plus facilement en présence de flux génique. Malheureusement, les données empiriques concernant ces deux modèles de spéciation sont très limitées (mais voir les revues d'Ortiz-Barrientos *et al.* 2002 et Servedio & Noor 2003 pour quelques exemples).

D'autres modèles ont été développés, le modèle 'sans gène' considère que l'accouplement préférentiel avec les individus adaptés peut être un sous-produit de l'adaptation locale (Rice 1984; Kondrashov 1986; Doebli 1996). Par exemple, chez l'épinoche (*Gasterosteus aculeatus*), la taille du corps contribuerait à une meilleure adaptation locale. Certains auteurs ont proposé que cette modification du phénotype induise un choix actif de partenaire (Nagel & Schluter 1998; Boughman *et al.* 2005).

1.2.2 *Barrières post-zygotiques*

Historiquement, les barrières d'isolement post-zygotique ont été décrites plus tôt que les barrières pré-zygotiques. Ces barrières d'isolement agissent après la formation des zygotes et se traduisent par une moindre viabilité et/ou fertilité des hybrides

1.2.2.1 Barrières post-zygotiques endogènes

Les modèles expliquant la mise en place des barrières post-zygotiques endogènes sont nombreux (Nei & Nozawa 2011), mais peu sont soutenus empiriquement. Seuls les trois grands modèles détaillés ci-dessous sont bien supportés par des observations. Notons que les autres modèles peuvent être interprétés comme des extensions au modèle de Dobzhansky-Muller (§ 1.2.2.1.3) comme par exemple, le modèle des gènes complémentaires multialléliques de Nei *et al.* (1983) (Nei & Nozawa 2011).

1.2.2.1.1 Modèle de sous-dominance des hybrides

Les remaniements chromosomiques tels que les inversions, les translocations, les fissions ou les fusions (Rieseberg 2001) ou même la polyploïdisation (Otto & Whitton 2000) peuvent créer de la sous-dominance chez les hétérozygotes [voir aussi la revue de Nei & Nozawa (2011)]. L'isolement post-zygotique est basé sur la réduction de la valeur sélective des hétérozygotes comparée aux génotypes parentaux. Sous ce modèle, les croisements entre deux populations fixées pour deux allèles différents à un locus sous-dominant produiront une dépression hybride dès la première génération. En effet, les remaniements chromosomiques sont responsables d'une production de gamètes déséquilibrés dans leur contenu en gènes,

c'est-à-dire présentant des duplications ou des délétions de gènes (Figure 2).

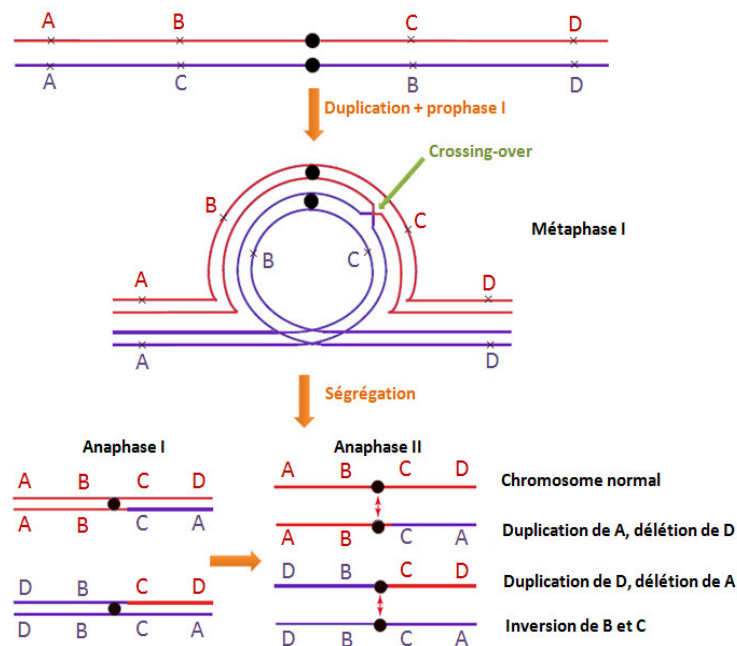


Figure 2 : Modèle de sous-dominance des hybrides par réarrangements chromosomiques. Un génotype possède une combinaison génétique classique (ABCD) tandis que le second a subi une inversion chromosomique (ACBD). Les événements de crossing-over entre ces deux combinaisons génétiques vont produire des gamètes déséquilibrés ne possédant pas l'ensemble des gènes. La perte de ces locus affectera alors la viabilité des gamètes. Modifié d'après Rieseberg (2001).

Sous ce modèle de sous-dominance, la fixation de ces nouveaux allèles « mutés » est cependant difficile car les nouvelles mutations apparaissent à l'état hétérozygote et sont contre-sélectionnées. Par conséquent, la probabilité que ces deux populations mettent en place une telle barrière post-zygotique nécessite que l'une des deux populations traverse une vallée adaptative par dérive. Cette probabilité est très faible (Gavrilets 2003) et nécessite des conditions extrêmes de goulots d'étranglement. Par exemple, dans les cas de péripatrie, la divergence est la conséquence de la fondation d'une petite population en marge de son aire de répartition. Le paradoxe est alors que plus cette barrière post-zygotique est forte, moins il est vraisemblable que celle-ci se fixe (Barton & Rouhani 1991; Spirito 1992).

1.2.2.1.2 Modèle d'Oka

Un autre modèle simple explique la mise en place de la stérilité hybride par des barrières post-zygotiques endogènes [Oka (1953; 1957; 1974)]. La portée de ce modèle a été récemment remise en lumière (Werth & Windham 1991; Lynch & Force 2000; Nei & Nozawa 2011). Sous ce modèle, deux lignées génétiques divergent à partir d'une lignée génétique ancestrale en allopatrie porteuse de deux gènes dupliqués redondants (allèles A_0 et B_0). Considérons que l'allèle A_0 de la population 1 mute vers un allèle non-fonctionnel A_1 et l'allèle B_0 de la population 2 mute vers un allèle non-fonctionnel B_2 . A la suite de la remise en contact des populations, l'hybride ($A_0A_1B_0B_2$) va produire quatre gamètes (A_0B_0 , A_1B_0 , A_0B_2 et A_1B_2). Les gamètes portant les deux allèles A_1 et B_2 sont alors non-viables (Figure 3).

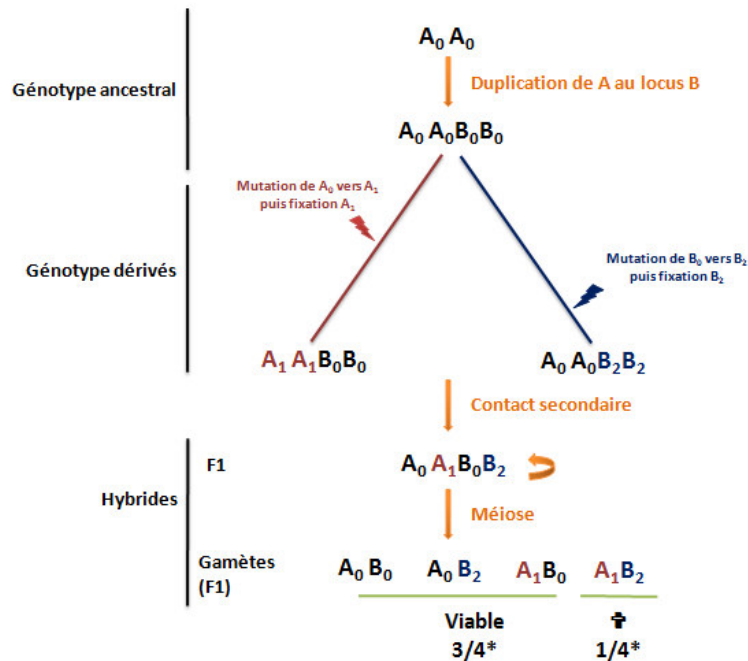


Figure 3 : Modèle de mise en place de barrières post-zygotiques selon Oka (1953; 1957; 1974). Chez un haplotype ancestral, un gène (A) est dupliqué sur une autre région du génome (B, ici on considère A et B indépendants). La dérive génétique de deux populations en allopatrie autorise la mutation vers des allèles non-fonctionnels et la fixation différentielle de ces allèles dans chaque population (A_0 vers A_1 dans la population 1 ; B_0 vers B_2 dans population 2). Sous le modèle d'Oka, seules les combinaisons contenant les deux allèles non-fonctionnels sont délétères, soit $1/4$ des gamètes de la F1 chez un organisme diploïde.

Un des grands avantages de ce modèle est qu'il permet d'expliquer la mise en place de barrières génétiques sans la nécessité de traverser des vallées adaptatives et ce chez les organismes porteurs de nombreux gènes dupliqués. Notons que le taux de mutation létale est un paramètre généralement mal connu. Chez la drosophile celui-ci serait voisin de 10^{-5} par génération rendant ce modèle particulièrement vraisemblable dans le cas des génomes polyplœides (Nei & Nozawa 2011). La récente mise en évidence de nombreux gènes dupliqués chez les espèces non-polyplœides (variation du nombre de copies), par exemple par Redon *et al.* (2006), suggère que ce modèle ne se résume pas aux cas de génomes polyplœides mais pourrait expliquer la mise en place de l'isolement reproductif chez de nombreuses espèces haploïdes ou diploïdes.

Ce modèle a reçu bien peu de soutien théorique, bien que soutenu par des évidences empiriques [(chez le riz Mizuta *et al.* (2010) ; Yamagata *et al.* (2010) ; chez la levure Scannell *et al.* (2006)] et malgré l'importance de la polyplœidisation chez les espèces végétales (2 à 4 % des plantes auraient ainsi spéciées ; Otto & Whitton, 2000), Par exemple, Coyne & Orr (2004) minimise l'importance de ce modèle en remarquant que très peu d'espèces animales sont polyplœides ce qui n'expliquerait donc pas leur divergence². Cependant, l'abondance des gènes dupliqués chez les espèces animales suggère que les barrières post-zygotiques décrites par ce modèle peuvent jouer un rôle important chez un grand nombre d'eucaryotes (Nei & Nozawa 2011).

² Des événements de polyplœidisation chez les espèces animales ont toutefois eu lieu au début de la divergence des vertébrés et ont pu avoir un rôle considérable dans les premières étapes de divergence des animaux (Ohno 1970, 1998; Hirsch *et al.* 2002; Jaillon *et al.* 2004).

1.2.2.1.3 Modèle de Dobzhansky-Muller

Un troisième modèle, de loin le plus populaire, a été proposé indépendamment par Dobzhansky (1937) puis par Muller (1942)³. De la même façon que sous le modèle d'Oka, ce modèle permet d'appréhender la mise en place d'un isolement post-zygotique intrinsèque sans nécessité de franchir les vallées adaptatives ou que les mutations soient délétères (Orr 1995). Ce modèle, dit de Dobzhansky-Muller, implique la fixation de plusieurs locus neutres ou avantageux dans leur fond génétique habituel mais délétères lorsqu'ils sont mélangés à d'autres allèles dérivés provenant d'un autre fond génétique. Dans sa forme la plus simple, la dépression hybride est alors due à des interactions épistatiques négatives entre deux locus bi-alléliques (A et B) (Figure 4).

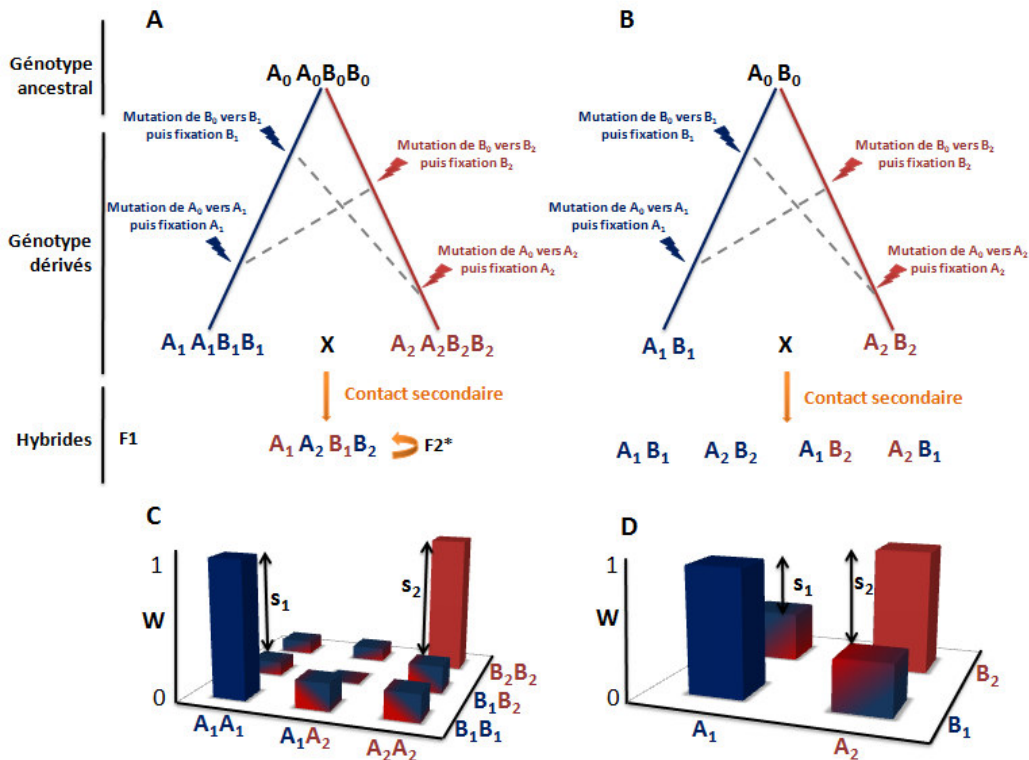


Figure 4 : Evolution de l'isolement reproductif sous le modèle de Dobzhansky-Muller pour un organisme diploïde (A) ou haploïde (B). A partir d'un génotype ancestral, des nouveaux allèles sont créés par mutation et se fixent par dérive génétique ou par sélection naturelle. Les allèles fixés peuvent alors être incompatibles entre les deux fonds génétiques. Ces incompatibilités, ici représentées par les lignes pointillées, sont alors révélées en cas de contact secondaire entre les deux populations. Les allèles d'incompatibilités sont ici considérés dominants et présentent des effets non-additifs. Les figures C et D représentent les valeurs sélectives associées aux génotypes de la F2 chez le diploïde et la F1 chez un haploïde et s_1 et s_2 représentent l'intensité de la sélection contre les hybrides. Notons que la distorsion de ségrégation chez un haploïde n'est observable que si $s_1 \neq s_2$. L'astérisque indique que la F2 n'est possible que si la génération F1 est viable, la production d'une F2 est donc rare si les allèles dérivés sont dominants (mais voir ci-dessous).

Considérons deux populations ayant divergé en allopatrie à partir d'un génotype ancestral fixé pour l'allèle A_0 au locus A et B_0 au locus B. Si un nouvel allèle (A_1) se fixe dans

³ Certains auteurs considèrent que Bateson était arrivé à un modèle similaire un quart de siècle plus tôt (Coyne & Orr 2004) et considèrent qu'il serait plus exact de dénommer ces barrières génétiques endogènes, des incompatibilités de Bateson-Dobzhansky-Muller. Cette dénomination est discutable car si Bateson a bien formulé des attendus de stérilité hybride *via* des interactions entre deux locus (Bateson 1909), les processus évolutifs pouvant conduire à un tel système n'étaient pas considérés par son modèle (Nei & Nozawa 2011).

la lignée génétique 1 et un nouvel allèle (B_2) se fixe dans la population 2. La formation d'hybrides lors d'un contact secondaire va permettre aux deux allèles dérivés de se retrouver combinés dans un même génotype. La valeur sélective des combinaisons porteuses des allèles A_1 et B_2 peut être alors réduite par rapport aux génotypes parentaux.

La vitesse de fixation des allèles A_1 et B_2 en allopatrie est une question primordiale pour la mise en place des incompatibilités de Dobzhansky-Muller (DMI pour la suite du texte). Soit une mutation neutre apparaissant par mutation chez un organisme diploïde, sa fixation est dépendante seule de la dérive génétique et le temps de remplacement de A_0 par A_1 est de $\frac{1}{\mu} + 2N$ où μ représente le taux de mutation et N la taille efficace de la population (Nei 1976). Ce temps nécessaire à la fixation des allèles incompatibles peut être donc particulièrement long par le simple effet de la dérive génétique si la taille efficace des populations est grande. Si la mutation est avantageuse pour une population, l'allèle va envahir la population beaucoup plus rapidement que par le simple effet de la dérive génétique. Pour des populations de petite taille, le temps de remplacement de l'allèle est alors approximativement égal au temps nécessaire à l'apparition de la mutation (Li & Nei 1977).

Chez les diploïdes, la dépression hybride est généralement plus forte dans les secondes générations d'hybridations (F2 ou rétrocroisements) qu'en F1 (Edmands 1999). En effet, si les allèles dérivés A_1 et B_2 sont récessifs et localisés sur des autosomes, la dépression hybride n'a lieu que chez les homozygotes ($A_1A_1B_2B_2$). Si les allèles sont portés par le chromosome sexuel, les barrières de type Dobzhansky-Muller vont avoir un effet plus fort chez le sexe hétérogamétique (XY ou WZ) que chez le sexe homogamétique conformément à la règle de Haldane (Haldane 1922; Wu & Davis 1993).

Des modèles mathématiques ont été développés pour prédire la vitesse de mise en place des incompatibilités de type Dobzhansky-Muller au cours du temps de divergence. Les prédictions issues de ces modèles suggèrent que les incompatibilités s'accumulent comme le carré de la divergence pour les incompatibilités à deux locus suggérant un effet d'emballement ou effet « boule de neige » (Orr 1995; Orr & Turelli 2001). Pour les incompatibilités impliquant trois locus, les incompatibilités s'accumuleraient comme le cube de la vitesse et ainsi de suite. Les études empiriques sont encore peu nombreuses, mais de récents résultats obtenus sur des espèces divergentes de *Drosophila* semblent valider cet attendu (Matute *et al.* 2011 mais voir aussi Gourbiere & Mallet 2010). Un résultat empirique similaire suggère aussi cet effet d'emballement dans l'accumulation des barrières génétiques de stérilité chez *Solanum* (Moyle & Nakazato 2010).

Les locus impliqués dans les DMI n'ont été identifiés que chez quelques espèces modèles telles que *Drosophila* (Sawamura & Yamamoto 1997; Presgraves *et al.* 2003; Brideau *et al.* 2006; Ferree & Barbash 2009; Phadnis & Orr 2009; Tang & Presgraves 2009), *Arabidopsis thaliana* (Bomblies *et al.* 2007), *Saccharomyces* (Lee *et al.* 2008; Chou *et al.* 2010), *Mus musculus* (Mihola *et al.* 2009). Si spontanément les DMI sont supposées dues à des mutations non-synonymes sur les gènes, notons que ces incompatibilités peuvent aussi être liées à une divergence dans des régions non-codantes (Barbash *et al.* 2003; Brideau *et al.* 2006; Maheshwari & Barbash 2011). Actuellement, les études empiriques sont trop peu nombreuses pour fixer un *a priori* sur les classes de gènes impliqués dans les DMI.

1.2.2.2 Barrières post-zygotiques exogènes

L'origine de l'isolement reproductif peut aussi être extrinsèque. Dans le cas de ces barrières, l'isolement reproductif est un sous-produit de l'adaptation locale (Dobzhansky 1937; Mayr 1963). Sous ce modèle, les hybrides produits par un croisement entre deux

populations différenciellement adaptées ne sont pas affectés intrinsèquement dans leur développement mais présentent une valeur sélective plus faible que leurs parents dans les deux habitats parentaux. La recombinaison peut générer de nouvelles combinaisons génétiques n'ayant jamais été préalablement testées par la sélection naturelle ou recréer des combinaisons transitoires dans l'évolution de ces populations.

Chez les lépidoptères *Heliconius melpomene* et *H. cydno*, deux espèces proches vivant en sympatrie en Amérique centrale, il existe des différences entre les ailes des deux espèces. L'une est colorée de noir, de rouge et de jaune et l'autre uniquement de noir et de blanc. Les hybrides peu fréquents présentent des ailes différentes des deux espèces parentales. On observe alors que la valeur sélective associée au phénotype hybride est réduite par rapport aux phénotypes parentaux. Un autre exemple a été décrit chez les pares d'espèces benthique et limnétique de l'épinoche (*Gasterosteus aculeatus*) (Hatfield & Schluter 1999; Rundle *et al.* 2000; Rundle 2002). Le groupe morphologique benthique consomme du plancton en eau profonde alors que le groupe limnétique consomme des invertébrés sur les eaux moins profondes du littoral et sont caractérisés par un corps plus gros et une structure différente de la mâchoire. Les hybrides F1 obtenus au laboratoire ne sont pas affectés dans leur valeur sélective au laboratoire, mais le sont fortement dans l'environnement (Hatfield & Schluter 1999). Les rétrocroisements avec les populations parentales rétablissent sensiblement leur valeur sélective dans un des deux environnements (eau profonde si rétrocroisé avec le groupe benthique ; littoral si rétrocroisé avec le groupe limnétique).

1.3 Zone hybrides et contact secondaire

Distinguer les différents types de barrières est difficile en situation d'allopatrie. En effet, la structure génétique peut induire une forte hétérogénéité de F_{ST} sur le génome par le simple effet de la dérive génétique avec, dans les cas les plus extrêmes, des fixations d'allèles différents entre les populations. Une solution consiste donc à étudier ces barrières dans des situations de contacts secondaires avec flux de gènes. Par définition, les zones d'hybridation (dont les zones hybrides *sensus stricto*) sont des zones étroites où les populations génétiquement distinctes se rencontrent, se croisent et produisent des hybrides (Barton & Hewitt 1985). Ces zones sont des lieux particulièrement adaptés pour déterminer l'intensité et la nature des barrières génétiques. En effet, à la différence d'une approche réalisée par croisements en laboratoire où le nombre d'évènements de recombinaison est limité, les zones d'hybridation produisent une grande variété de génotypes issus de nombreuses générations de recombinaison. En conséquence, il est théoriquement possible de distinguer des effets de locus physiquement proches, même dans les zones présentant un faible taux de recombinaison (Rieseberg *et al.* 1999). Ces zones sont donc considérées comme le « laboratoire naturel de la divergence » (Hewitt, 1988) et permettent de quantifier le nombre d'hybrides observés par rapport aux nombre d'hybrides attendus sous l'hypothèse d'une absence de mécanismes d'isolement reproductif. Deux grands types de zone hybride existent : les zones de contact primaire correspondant à la zone de chevauchement des aires de répartition de deux populations parapatriques ; ou des zones de contact secondaires provenant de la remise en contact de deux populations allopatrique. La remise en contact secondaire résulte d'un changement du régime de migration de l'allopatrie ($m=0$) vers la parapatrie ou sympatrie ($0 < m \leq 0.5$). D'après la liste des 106 zones hybrides référencées dans Barton & Hewitt (1985), 37% d'entre elles impliquent des populations présentes initialement dans un environnement fragmenté et pour lesquelles la remise en contact secondaire est très probable. La zone hybride en Europe de l'Est entre les deux espèces de crapauds sonneurs *Bombina bombina* & *B. variegata* constitue un exemple bien connu (Szymura & Barton 1986; Szymura 1993;

MacCallum *et al.* 1998; Kruuk *et al.* 1999b). La divergence serait survenue lors des périodes glaciaires où les populations étaient vraisemblablement isolées dans différents refuges (Szymura & Barton 1986; Hofman *et al.* 2007). Sans une bonne connaissance de l'histoire évolutive des populations, il apparaît donc difficile de déterminer si le contact est primaire ou secondaire. Cependant la divergence en allopatrie apparaît plus vraisemblable à mesure que le nombre d'incompatibilités détectées est grand (Coyne & Orr 2004). Il est en effet difficile d'accumuler un grand nombre d'incompatibilités en présence de flux génique ce qui rend peu compatible l'hypothèse parapatrique de contact primaire.

Dans les zones de contact secondaire, le temps de divergence entre les populations est un paramètre-clef de la prédiction de l'intensité de l'isolement reproductif. Si le temps de divergence en allopatrie est trop court pour mettre en place un isolement reproductif, alors la recombinaison va tendre vers une réhomogénéisation génétique complète des populations. Inversement, si le temps de divergence a été suffisamment long pour autoriser la mise en place de nombreuses barrières génétiques, alors on peut attendre une absence totale de flux génique. Cette imperméabilité au flux génique suggère alors que les hybrides produits sont non viables ou stériles. L'intégrité des génomes ne pouvant plus être altérée par la recombinaison, ces deux populations peuvent alors être considérées comme deux espèces biologiques vraies. Cette contre sélection totale des hybrides est en effet nécessaire à la définition des espèces (Barton 2010). Les deux situations (flux génique libre ou absence totale) ne représentent évidemment que deux situations extrêmes. L'isolement reproductif du aux barrières peut être incomplet. Des situations intermédiaires sont alors observées avec une hétérogénéité du flux génique sur l'ensemble du génome (Figure 5). On peut distinguer trois type de flux de gènes : (1) au locus du gène d'isolement reproductif, la migration efficace (Nm) est supposée nulle ($Nm=0$), (2) dans les régions neutres en déséquilibre de liaison avec le locus d'isolement, Nm augmente graduellement avec la distance génétique et (3) aux locus neutres indépendants des gènes d'isolement reproductif la migration n'est pas influencée par des processus sélectifs (encadré I). Cette caractéristique explique l'appellation « barrières semi-perméables » au flux génique (Harrison 1986, 1990) et plus récemment celui « d'îlots génomiques » (Turner *et al.* 2005).

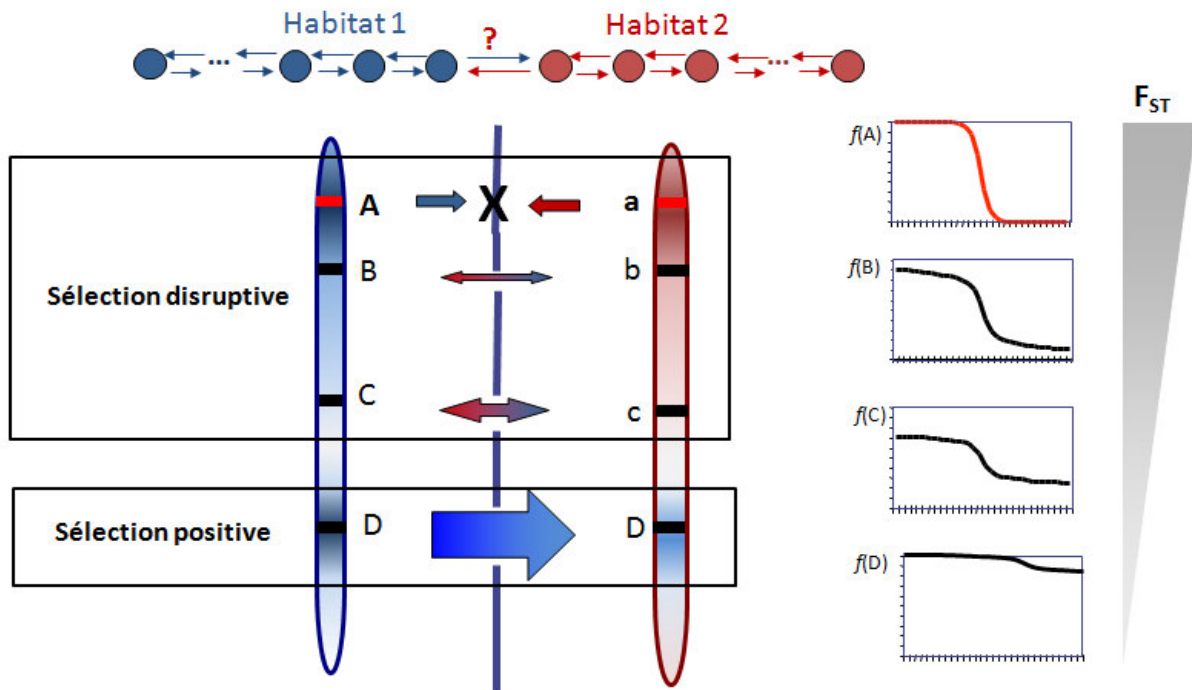


Figure 5 : Barrières génétiques et semi-perméabilité des génomes. Considérons le niveau de flux génique entre deux génomes différemment adaptés à deux habitats composés de 15 demeures de l'habitat 1 et 15 demeures de l'habitat 2 alignés sous un modèle en pas japonais en une dimension (Kimura 1953). Une barrière génétique (locus A) confère une adaptation différentielle à deux habitats (allèle A dans l'habitat 1, a dans l'habitat 2) se traduisant par un cline géographique particulièrement marqué au niveau de l'écotone. Aux locus neutres (B et C), l'introgession est de plus en plus forte à mesure que la distance génétique entre le marqueur et le locus sous sélection augmente (effet d'autostop génétique, encadré I). Des locus avantageux (D) dans l'autre fond génétique peuvent rapidement être introgressés dans l'autre fond génétique. Modifié d'après Bierne (2001) et Faure (2008).

Encadré I : autostop génétique

Lorsqu'une mutation favorable apparaît, son avantage sélectif favorisera une augmentation rapide de la fréquence de cet allèle jusqu'à sa fixation. Cette augmentation rapide entraîne une augmentation de la fréquence des allèles à proximité de ce locus sous sélection (Maynard Smith & Haigh 1974). Cet effet d'entraînement appelé, autostop génétique, se traduit par un balayage sélectif c'est-à-dire une perte locale de la diversité génétique au voisinage du locus sélectionné. La taille de l'autostop dépend de deux forces évolutives : la sélection qui tend à augmenter la fréquence de tous les allèles originaires du même fond génétique que la mutation favorable et la recombinaison qui tendra à éroder cette association. Cette érosion dépend aussi du temps depuis l'événement de balayage sélectif (Kim & Stephan 2002). Lorsque les populations sont de grande taille, l'autostop génétique peut réduire la diversité génétique bien plus facilement que par la seule dérive génétique.

L'autostop local dans un environnement hétérogène correspond à une situation où une mutation favorable apparaît dans un dème (mais est défavorable dans un autre) entraînant localement son environnement génomique (Figure 6). Cet entraînement se traduit par une

forte différenciation génétique entre les populations au locus d'adaptation et son environnement génomique immédiat. La diminution de la différenciation génétique (ex. F_{st}) est graduelle à mesure que la distance avec le gène sous sélection augmente (p. ex. Charlesworth *et al.* 1997). Un autre patron a été décrit, où l'on considère une mutation intrinsèquement favorable dans deux populations structurées. La fixation de cette mutation favorable dans les deux populations donnera un autre patron de différenciation produisant deux pics de différenciation de part et d'autre du locus sous sélection (« autostop global en population structurée » Figure 6 ; plus de détails dans Faure *et al.* 2008 ; Bierne 2010). Les différences dans les patrons d'autostop sont une dimension supplémentaire à intégrer dans le large cadre conceptuel associé à l'interprétation des scans génomiques. En effet, le choix du modèle qui sous-tend ces patrons peut influencer la localisation du locus d'isolement et conduire à une identification biaisée.

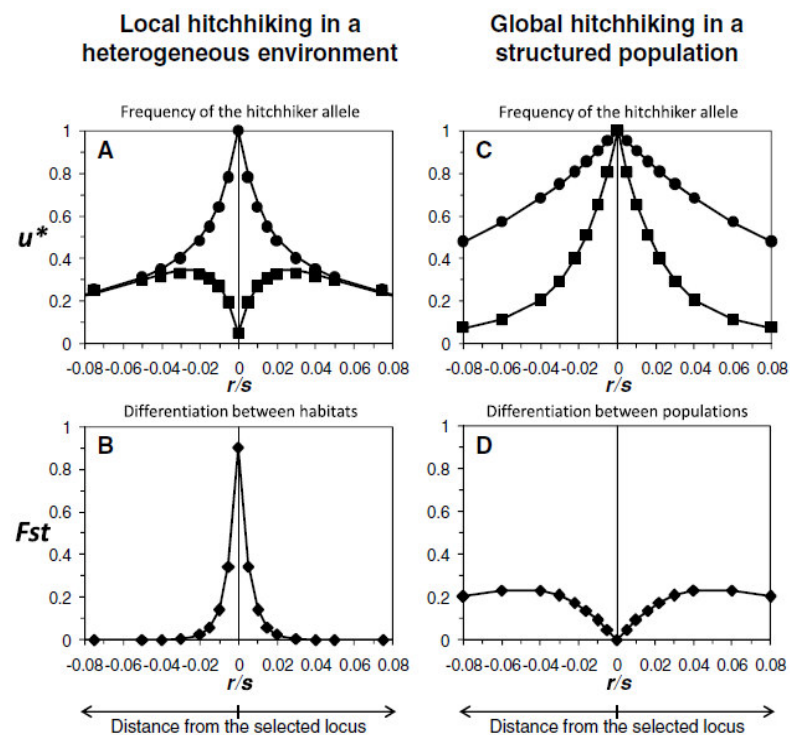


Figure 6 : Effet de l'autostop local dans des environnements hétérogènes et de l'autostop global en populations structurées sur les régions chromosomiques liées au locus sous sélection positive d'après Bierne 2010. La distance physique est ici présentée en fonction du ratio entre la sélection et la recombinaison. Si une mutation avantageuse conférant l'adaptation à un habitat 1 apparaît, elle augmente rapidement en fréquence (u^*) et entraîne son environnement génomique (A). Cela se traduit par un pic de différenciation (F_{st}) à ce locus et une réduction graduelle du F_{st} à mesure que la distance physique augmente (B). A contrario, si deux génomes sont structurés spatialement et qu'une mutation favorable apparaît dans un dème et diffuse dans un autre dème, elle va fixer dans les deux dèmes générant un balayage sélectif au locus (C et D), néanmoins la diversité génétique sera de moins en moins balayée à mesure que le déséquilibre de liaison avec le locus diminue.

1.4 Etude des barrières/Interprétations des scans génomiques

Les remises en contact secondaire de population semi-perméables revêtent un intérêt tout particulier en biologie évolutive car ils permettent la compréhension des mécanismes en cours lors de l'établissement des barrières d'isolement reproductif (Barton & Hewitt, 1985).

La recherche de marqueurs dits « outliers », c'est à dire s'écartant des attendus neutres de différenciation génétique (ex : F_{ST}), à partir d'un échantillonnage d'un grand nombre de marqueurs (scan génomique) peut donc être une méthode efficace pour rechercher les régions génomiques contribuant à l'isolement reproductif entre deux populations. Cette approche a d'abord été initiée par Cavalli-Sforza (1966) puis développée par Lewontin & Krakauer (1973). Ces auteurs ont proposé une méthode statistique permettant de tester l'hétérogénéité de l'indice de fixation de Wright (F_{ST} ⁴) entre locus. Ce test suppose que si quelques marqueurs présentent une différence significative par rapport à la distribution supposée neutre du F_{ST} , c'est qu'ils sont très probablement influencés par la sélection. Ce test a reçu beaucoup de critiques (Nei & Maruyama 1975; Robertson 1975a, b; Tsakas & Krimbas 1976) car il serait biaisé par de forts taux de mutation, des tailles efficaces de populations inégales ou encore par l'existence de patrons de différenciations particuliers (structure génétique, isolement par la distance). Néanmoins, les récentes modifications apportées à ce test par l'utilisation des méthodes basées sur des simulations (Beaumont & Nichols 1996; Vitalis *et al.* 2001), ou de l'inférence Bayésienne (Beaumont & Balding 2004; Foll & Gaggiotti 2008; Gautier *et al.* 2009; Bazin *et al.* 2010; Bonhomme *et al.* 2010) ont largement contribué à l'amélioration de la performance des algorithmes de scan génomique (Pérez-Figueroa *et al.* 2010 ; mais voir aussi § 1.4.1 et Meirmans, 2012).

Généralement les marqueurs révélés ne sont pas directement sous sélection mais liés physiquement ou génétiquement aux locus réellement sous sélection. La grande majorité des auteurs considère que la sélection affecte une partie du génome suffisamment grande à travers l'effet d'autostop génétique (voir encadré I) pour que les marqueurs neutres liés se comportent de la même façon que le marqueur sous sélection (mais voir Faure *et al.* 2008 ; Bierne 2010) . En conséquence, la proportion d'outlier détecté (généralement 2-10%, Nosil *et al.* 2009) est souvent interprétée comme la part du génome affectée par la sélection naturelle. Un calcul simple mais évocateur a été formulé par Bierne *et al.* (2011), considérant cette proportion d'outliers et la longueur de l'autostop génétique généralement observé pour les organismes où le génome est disponible - environ 5 kb – [*Arabidopsis lyrata* (Turner *et al.* 2010) ; *Drosophila melanogaster* (Kolaczowski *et al.* 2010)]. Des milliers de barrières génétiques devraient être présentes à l'intérieur des génomes pour expliquer une telle abondance en outliers. Cette hypothèse étant volontairement extravagante, nous étudierons pourquoi la détection d'outliers ne sous-entend pas qu'ils sont forcément sous sélection naturelle.

1.4.1 Outliers non-affectés par la sélection naturelle (Faux positifs)

Le rejet d'un test de neutralité de Lewontin et Krakauer n'est pas aussi robuste qu'un test direct de sélection positive [comme par exemple un test de McDonald-Kreitman (Nielsen 2005)]. La conséquence majeure est qu'une grande partie des outliers peut être détectée comme non-neutre et pourtant ne pas être affectée par la sélection naturelle, que ce soit intrinsèquement ou au travers de l'autostop génétique. Ces faux positifs peuvent provenir de différentes hypothèses alternatives qui doivent être considérés (revue dans Bierne *et al.* 2011).

⁴ Le F_{ST} correspond à la part variance en fréquence allélique distribuée entre les populations en comparaison à la variance totale disponible (c'est-à-dire distribuée intrapopulation ou interpopulation).

1. L'existence de subdivisions spatiales des populations peut augmenter la variance de différenciation génétique neutre (F_{ST}) ce qui a tendance à dépasser la variance attendue par un modèle neutre basé sur des simples modèles en îles (Robertson 1975a ; Excoffier *et al.* 2009 ; Bonhomme *et al.* 2010 ; voir aussi Meirmans, 2012). Plus précisément, Fourcade *et al.* (sous presse) suggèrent que les paysages fractals tels que les réseaux hydrographiques peuvent conduire à une surestimation très importante du nombre de marqueurs sous-sélection. Ces réseaux complexes semblent expliquer la très forte proportion d'outliers détectés chez les poissons [p. ex. 9 à 40% chez les saumons du Pacifique (Aguilar 2006; Heath *et al.* 2006; Spies *et al.* 2007; Narum *et al.* 2008)].
2. Un autre biais majeur est lié à l'existence d'un effet « de surf » dans les populations en expansion. Quand une mutation neutre arrive au niveau du front de colonisation de cette population, la fréquence de l'allèle muté peut fortement augmenter par le simple effet de la dérive génétique car la colonisation de l'habitat est effectuée par un nombre très restreint d'individus. Cet effet de surf génère des patrons relativement similaires à la sélection locale (Klopfstein *et al.* 2006; Hofer *et al.* 2009).
3. De la même façon, la contre-sélection des mutations délétères peut aussi contribuer à une augmentation de la différenciation génétique en raison d'une diminution de la taille efficace à l'intérieur des sous populations (Charlesworth *et al.* 1997).
4. Enfin, le F_{ST} étant une méthode fréquentiste, elle nécessite un échantillonnage suffisamment abondant des allèles pour être représentatif de la variabilité de la population et ainsi minimiser le biais d'échantillonnage. Or les populations généralement comparées présentent des effectifs généralement assez faibles, surtout lorsque des données hauts-débits de type NGS (Next Generation Sequencing) sont intégrés aux scans génomiques.

Plusieurs solutions existent pour éviter la détection de ces faux positifs, telles que la comparaison de paires de populations ou l'utilisation de modèles hiérarchiques ou la prise en compte de la démographie, même si ces approches s'accompagnent généralement d'un risque accru de faux négatifs (Foll & Gaggiotti 2008). La comparaison de populations recombinantes dans les zones ponctuelles d'hybridation est une des situations les plus favorables pour écarter les faux positifs. En effet, la recombinaison va tendre à réduire la variance neutre aux marqueurs neutres, mais pas dans les zones génomiques des barrières d'isolement reproductif. En conséquence, la semi-perméabilité des génomes en zone hybride contribue à un meilleur contraste entre les zones génomiques affectées par la sélection et les autres régions du génome.

1.4.2 *Outliers affectés par la sélection naturelle*

Les outliers intrinsèquement sous sélection ou influencés par la sélection affectant un locus voisin correspondent aux « vrais positifs » recherchés par les utilisateurs des algorithmes de scans génomiques. Toutefois, là encore, l'interprétation des outliers est particulièrement dépendante du cadre conceptuel sous lequel les résultats sont mis en relief.

1.4.2.1 Interprétation écologique

Les scans génomiques sont fortement utilisés pour comparer les génomes de populations d'habitats différents et ainsi identifier les zones génomiques responsables de l'adaptation locale. Considérons ici que les outliers révélés sont des marqueurs seulement influencés par la sélection naturelle disruptive. Sous cette hypothèse, les auteurs interprètent

classiquement les outliers comme des évidences de marqueurs co-localisant avec les barrières exogènes responsables de l'adaptation locale. Toutefois, bien peu d'études ont fait le lien entre génotype, phénotype, valeur sélective et variations environnementales (mais voir Lenormand *et al.* 1999 ; Hoekstra *et al.* 2004 ; Wheat *et al.* 2006 ; Schmidt *et al.* 2008 ; Storz & Wheat *et al.* 2010), une raison invoquée est que la cartographie des barrières s'avère ensuite complexe (Bierne *et al.* 2011). En effet, plusieurs hypothèses alternatives peuvent être formulées pour expliquer pourquoi un grand nombre d'outliers ne co-localisent pas. Etant donné (i) la faible taille des zones génomiques affectées par l'effet d'autostop, (ii) le grand nombre de marqueurs outliers détectés (2-10%) et (iii) l'échantillonnage relativement faible de marqueurs (généralement quelques centaines répartis sur des génomes de plusieurs méga ou gigabases), la co-localisation des outliers est une hypothèse peu vraisemblable. Deux interprétations sont alors possibles, soit l'adaptation locale est due à un très grand nombre de gènes, soit d'autres processus que la seule adaptation locale peuvent expliquer ces forts niveaux de différenciation.

1.4.2.2 Interprétation non-exclusivement écologique

Si les outliers correspondent bien à des marqueurs en déséquilibre de liaison avec les habitats, l'interprétation écologique de tous les outliers peut s'avérer erronée. En effet, après une remise en contact secondaire des populations, l'expression des incompatibilités de type Dobzhansky-Muller chez les hybrides peut générer les mêmes patrons de différenciation aux locus de barrières qu'attendus par l'adaptation locale (p. ex. Kruuk *et al.* 1999a). En effet, la contre-sélection des hybrides et la sélection écologique vont toutes les deux contribuer au maintien de la différenciation génétique à ces locus. En revanche, les conséquences de cette sélection sur l'environnement génomique immédiat du locus sont inconnues, aucune étude théorique n'a formellement caractérisée les patrons d'autostop génétique dans le cas de remise en contact secondaire de population⁵.

Les deux interprétations écologiques et non-écologiques ne sont pas mutuellement exclusives, un même scan génomique peut révéler à la fois des barrières endogènes et exogènes. La divergence allopatrique entre deux populations peut aisément conduire à la mise en place d'adaptations locales différentes à certains locus et d'incompatibilités à d'autres. Chez de nombreuses paires d'espèces, les deux types de barrières exogènes et endogènes sont actives de façon concomitante (Barton & Hewitt, 1985 ; Hewitt 1988 ; Table 1 dans Bierne *et al.* 2011). L'interprétation la plus parcimonieuse envisage la contribution à la fois des barrières écologiques et non-écologiques. Notons toutefois que (i) la seule détection des barrières exogènes devient de plus en plus délicate à mesure que la période de divergence est longue et (ii) qu'une connaissance relativement exhaustive⁶ de l'histoire évolutive de l'organisme d'étude est un préalable à toute interprétation des outliers. L'étude des barrières génétiques dans la zone hybride Européenne de la souris (*Mus musculus musculus*/*Mus musculus domesticus*) n'a pour l'heure révélé que des barrières endogènes et aucune barrière exogène.

⁵ Par la suite, on considérera que les patrons d'autostop génétique générées par des barrières de type Dobzhansky-Muller sont identiques à ceux attendu sous un autostop local (encadré I). Cette hypothèse bien que probable (la sélection disruptive tend en à maintenir des allèles différents au locus induisant un pic de différenciation à ce locus et la recombinaison tend à éroder graduellement cette différenciation sur les régions génomiques flanquantes de la barrière génétique) reste à valider formellement.

⁶ L'existence ou non de périodes d'allopatrie et de remise en contact secondaire est une information cruciale à la formulation des attendus. Les méthodes d'inférences bayésiennes telles que les méthodes ABC permettent aujourd'hui de formuler un grand nombre de scénarios et tester si des périodes de divergence allopatrique sont probables (voir encadré IV pour un exemple d'application).

2. EVOLUTION DES POPULATIONS DE PATHOGENES

Le fort clivage entre partisans et opposants aux hypothèses de divergence sympatrique a longtemps marqué l'histoire de notre discipline et a été à la source de nombreux projets de recherche, à la fois théorique ou empirique (Felsenstein 1981; Johnson & Gullberg 1998; Kirkpatrick & Ravigné 2002; Gavrillets 2003). Selon Coyne & Orr (2004), quatre arguments doivent être simultanément validés pour écarter l'hypothèse nulle d'une divergence allopatrique : les populations doivent (1) être sympatriques ou au moins l'avoir été au moment du processus initial de divergence, (2) présenter un isolement reproductif « substantiel », (3) être observé entre des groupes phylogénétiquement proches ne résultant pas d'événements d'hybridation et (4) l'histoire évolutive de ces groupes doit pouvoir réfuter l'hypothèse d'une divergence en allopatrie. Selon ces auteurs, seuls 3 exemples empiriques soutiennent complètement ce modèle : les poissons Tilapiinae de la famille des Cichlidés vivant dans les cratères de lacs du Cameroun (Schliewen *et al.* 1994) ; l'ombre chevalier dans le lac Galtaból en Islande (Gislason *et al.* 1999) et les guêpes du genre *Apocryptophagus* parasitant les figuiers (Weiblen & Bush 2002). Face à la difficulté de réfuter l'hypothèse d'allopatrie, de nombreux biologistes de l'évolution ont renoncé à rechercher des exemples de spéciation sympatrique et préféré quantifier la contribution de l'écologie dans la divergence (« spéciation écologique »)⁷. Les hypothèses de spéciation écologique expliquant l'évolution de l'isolement reproductif entre populations en réponse à une sélection naturelle divergente par les habitats (Funk 1998; Schluter 2000; Schluter 2001; Via 2001; Rundle & Nosil 2005; Schluter 2009; Schluter & Conte 2009) ont connu un spectaculaire essor au cours de la dernière décennie. Désormais, l'hypothèse de spéciation écologique prédomine par rapport aux hypothèses plus neutres du modèle de Dobzhansky-Muller, au moins en nombre total de citations (depuis 2007, Fig 1 dans Hendry 2009). Les pathogènes forment un ensemble d'organismes pour lesquels la diversification est essentiellement interprétée du point de vue de l'adaptation à l'hôte. Ce lien étroit entre pathogène et hôte a bien souvent autorisé des interprétations dans le cadre conceptuel de la spécialisation écologique (de Meeus *et al.* 1998; Schluter 2000; Rundle & Nosil 2005; Giraud *et al.* 2006; Giraud *et al.* 2010; Gladieux *et al.* 2011)

2.1 Modèles théoriques : forte adaptation à l'habitat attendue

Pour un pathogène, l'hôte n'est pas seulement une ressource alimentaire mais représente aussi un habitat et un moyen d'être véhiculé (Combes 1995). Si le pathogène est bel et bien soumis aux variations continues d'un environnement biotique et abiotique bien plus complexe, les modèles théoriques considèrent généralement l'hôte comme le facteur déterminant de la diversification des pathogènes (p. ex. McCoy 2003).

La théorie de la Reine rouge formulée par Van Valen (1973) est l'une des théories les plus populaires pour décrire les relations biotiques entre les organismes dans les écosystèmes. Cette théorie suggère que toute modification d'une espèce modifie l'environnement de toutes les autres espèces, en changeant le paysage adaptatif (cf § 3.3.1). Transposé aux pathogènes, ce modèle suggère que toute modification de l'hôte va avoir un impact sur les populations de pathogènes et réciproquement. Cette co-évolution a ainsi été nommée « Théorie de la Reine

⁷ Si intuitivement ce scénario évolutif sous-entend spontanément la divergence en sympatrique, notons ici que les exemples les plus convaincants sont apparus en allopatrie (Vines & Schluter 2006; Langerhans *et al.* 2007; Elias *et al.* 2012).

rouge⁸ » car les espèces doivent perpétuellement évoluer pour rester dans la course. Cette théorie semble particulièrement pertinente pour l'étude des écosystèmes naturels où l'adaptation fut relativement ralentie par la faible taille des populations de pathogènes. En effet, la forte variabilité locale en espèces végétales, l'hétérogénéité des fonds génétiques de leurs hôtes et la faible densité en hôte disponible sont autant de facteurs limitant l'établissement de populations de grande taille et donc de l'adaptation des populations de pathogènes dans les environnements naturels (Stukenbrock & McDonald 2008). Les pressions de sélection exercées par les pathogènes sur les populations de leur hôte autorisent une lente mise en place de nouveaux systèmes de résistance. Alternativement, la mise en place des agrosystèmes est intrinsèquement favorable au pathogène. Leurs hôtes sont génétiquement homogènes et cultivés en forte densité sur des grandes surfaces permettant l'établissement de populations de grande taille (Stukenbrock & McDonald 2008). Dans ces conditions, l'adaptation des populations de pathogènes est attendue de manière beaucoup plus rapide que celles de leurs hôtes pour différentes raisons.

Premièrement, la forte fécondité des pathogènes pouvant produire plusieurs milliers de descendants par jour (Alexopoulos *et al.* 1996; Andrivon *et al.* 2007) autorise la création de nombreux allèles mutés et leurs tests par la sélection naturelle. Les pathogènes sont généralement décrits comme particulièrement plastiques (Dobrindt & Hacker 2001; Raffaele & Kamoun 2012). Ces « mutations » peuvent être générées par des événements de mutations classiques mais aussi par des mécanismes plus ou moins spécifiques tels que l'insertion d'éléments transposables ou par l'existence d'un mécanisme de défense contre ces éléments transposables générant de nombreuses mutations ponctuelles (RIP pour 'Repeat-Induced Point mutations', Selker *et al.* 1987 ; Cambareri *et al.* 1989 ; Grayburn & Selker, 1989 ; Cambareri *et al.* 1991). La forte fécondité des populations de pathogènes permet alors de compenser le fort fardeau génétique associé à l'apparition de ces nombreuses mutations. De plus, en l'absence d'un choix actif de partenaire, les descendants sont généralement dispersés au hasard du vent ou de la pluie. Dans de rares cas, le migrant se retrouve dispersé sur un hôte où il peut établir une interaction compatible. A l'échelle du pathogène, son hôte peut donc être perçu comme une « oasis dans le désert » (Brown *et al.* 2002).

Deuxièmement, les pathogènes sont soumis à de fortes pressions de sélection par la plante. Par exemple, les résistances majeures de type Gène-pour-Gène génèrent de fortes pressions de sélection sur les populations possédant l'allèle avirulent correspondant reconnu par la plante (encadré II). L'intensité de cette sélection favorise l'augmentation en fréquence des allèles favorables à l'adaptation locale et limite l'immigration des allèles maladaptés par l'inviabilité des migrants provenant des autres habitats (Nosil *et al.* 2005; Giraud 2006b).

Troisièmement, chez les pathogènes, la reproduction est majoritairement assurée de manière clonale. Par exemple, chez les champignons, la plupart de ces organismes ne possède qu'un seul cycle de reproduction sexué par an (Giraud *et al.* 2010). La grande partie de leur fécondité est alors assurée par de multiples cycles de reproduction asexuée. Chez les organismes présentant un mode de reproduction de ce type (dit mixte), la reproduction sexuée est favorable à l'établissement de nouvelles combinaisons génétiques qui pourront être testées par la sélection. La reproduction sexuée ayant lieu après la sélection de ces combinaisons génétiques par l'hôte, la fréquence des meilleurs génotypes peut ensuite augmenter fortement au cours des différents cycles de reproduction asexué.

⁸ Le nom de Reine rouge est emprunté au roman « de l'autre côté du miroir » écrit par Lewis Carroll au XIX^{ème} siècle. Dans une scène, Alice et la Reine rouge se lancent dans une course effrénée. Alice demande alors : « Mais, Reine Rouge, c'est étrange, nous courons vite et le paysage autour de nous ne change pas ? » Et la reine répondit : « Nous courons pour rester à la même place. »

Encadré II : Le modèle Gène-pour-Gène

Les plantes présentent deux grands types de résistance contre les attaques par les pathogènes : des résistances quantitatives, des combinaisons génétiques multiples de gènes « mineurs » et les résistances qualitatives dont l'effet est due à un seul gène « majeur ». Cette dernière classe a été formalisée par Flor (1971) sous le nom de modèle Gène-pour-Gène. Sous ce modèle, chaque gène de résistance (R) chez l'hôte interagit spécifiquement avec un gène d'avirulence (Avr) chez le pathogène. L'interaction est incompatible, c'est-à-dire ne conduit pas au développement de la maladie, lorsque les deux protagonistes sont porteurs d'au moins un allèle dominant et fonctionnel du gène R et du gène *Avr*⁹. En effet, dans la plupart des cas l'allèle de résistance est dominant sur l'allèle de sensibilité. Dans tous les autres cas, l'interaction est compatible et se traduit par l'expression de la maladie (Dangl & Jones 2001 ; Jones & Dangl, 2006 ; tableau 1).

Tableau 1 : Relation Gène-Pour-Gène dans le cas d'un agent pathogène haploïde et d'une plante diploïde (Flor, 1971). La plante résistante ne reconnaît le pathogène que s'il est porteur d'un allèle d'avirulence (*Avr*). Cette reconnaissance génère une réaction hypersensible et l'induction d'une large gamme de gènes impliqués dans les réactions de défense se traduisant par une relation de type incompatible. En absence de la production de la protéine R ou si la conformation de la protéine d'avirulence est altérée (*avr* ou *vir*), la plante ne perçoit plus le pathogène et ne peut alors pas déclencher sa réponse immunitaire.

Pathogène	Plante	
	Résistante (RR ou Rr)	Sensible (rr)
Avirulent (<i>Avr</i>)	Réaction Hypersensible	Maladie
Virulent (<i>avr</i> ou <i>vir</i>)	Maladie	Maladie

La reconnaissance de l'allèle d'avirulence par l'allèle de résistance a fait l'objet d'une littérature abondante (revues de Dangl & Jones 2001, Jones & Dangl 2006). Cette reconnaissance entre la protéine de résistance et la protéine d'avirulence est rarement directe mais fait généralement appel au minimum à une troisième protéine appelée « protéine de garde » qui interagit avec les deux précédentes (Van der Biezen & Jones 1998; Dangl & Jones 2001; Van der Hoorn *et al.* 2002).

Les allèles de virulence des pathogènes et les allèles de résistance sont généralement considérés comme générant une perte de valeur sélective (coût de fitness) pour les individus porteurs. En absence d'un avantage sélectif à la résistance ou à la virulence, ces allèles sont attendus en basse fréquence (Figure 7). Dans un modèle de co-évolution (Figure 7), ces coûts sont en effet nécessaires pour maintenir le polymorphisme dans les gènes de résistance et de virulence, comme typiquement observé dans la nature (Burdon & Thrall 1999; Thrall & Burdon 2003). En absence de coût de fitness, les allèles de virulence et de résistance augmentent en fréquence jusqu'à leur fixation.

Dans le cadre du contournement des résistances présentes dans les agrosystèmes, les mêmes attendus sont généralement considérés. Seulement dans ce contexte, l'augmentation des fréquences en hôtes résistants est essentiellement due à l'introgession de résistance

⁹ Chez *V. inaequalis*, la nomenclature généralement utilisée oppose l'avirulence (*avr*, p. ex. *avrRvi6*) et la virulence (*vir*, p. ex. *virRvi6* ; voir par exemple Guérin & Le Cam 2004). Pour ne pas créer de confusion dans la suite du manuscrit, je vais maintenir cette nomenclature *avr/vir*.

génétique *via* les programmes d'amélioration et au déploiement de ces nouvelles variétés dans les agrosystèmes. Le contournement est essentiellement expliqué de la même façon que sous le modèle de coévolution, par un changement d'état du locus d'avirulence. *A contrario*, la baisse de fréquence des allèles de résistances contournées n'est pas liée au coût intrinsèque de la résistance, mais plus vraisemblablement au désintérêt pour les cultivars contournés.

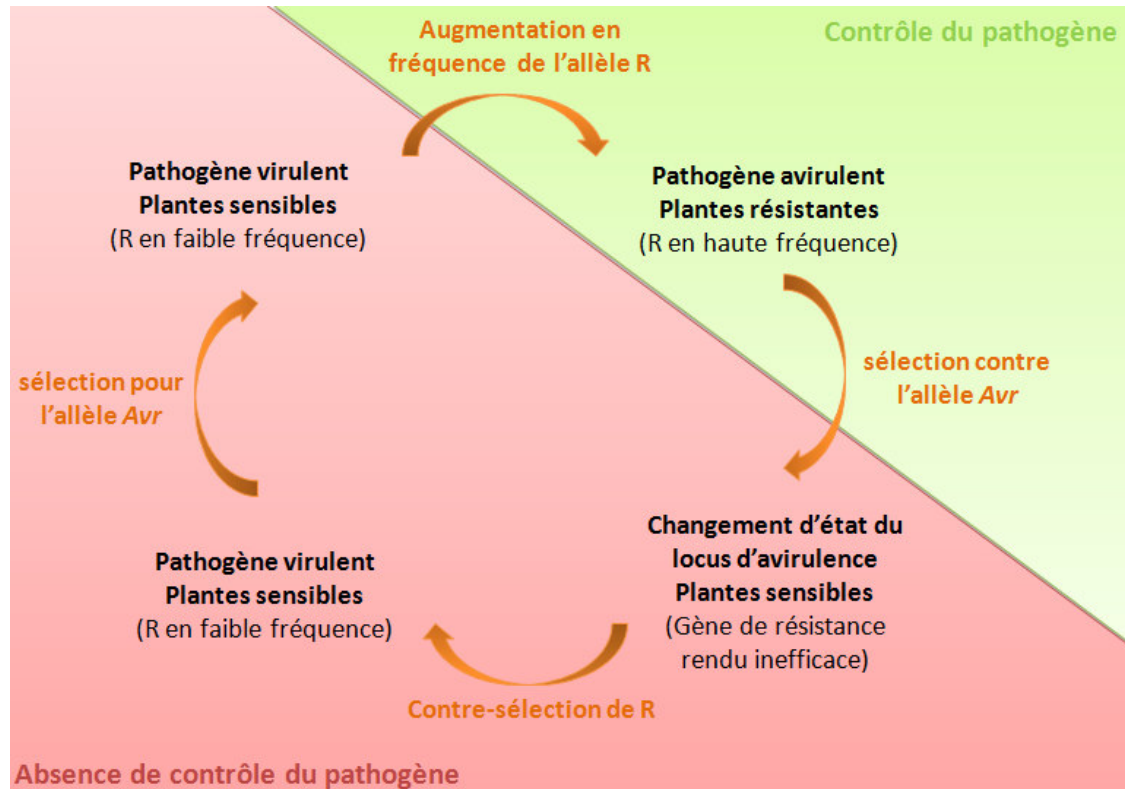


Figure 7 : Coévolution des gènes d'avirulence du pathogène et de résistance des plantes. Modifié d'après Jones & Dangl (2006). Un pathogène porteur d'un gène d'avirulence Avr est reconnu par la plante (en haut). L'avantage sélectif conféré par l'échappement à son pathogène va conduire à une augmentation en fréquence de l'allèle de résistance. La sélection étant très forte à ce locus, toute mutation du pathogène permettant d'échapper à la reconnaissance sera avantagée (à droite). L'efficacité du gène de résistance s'érode rapidement, le rendant plus coûteux qu'avantageux. La fréquence de l'allèle de résistance va alors tendre à diminuer (en bas). La perte de la résistance va alors permettre une réaugmentation en fréquence de l'allèle d'avirulence (considéré comme avantageux pour les pathogène).

Quatrièmement, chez de nombreux champignons pathogènes de la division des Ascomycètes et Basidiomycètes, les gamètes ne peuvent être dispersés au delà de leur hôte (Giraud *et al.* 2010). En conséquence, pour les champignons hétérothalliques où la reproduction a lieu entre des organismes de signes sexuels opposés, l'adaptation à l'hôte a un effet pléiotrope sur le choix de partenaire sexuel. Sous ce modèle, seuls les individus porteurs de la même adaptation locale peuvent contribuer à la génération ultérieure comme attendu avec un choix actif de partenaire. En conséquence, les barrières pré-zygotiques exogènes pourraient être particulièrement efficaces pour maintenir l'intégrité des génomes et autoriser la divergence même en sympatrie (Giraud 2006a, b; Giraud *et al.* 2008; Giraud *et al.* 2010).

2.2 Mécanismes évolutifs conduisant aux émergences

Les hypothèses écologiques de divergence des pathogènes ont majoritairement été considérée sous le prisme de l'extension de sa gamme d'hôte. Dans une revue assez exhaustive, Stukenbrock et McDonald (2008) décrivent 4 grands mécanismes évolutifs pouvant conduire à l'émergence des pathogènes dans les agrosystèmes.

Le suivi de l'hôte au cours de la domestication (« host-tracking¹⁰ ») est un de ces scénarios évolutifs les mieux décrits. Selon ce scénario de co-divergence, la sélection des génotypes d'hôtes présentant les caractères agronomiques les plus intéressants (p. ex. meilleure productivité, résistance aux maladies) aurait simultanément sélectionné les pathogènes les mieux adaptés à ces génotypes d'hôte. Notons que si intuitivement notre vision de la domestication concerne des espèces végétales et animales domestiquées depuis plusieurs milliers d'années (Diamond 1997, 2002), certaines espèces végétales ont été domestiquées beaucoup plus récemment [p. ex. le kiwi ou le triticale, Stukenbrock & McDonald (2008)]. Bien qu'ayant reçu un écho certain au sein de la communauté des phytopathologistes (p. ex. dans Stukenbrock & McDonald 2008), ce scénario évolutif n'a été validé que sur bien peu d'espèces de pathogènes. Parmi les rares exemples, l'émergence de l'agent de la pyriculariose du riz, *Magnaporthe oryzae*, puisque son origine géographique et son temps de divergence avec les autres populations infectant des espèces sauvages (5000-7000 ans) est compatible avec les résultats archéologiques sur la domestication du riz (Couch *et al.* 2005). Ce scénario évolutif a été proposé chez quelques autres espèces : *Mycosphaerella graminicola* sur blé (Stukenbrock *et al.* 2007; Torriani *et al.* 2011), *Ustilago maydis* sur maïs (Munkacsi *et al.* 2008 mais voir aussi Munkacsi *et al.* 2007), *Venturia inaequalis* sur pommier (Gladieux *et al.* 2010; Lê Van *et al.* 2012), *Phytophthora infestans* sur pomme de terre (Gomez-Alpizar *et al.* 2007).

L'adaptation à un nouvel hôte peut conduire à une extension de gamme d'hôte ou un saut d'hôte, c'est-à-dire respectivement la colonisation d'un nouvel hôte en conservant la capacité à infecter l'hôte ancestral ou en perdant cette capacité. Sous ce modèle, un nouveau pathogène peut émerger à partir de populations sauvages ou d'autres plantes cultivées. Ce scénario évolutif est étroitement lié au déplacement de populations localement adaptées dans des environnements où de nouvelles espèces d'hôtes sont présentes. Un des exemples les plus célèbres est l'infection des chataigniers nord-américain par le champignon *Cryphonectria parasitica* à la suite de l'import de chataigniers Asiatiques infectés (Anagnostakis 1987). Plus récemment, Zaffarano *et al.* (2006; 2008) ont démontré que le centre d'origine de l'agent pathogène de l'orge *Rhynchosporium secalis* se situerait en Scandinavie ce qui ne coïncide

¹⁰ Parfois référencé sous le nom de domestication du pathogène (Stukenbrock & McDonald 2008). Pour un pathogène, cette dénomination peut apparaître contradictoire car la domestication se rapporte théoriquement à des effets bénéfiques pour l'homme.

pas avec la zone de domestication de son hôte dans le Croissant fertile.

Le transfert horizontal de gènes a aussi été largement décrit comme un scénario évolutif contribuant à l'émergence de nouveaux agents phytopathogènes. Sous ce scénario, l'acquisition de gènes nécessaire à la pathogénie est effectuée par l'échange de gènes spécifiques ou de régions génomiques (« îlots de pathogénéité ») entre deux espèces reproductivement isolées. Un organisme non-pathogène ou faiblement pathogène peut ainsi devenir virulent ou plus agressif. Notons toutefois que cette définition est imparfaite car les bactéries échappent à la définition du concept biologique d'espèce. Or les exemples les plus nombreux et les plus convaincants sont observés chez les pathogènes procaryotes (de la Cruz & Davies 2000). Chez les eucaryotes, le transfert horizontal de gènes de la voie de biosynthèse des caroténoïdes entre un champignon et le puceron vert du pois est un des cas les plus frappants (Moran & Jarvik 2010). Ce transfert horizontal de gène permet la synthèse de torulène par le puceron, un pigment rouge absent de toutes les autres espèces du règne animal. Les individus rouges semblent mieux échapper à la prédation et contribuerait ainsi à une meilleure valeur sélective du parasite. Plus classiquement en phytopathologie, les autres cas décrits suggèrent des transferts horizontaux de gènes codant des toxines (Walton 2000; Friesen *et al.* 2006) ou des échanges de chromosomes surnuméraires chez les champignons (Han *et al.* 2001; Hatta *et al.* 2002; Hatta *et al.* 2006; Ma *et al.* 2010).

Bien que pourtant très décrite chez leurs hôtes (Rieseberg 2001), la contribution de l'hybridation interspécifique dans l'évolution des populations de pathogènes a été relativement ignorée en phytopathologie. Pourtant les exemples d'hybridation interspécifiques semblent être nombreux et ont vraisemblablement contribué à l'évolution de la pathogénie des populations de pathogènes (Schardl & Craven 2003). A la différence du transfert horizontal, l'hybridation interspécifique implique des génomes entiers et peut ainsi aboutir au changement du nombre de chromosomes, du niveau de ploïdie ou des réarrangements chromosomiques importants (Schardl & Craven 2003). Les hybrides produits peuvent acquérir la capacité à infecter un nouvel hôte ou alors combiner les gammes d'hôtes de leurs deux parents (Brasier 1995). Chez les champignons et les oomycètes, quelques exemples d'hybridation interspécifiques sont bien décrits. Ainsi *Phytophthora alni*, l'agent responsable de dépérissement de l'aulne, aurait émergé par allopolyploïdisation de deux génomes diploïdes de *Phytophthora* ayant des gammes d'hôtes différentes ne comprenant pas l'aulne (Brasier *et al.* 1999). *Phytophthora andina*, un pathogène des Solanacées cultivées dans les Andes, a aussi émergé par une hybridation entre *P. infestans*, l'agent du mildiou de la pomme de terre et une autre espèce inconnue phylogénétiquement proche de *P. murabilis* et de *P. ipomoeae* (Goss *et al.* 2011). Chez les ascomycètes *Ophiostoma ulmi* et *O. novo-ulmi*, agents responsables de la graphiose de l'orne, des hybrides interspécifiques entre ces deux espèces ont aussi été caractérisés (Brasier *et al.* 1998). Ces événements ont aussi été décrits chez le basiomycète *Melampsora x colombiana*, un hybride de génomes de *M. occidentalis* et *M. medusae* partageant la même gamme que leurs parents, le peuplier (Newcombe *et al.* 2000). Chez les champignons, la modification du niveau de ploïdie a aussi été décrites chez *Verticillium longisporum* (Inderbitzin *et al.* 2011) ou *Botrytis allii* (Staats *et al.* 2005) infectant respectivement les crucifères et les oignons. L'importance de l'allopolyploïdisation chez les champignons est probablement liée au phénomène de la parasexualité. La capacité des champignons à tolérer la présence de plusieurs noyaux différents dans leurs cellules est un facteur favorable à l'allopolyploïdisation par caryogamie de génomes divergents (Schardl & Craven 2003).

2.3 La remise en contact secondaire des populations de pathogènes : une hypothèse négligée

Si l'hybridation de populations présentant la même gamme d'hôte peut apparaître compatible avec un modèle écologique, l'hybridation entre des espèces présentant une gamme d'hôte différente est plus difficile à appréhender. Les pathogènes sont probablement bien plus en contact entre eux dans l'environnement que communément admis, *via* des co-infections d'hôtes actuellement non-décrits (p. ex. en phase saprophytique, voir Morris *et al.* 2009). De plus, les exemples relativement nombreux d'hybridations interspécifiques observés chez les champignons phytopathogènes suggèrent qu'à l'échelle intraspécifique la probabilité que les populations divergentes de pathogènes s'hybrident est encore plus forte qu'entre des espèces différentes puisque ces populations infectent généralement la même gamme d'hôtes. Étonnamment, les hypothèses de remise en contact secondaire de populations divergentes sont largement ignorées chez les pathogènes au profit des hypothèses écologiques précédemment citées. Pourtant ces zones d'hybridation sont généralement considérées comme « le laboratoire naturel de la spéciation » (Hewitt 1988) et permettent d'étudier la contribution de l'ensemble des barrières (exogènes et endogènes) à l'isolement reproductif. Or puisque les barrières endogènes sont considérées comme le modèle nul de divergence pour l'ensemble des espèces, la réfutation de ces hypothèses est un préalable à toute interprétation écologique (Coyne & Orr 2004). En phytopathologie, plusieurs expressions traduisent cette vue générale sur le rôle de l'adaptation locale, la confusion de l'usage du mot « adaptation » à la place de spécialisation ou de diversification ; ou encore l'utilisation de l'expression 'incipient species' pour désigner le processus qui conduira à la naissance de nouvelles espèces, sans tenir compte que cette divergence peut être rapidement balayée par un contact secondaire faisant suite à un changement des aires de répartition des populations. Historiquement, la phytopathologie est une discipline qui est étroitement liée à la botanique et à l'amélioration des plantes. Ce regard particulier a largement favorisé une étude restrictive des organismes pathogènes en se focalisant uniquement sur son interaction avec son hôte et donc sur les pressions de sélection générées par l'environnement.

Les pathogènes possèdent pourtant des caractéristiques biologiques compatibles avec l'hypothèse de divergence suivie d'une remise en contact secondaire. D'abord, de nombreuses espèces de pathogènes ont des spores intrinsèquement peu favorables à un long temps de vol et sont ainsi dispersés sur des distances relativement courtes (McDonald & Linde 2002). Même si de nombreuses espèces de phytopathogènes se dispersent à une plus longue distance (Brown & Hovmøller 2002), il n'en demeure pas moins que leur capacité de migration dépasse rarement le millier de kilomètres (McDonald & Linde 2002). En considérant simultanément la redistribution mondiale des plantes domestiquées dans des zones différentes du monde et la présence de barrières géographiques à la dispersion entre continents (mers, montagnes), la divergence des populations en absence ou quasi-absence de flux génique peut autoriser la mise en place de barrières post-zygotiques endogènes et la remise en contact de ces populations est ensuite largement dépendante de la redistribution de ces populations, essentiellement d'origine anthropique (transport de plants contaminés, exportations des productions, connexions des zones cultivées).

Récemment, les hypothèses de contact secondaire ont été testées plus rigoureusement par Giraud & Gourbière (2012) chez les champignons. À partir d'une revue relativement exhaustive des cas décrits d'isolement reproductif entre les populations de champignons (Le Gac & Giraud 2008), Giraud & Gourbière (2012) suggèrent que d'accumulation des barrières génétiques au cours du temps chez les Ascomycètes et les Basidiomycètes ne suit pas le modèle classique d'accumulation des barrières par l'effet boule de neige que ce soit chez les

taxons allopatriques ou sympatriques. *A contrario*, les hypothèses les plus probables sont une décroissance linéaire (incompatibilités simples sans épistasie) chez les Ascomycètes et d'une augmentation très rapide de l'isolement reproductif en fonction de la divergence chez les Basidiomycètes calculée sur la base de séquences au locus ITS ('Internal Transcribed Spacer') [modèle « snowdown » suggérant la mise en place d'une barrière écologique forte réduisant fortement le flux génique, voir Gourbière & Mallet (2010)]. Les auteurs suggèrent donc que le modèle classique basée sur l'effet d'emballement ne correspond pas aux données observées d'isolement reproductif partiel des champignons. L'isolement reproductif des ascomycètes étudiés semblent mieux correspondre à un modèle d'isolement post-zygotique linéaire dû à des remaniements chromosomiques. *A contrario*, les observations sur basidiomycètes corrèlent mieux avec un isolement reproductif essentiellement expliqué par des barrières génétiques pré-zygotiques. Bien que les effectifs soient relativement importants (69 paires d'Ascomycètes « sympatriques » et 20 allopatriques, 47 paires de Basidiomycètes sympatriques et 57 allopatriques), ces effectifs restent trop faibles pour créer une forte différence entre les 3 modèles testés, particulièrement dans le cas des Ascomycètes (Giraud & Gourbière 2012). De plus, cette étude ne distingue pas formellement les champignons pathogènes des autres champignons, ne permettant pas de formaliser des attendus propres aux pathogènes. Les attendus sont vraisemblablement différents entre les espèces d'ascomycètes, par exemple, chez l'ascomycète *Saccharomyces cerevisiae*, les barrières semblent bien s'accumuler avec le temps de divergence au minimum de manière linéaire (Gourbière & Mallet 2010) voire quadratique (Matute *et al.* 2011). Enfin, la distance génétique est exclusivement basée sur la séquence du seul locus ITS. L'acquisition de données de séquençage haut débit permettra de mieux estimer le taux de substitution basée sur l'ensemble du génome et ainsi de mieux estimer le temps de divergence entre les différents taxons.

Si l'étude récente de Giraud & Gourbière (2012) permet de formaliser des attendus particuliers aux ascomycètes et basidiomycètes, les informations sont encore très parcellaires. Une des raisons est que cette façon d'étudier globalement la mise en place des barrières génétiques nécessite d'utiliser la distance génétique comme prédicteur du temps de divergence (Gourbière & Mallet 2010) et est forcément basée sur une simplification des attendus en séparant des hypothèses de barrières post-zygotiques endogènes d'un côté et des hypothèses pré-zygotiques exogènes de l'autre. La théorie prévoit pourtant que ces deux dernières hypothèses sont pourtant loin d'être exclusives (Coyne & Orr 2004; Barton & de Cara 2009; Bierne *et al.* 2011), *a fortiori* lorsque les pressions de sélection de l'environnement sont particulièrement fortes comme attendus chez les organismes pathogènes (Bierne *et al.* 2011).

3. LES HYPOTHESES NON-ECOLOGIQUES : UN NOUVEAU CADRE CONCEPTUEL A TRANSPOSER CHEZ LES PATHOGENES

La description de la nature des barrières génétiques est particulièrement complexe car les barrières exogènes et endogènes sont dépendantes des mêmes forces évolutives et génèrent des patrons relativement similaires (voir ci-dessous). Deux grandes méthodes permettent toutefois de renseigner leur nature : l'utilisation de populations naturelles isolées de zone hybrides d'une part et l'utilisation d'approches de cartographie génétique à partir de croisements *in vitro* d'autre part. Dans les deux cas, l'interprétation des patrons peut être particulièrement délicate et nécessite de se pencher sur les attendus théoriques associés à

chacune de ces barrières génétiques.

Les attendus théoriques notamment proposés à partir de la théorie des clines n'ont pas, ou que bien partiellement, été transposé chez les pathogènes. Pourtant, ces attendus théoriques semblent suggérer que les barrières endogènes sont particulièrement attendues chez les pathogènes. Si l'étude des clines apparaît complexe chez les pathogènes étant donné la structure mosaïque des habitats des pathogènes, des méthodes récentes semblent prometteuse pour discriminer les locus liés à l'isolement reproductif des pathogènes mais aussi les locus qui sont fortement introgressés dans un génome étranger (voir ci-dessous).

3.1 Théorie des clines

3.1.1 *Clines géographiques*

La théorie des zones hybrides est très étroitement liée à la description des clines géographiques de fréquence allélique. A l'échelle d'un locus, le cline repose sur un équilibre entre deux forces antagonistes, la migration qui tend à homogénéiser les fréquences alléliques et la sélection qui tend à ré-enrichir la population en géotypes parentaux. La largeur de ce cline géographique est dépendante du ratio entre ces deux forces antagonistes. La théorie des clines fut initialement développée pour formuler les attendus d'isolement reproductif post-zygotique endogène (Key 1968; Bazykin 1969; Barton 1983; Barton & Shpak 2000). A contrario, les attendus basés sur des barrières écologiques d'isolement ont été formulés plus tardivement (Gaggiotti *et al.* 1997; Gavrilets 1997c; Kruuk *et al.* 1999a). Un des grands résultats de ces simulations est que la forme des clines aux locus de barrières exogènes est indissociable des clines de barrières post-zygotiques endogènes car les variations de fréquence allélique dans l'espace sont dépendants des deux mêmes forces évolutives (Kruuk *et al.* 1999a). En effet, la seule différence entre barrières endogènes et exogènes est le mode de sélection, dans le premier cas, la sélection contre les hybrides est intrinsèque à l'organisme, dans le second, la sélection est générée par l'environnement. La conséquence principale est que l'étude des fréquences alléliques à certains locus marqueurs de barrières (p. ex. les outliers obtenus par scans génomiques), n'est que faiblement informative sur la nature des barrières génétiques en présence.

La théorie des clines monolocus a largement contribué à l'établissement de la théorie plus intégrative des clines multilocus. Sans rentrer dans les détails, lorsque plusieurs locus sont impliqués dans le maintien de la différenciation des génomes, des interactions entre les locus sont mises en place et génèrent du déséquilibre de liaison (Slatkin 1975; Barton 1983). Le couplage (Φ) entre deux barrières est donc dépendante de deux forces évolutives le taux de sélection et le taux moyen de recombinaison entre deux locus (Barton 1983 ; Kruuk 1999). Si Φ est faible (faible sélection et faible liaison génétique), les locus se comportent de manière indépendante du reste du génome comme attendu pour la théorie des clines monolocus. A contrario, si Φ est fort (sélection forte et/ou liaison forte) et le nombre de barrières relativement nombreux, alors le génome se comporte comme un bloc non-recombinant, on parle alors de génome « congelés » (Turner 1967; Kruuk *et al.* 1999a). Des résultats plus récents (Barton & de Cara 2009) suggèrent un couplage d'un grand nombre de barrières génétiques permettant ainsi d'augmenter la variance génétique et de renforcer l'isolement reproductif. En outre, ce couplage peut résulter de la mise en place exclusive des barrières post-zygotiques et ne nécessite pas la préexistence d'une barrière génétique pré-zygotique exogène comme généralement supposée (Turelli *et al.* 2001; Kirkpatrick & Ravigné 2002; Schemske 2010). Par exemple, il a été récemment démontré que l'isolement reproductif post-

zygotique a évolué avant l'isolement pré-zygotique chez les populations d'eau douce et d'eau salé du poisson *Lucania parva* (Kozak *et al.* 2012). La théorie prévoit aussi que dans le cadre d'une zone de tension, les barrières génétiques endogènes et exogènes sont attendues couplées au niveau de l'écotone, c'est-à-dire la frontière entre les deux habitats (voir Bierne *et al.* 2011). Ce couplage favorise le renforcement de l'isolement reproductif au niveau de la zone de contact, ce couplage étant d'autant plus favorisé que le différentiel de sélection entre les deux habitats est grand. En conséquence, l'étude des marqueurs en déséquilibre de liaison avec l'adaptation locale peut être un bon moyen de détecter l'ensemble des barrières génétiques contribuant au maintien de l'intégrité des génomes parentaux. Chez les pathogènes, le différentiel de sélection étant considéré important, le couplage des barrières exogènes avec les barrières endogènes est particulièrement attendu. Réciproquement, cet attendu suggère que la détection des seules barrières génétiques exogènes est particulièrement complexe chez les organismes subissant des fortes pressions de sélection de l'environnement comme les pathogènes.

3.1.2 Clines génomiques

Les modèles de clines géographiques sont cependant difficile à appliquer car ils nécessitent (i) une structure simple et continue de l'habitat, excluant les situations complexes d'habitats mosaïques (Vines *et al.* 2003; Payseur 2010), (ii) une bonne échelle d'échantillonnage car les résultats peuvent varier en fonction de l'échelle utilisée (Patton 1993; Sites *et al.* 1995). Etant donné que la plus grande proportion des hybrides est générée au centre de la zone hybride, la distance géographique peut être considérée comme une mesure de substitution du degré d'hybridation. Une autre approche consiste donc à utiliser directement la proportion des génomes parentaux afin de prédire l'introgession de différents locus. Cette proportion peut être caractérisée par l'indice hybride (h) de chaque individu dans la zone hybride. En effet, si les génomes parentaux sont connus, la proportion des allèles hérités d'un parent A et d'un parent B peut être estimée. Considérons simplement que $h=0$ correspond à tous les allèles hérités du parent A et $h=1$, tous les allèles sont hérités de B. Si les allèles sont fixés dans chacune des populations et les marqueurs sont co-dominants, h correspond simplement à la fréquence des allèles d'un des deux parents (Barton & Gale 1993). Lorsque les marqueurs sont encore polymorphes dans les populations et/ou dominants, des méthodes de maximum de vraisemblance ou bayésiennes sont disponibles pour estimer h (Pritchard *et al.* 2000a; Falush *et al.* 2003; Buerkle 2005; Falush *et al.* 2007; Durand *et al.* 2009).

Les méthodes basées sur des clines génomiques ont été largement promues par le laboratoire d'Alex Buerkle (Gompert & Buerkle 2009; Gompert & Buerkle 2010; Gompert & Buerkle 2011, 2012; Gompert *et al.* 2012). Les modèles développés permettent de générer une distribution nulle de l'introgession sur l'ensemble du génome à l'aide d'un modèle paramétrique simple, une méthode par permutation, ou plus récemment par l'utilisation d'une approche bayésienne. Ces modèles peuvent être utilisés pour identifier des outliers qui se caractérisent par une introgession significativement différente du reste du génome, permettant ainsi de caractériser des locus fortement introgessés en contexte génomique « étranger » et donc potentiellement marqueurs de régions liées à l'adaptation ou au contraire les locus présentant un cline marqué et candidats à l'isolement reproductif.

Ces clines génomiques représentent une alternative intéressante lorsque la structure spatiale des populations est particulièrement complexe à appréhender comme pour le cas des populations de pathogènes. En effet, les populations de pathogènes sont généralement présentes sur des habitats particulièrement découpés, au moins pour les pathogènes infectant

le compartiment cultivé car le déploiement de ces populations est dépendant même de la structure mosaïque des agrosystèmes. De plus, les populations d'hôtes sont généralement mal caractérisées et les attendus de spécialisation des populations de pathogènes difficile à formuler.

3.2 Détection des barrières par cartographie génétique

La divergence des populations et la mise en place des barrières génétiques peuvent aussi être caractérisée par une approche de cartographie génétique. Bien que cette stratégie soit souvent utilisée pour rechercher les caractères monogéniques ou polygéniques ('quantitative trait loci') associés à l'adaptation locale, c'est-à-dire les barrières exogènes ; ces approches peuvent aussi détecter les barrières endogènes (Hall & Willis, 2005 ; ci-dessous). La construction d'une carte nécessite la réalisation d'un croisement *in vitro* entre deux souches divergentes, l'obtention d'un nombre suffisant d'hybrides entre deux populations divergentes puis l'échantillonnage d'un grand nombre de marqueurs aléatoirement répartis sur le génome. Outre la localisation des marqueurs sur les groupes de liaisons, cette approche permet surtout de détecter des marqueurs s'écartant significativement des ratios attendus de ségrégation mendélienne. De nombreuses études ont mis en évidence que les marqueurs distordus ne sont pas aléatoirement répartis sur les groupes de liaisons, suggérant que certains locus pourraient être à l'origine de ce processus (Jiang *et al.* 2000; Fishman *et al.* 2001; Schwarz-Sommer *et al.* 2003; Myburg *et al.* 2004; Solignac *et al.* 2004; Hall & Willis 2005) et que le nombre de ces zones génomiques est proportionnel au temps de divergence entre les deux populations étudiées (Zamir & Tadmor 1986; Causse *et al.* 1994; Palopoli & Wu 1996; Jenczewski *et al.* 1997; Whitkus 1998; Harushima *et al.* 2001; Taylor & Ingvarsson 2003). Une des limites de cette approche est que la localisation des barrières peut être fortement affectée si la sélection est forte car la distorsion générée peut fortement affecter la localisation des marqueurs sur la carte génétique.

Une seconde limite est que les écarts observés par rapport à la ségrégation mendélienne attendue peuvent être la résultante de barrières exogènes (maladaptation des hybrides) mais aussi de barrières endogènes (de la stérilité ou mortalité hybride à cause des épistasies négatives de type Dobzhansky-Muller, des réarrangements chromosomiques ; Rieseberg *et al.* 1995) ou encore de biais méiotiques (Pardo-Manuel de Villena *et al.* 2000). Toutefois chez certains organismes comme les microorganismes dont le cycle de vie peut être maîtrisé *in vitro*, l'extraction de l'organisme de son environnement naturel permet de s'affranchir des pressions de sélection environnementales et permet ainsi la seule détection des barrières endogènes.

3.3 Conséquences potentielles des remises en contact secondaire pour les populations de pathogènes

3.3.1 Hybridation & paysage adaptatif

Les conséquences des remises en contact secondaire sont largement ignorées chez les pathogènes. Pourtant dans ces zones, la recombinaison génère de nouvelles combinaisons génétiques qui n'ont pas été testées par la sélection. L'intégration de l'évolution de l'agressivité des pathogènes (mais aussi l'acquisition de virulence et, dans le cas des agrosystèmes, de résistances aux pesticides) dans des concepts de paysage adaptatif est

pourtant particulièrement informative pour formuler des attendus sur l'évolution des populations de pathogènes et intégrer les hypothèses écologiques et les hypothèses plus neutres.

L'hybridation entre populations divergentes peut être un formidable moteur de diversité. Si la plupart des combinaisons génétiques sont attendues au niveau des « trous » du modèle de Gavrilets (1997a) de part l'expression des incompatibilités de Dobzhansky-Muller (voir encadré III), certaines combinaisons peuvent explorer des nouvelles zones du paysage et accéder à des meilleurs valeurs sélectives que les populations parentales. Chez les pathogènes, la fécondité des organismes étant particulièrement forte, l'exploration du paysage adaptatif peut être rapide dans les zones d'hybridation. Le taux de mutation est lui-même généralement considéré comme particulièrement fort chez ces organismes, ce qui sous-entend qu'après découverte d'un nouveau pic adaptatif, les génotypes pourraient rapidement accéder à l'optimum local par mutation.

Encadré III : Paysage adaptatifs

Un paysage adaptatif correspond à l'ensemble des valeurs sélectives des individus ayant toutes les combinaisons génétiques possibles. Sous cette représentation, un génotype correspond à un point et une population à un nuage de points. Son espace est multidimensionnel, du même ordre de grandeur que le nombre de locus, et discret, puisque le nombre d'allèle à chaque locus est fini. Par commodité, Wright (1931) a représenté cet espace comme une surface continue et dans un espace tridimensionnel où les deux premières dimensions correspondent aux génotypes possibles et la troisième à la valeur sélective. Sa représentation classique génère des montagnes et des crêtes (génotypes avec des valeurs sélectives élevées) ou des vallées (valeurs sélectives faibles). Un des grands enseignements de cette théorie est que le déplacement des individus dans ce paysage n'est pas aléatoire, les génotypes se déplaçant beaucoup plus facilement vers un optimum local par des petits changements génétiques (pics locaux) que vers un autre pic adaptatif distant. En effet, puisque la sélection ne peut pas conduire vers une réduction de la valeur sélective, la redescente vers les vallées adaptatives est beaucoup plus rare et nécessite soit un fort goulot d'étranglement où la dérive génétique est plus forte que la sélection, soit la recombinaison qui a pour effet de redéployer totalement les individus d'une population dans l'espace possible des génotypes. Néanmoins, la traversée de ces vallées adaptatives peut autoriser l'exploration de nouveaux pics adaptatifs dont certains peuvent présenter une plus forte valeur adaptative que les pics locaux initiaux et ainsi contribuer à l'évolution des populations.

Un modèle bien plus complexe a été proposé par Gavrilets (1997a) pour tenir compte de la plus grande connectivité entre les génotypes possibles en intégrant un plus grand nombre de dimensions. Gavrilets suggère que les individus à forte valeur sélective sont beaucoup plus connectés entre eux par des modifications d'un seul allèle que ce que Wright proposait, ce qui impliquerait de nombreux « plateaux » dans le paysage, formant un réseau d'individus à forte valeur sélective. Il dénomme ce modèle « le paysage adaptatif à trous ». Ces « trous » sont produits par des individus à très faible valeur sélective directement connectés à des individus à plus forte valeur sélective comme attendu dans le cas de barrières de type Dobzhansky-Muller. En effet, considérons le modèle classique d'une épistasie négative dans un modèle à deux locus bi-alléliques avec $A_1A_1B_1B_1$ et $A_2A_2B_2B_2$, ces deux génotypes sont largement connectés entre eux par toutes les combinaisons délétères exprimant les incompatibilités A_1B_2

ou A_2B_1 . Ce modèle permet donc de connecter des attendus d'adaptation locale et des hypothèses plus neutres de divergence. La divergence favorise l'établissement de génotypes de forte valeur sélective mais voisins de trous car pouvant générer facilement des individus porteurs de faible valeur sélective par peu d'événements de « mutation ». Selon Gavrillets, le paysage adaptatif « classique » ne serait observable que si les génotypes n'ont pas encore atteint le réseau des génotypes à forte valeur sélective. Bien que particulièrement difficile à appréhender, ce dernier modèle a un intérêt particulier dans son rôle intégratif des barrières génétiques endogènes et exogènes.

3.3.2 *Importance d'intégrer ces hypothèses alternatives*

Le cadre conceptuel de l'évolution des populations de pathogènes est essentiellement lié aux pressions de sélection de son hôte. L'intégration des hypothèses alternatives, considérées pourtant comme générales chez les autres organismes biologiques, peut permettre d'expliquer certains autres attendus chez les populations de pathogènes.

Premièrement, l'utilisation des concepts de génétique des populations en phytopathologie a été réalisée plus tardivement que chez d'autres organismes. La structure des populations a été largement reportée chez de nombreux pathosystèmes au cours de la dernière décennie. L'étape suivante est généralement d'identifier les gènes responsables de cette différenciation notamment par l'utilisation d'approches basées sur l'hétérogénéité de F_{ST} en fonction des régions du génome. Cependant, les barrières exogènes ne sont pas seules à empêcher le flux de gènes et à présenter des forts indices de fixation. Les barrières endogènes s'accumulant au fur et à mesure que le temps de divergence est grand, la détection d'un grand nombre de barrières endogènes en déséquilibre de liaison avec les barrières exogènes est attendue le modèle de Lemaire *et al.* (en préparation & encadré IV). Or la recherche des bases génétiques de la différenciation est l'étape la plus longue. Considérons par exemple qu'une structure génétique a été préalablement détectée entre deux populations infectant deux hôtes différents, un résistant (porteur d'un gène de résistance majeur) et un sensible. Si les deux populations infectant ces hôtes ont bien *a priori* des allèles différents au locus d'avirulence, la divergence en allopatrie entre ces deux populations peut avoir autorisé la fixation de nombreux autres allèles incompatibles entre les deux fonds génétiques et contribuer au maintien de la différenciation génétique. La première conséquence de cette situation est que le temps nécessaire à la découverte du gène d'avirulence sera très sensiblement allongé. La seconde est que la contribution du gène d'avirulence à ce système sera surestimée, notamment concernant le coût de fitness associé à la virulence.

Deuxièmement, la recombinaison de populations génétiques divergentes peut être un facteur d'introgession de différents allèles entre les fonds génétiques. Le transfert de certains gènes impliqués dans la virulence, dans les facteurs d'agressivité peut ainsi sensiblement modifier les caractéristiques épidémiques de la maladie. Le modèle de Lemaire *et al.* (en préparation) associé au cadre conceptuel des zones hybrides permet ainsi de formuler des attendus sur l'évolution des populations de pathogènes.

Encadré IV : Scénarios évolutifs d'émergence des populations virulentes à partir du compartiment non-cultivé (De Gracia, 2011 ; Lemaire *et al.* en préparation)

Le modèle de Lemaire *et al.* (en préparation) suggère que le compartiment non-cultivé peut être une source importante de virulences. En effet, les auteurs ont utilisé une approche ABC (« Approximate Bayesian Computation » ; Tavaré *et al.* 1997 ; Beaumont *et al.* 2002) pour tester six scénarios différents d'émergence de la virulence *virRvi6* sur les pommiers porteurs du gène de résistance *Rvi6* (3 scénarios de mutations vers la virulence, et 3 scénarios de migration de la virulence à partir du compartiment non-cultivé, Figure 8).

Cette approche ABC permet de réfuter l'hypothèse d'une acquisition d'une nouvelle virulence à partir de la population *avrRvi6* résidente à l'intérieur des agrosystèmes (scénarios « 1-3 »). Cette approche valide l'hypothèse d'une migration de souches préexistantes à partir de *M. x floribunda* présent hors des agrosystèmes (scénarios 4-6). Le scénario le plus vraisemblable (scénario 6) suggère au moins trois événements différents de migration vers les agrosystèmes, plutôt qu'un ou deux événements (scénarios 4-5). En supposant une génération de reproduction sexuée par an et un taux de mutation de 5×10^{-4} , le temps de divergence entre les populations *avrRvi6* sur pommier et *virRvi6* sur *M. x floribunda* a été estimée à environ 1000 ans (mode : 973 ans, intervalle de confiance à 95% : 717-994 ans ; De Gracia, 2011). De nouvelles simulations suggèrent toutefois que cette divergence pourrait être un peu plus ancienne (1000-10000 ans, Lemaire *et al.* en préparation).

Sous ce modèle, les populations avirulentes résidentes dans les agrosystèmes et les souches virulentes préexistantes dans le compartiment non-cultivé présentent un certain degré de divergence. En conséquence, le changement rapide du régime de migration entre avirulentes et virulentes (allopatrie vers parapatrie) associé à cette émergence peut autoriser la remise en contact des populations et l'établissement d'une zone d'hybridation (Figure 9).

Comme nous l'avons précédemment vu, la description de la nature des barrières d'isolement reproductif est particulièrement complexe. Les approches de génomique des populations basées sur le génotypage d'un grand nombre de locus de populations dans les zones d'hybridation naturelles permettent de mieux discriminer la localisation des barrières génétiques grâce à un plus grand nombre de générations de recombinaison que pour les approches de cartographie génétique basées sur des croisements. *A contrario*, cette dernière méthode semble plus propice à une meilleure caractérisation de la nature de ces barrières si les croisements sont obtenus *in vitro*. Ces deux méthodes seront simultanément utilisées (article 1), d'abord en ciblant certains outliers par génomique des populations puis ensuite en caractérisant leur ségrégation sur une F1 obtenue *in vitro*. Ce travail sera réalisé pour caractériser les barrières d'isolement reproductif entre deux populations, la population *avrRvi6* native sur pommier et la population *virRvi6* originaire du compartiment non-cultivé (figure 9). Avant de décrire les résultats obtenus, les particularités du pathosystème *V. inaequalis*/pommier et ensuite de manière plus précise, l'histoire de l'introggression de la résistance *Rvi6* et de son contournement seront décrites dans le chapitre suivant.

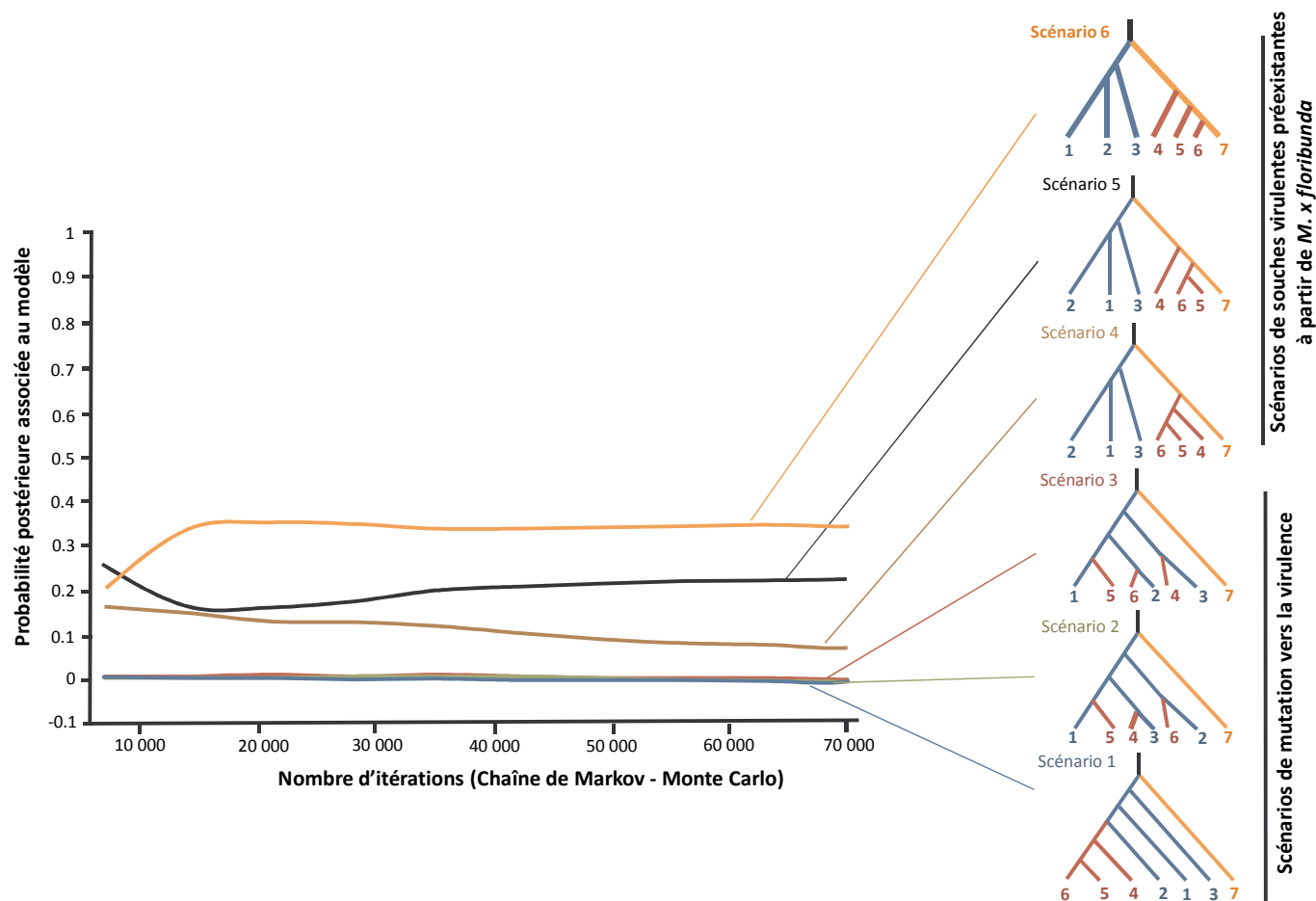


Figure 8 : Comparaisons des probabilités postérieures associées aux différents scénarios d'émergence de la virulence-Rvi6 dans les agrosystèmes [modifié d'après De Gracia (2011) et Lemaire et al. en préparation]. Les populations isolées de pommier domestique (1-6) ont été échantillonnées dans trois zones géographiques différentes (Danemark, Pologne et France) présentant des hôtes Rvi6 et non-Rvi6 en sympatrie. Le groupe 7 comprend des accessions isolées de *M. x floribunda* à travers l'Europe. Les branches bleues correspondent à des populations avrRvi6, les branches rouges des populations virRvi6 isolées de *M. x domestica* et orange les accessions de *M. x floribunda*. Les numéros des populations correspondent respectivement à : 1 - avrRvi6 du Danemark ; 2 - avrRvi6 de Pologne ; 3 - avrRvi6 de France (Yvrandes) ; 4 - virRvi6 de France (Yvrandes) ; 5 - virRvi6 du Danemark ; 6 - virRvi6 de Pologne ; et 7 - virRvi6 isolées de *M. x floribunda*.

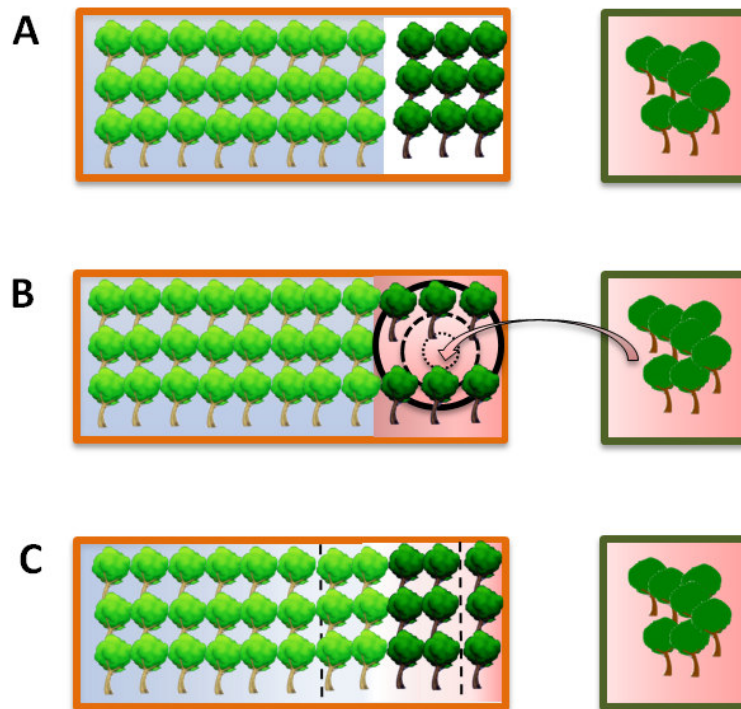


Figure 9 : Emergence de virulences à partir du compartiment sauvage et attendus de contact secondaire. L'introgession d'un gène de résistance R à partir d'un hôte sauvage aboutit à l'acquisition d'une résistance envers les populations Avr natives (en bleu) dans le compartiment cultivé (cadre orange). Les populations Avr étaient initialement présentes sur un hôte sensible (arbres vert clair). Initialement ce gène de résistance est toutefois déjà contourné sur un hôte du compartiment sauvage (cadre vert) (A). La migration des souches virulentes préexistantes peut alors autoriser l'infection des hôtes résistants dans les agrosystèmes. En l'absence de compétition, les souches virulentes augmentent rapidement en fréquence sur les hôtes résistants (B). Si l'isolement reproductif est incomplet, la remise en contact des deux populations peut alors s'accompagner de flux génique entre les deux populations (C). Les deux lignes pointillées délimitent la zone d'hybridation.

4. PATHOSYSTEME *VENTURIA INAEQUALIS*/ POMMIER

L'agent pathogène *Venturia inaequalis* infecte une large gamme d'hôte de la famille des *Rosaceae*. Son hôte le plus étudié est le pommier domestique (*Malus x domestica*). Cependant ce champignon pathogène a aussi été décrit sur d'autres hôtes du genre *Sorbus*, *Cotoneaster*, *Crataegus*, *Heteromeles*, *Kageneckia*, *Pyracanthae* et *Eryobotrya* (Menon 1956; Raabe & Gardner 1972; Sivanesan 1977; Le Cam *et al.* 2001). Dans ce chapitre, je me focaliserai sur l'interaction entre *V. inaequalis*, le pommier domestique et les espèces du genre *Malus*.

4.1 Pommier domestique

Le genre *Malus* serait apparu vers la fin de la période du Tertiaire, à l'époque Eocène, il y a environ 30 millions d'années dans une zone se situant vraisemblablement au sud de la Chine (Phipps *et al.* 1991; Juniper & Mabberley 2006). Ce genre compterait environ 40 espèces différentes, mais ce nombre varie énormément selon les auteurs (8 à 79, Juniper & Mabberley, 2004). Cette méconnaissance tire probablement sa source d'une description morphologique des espèces. Cette différence résulte aussi de l'importance de l'hybridation et de la polyploidisation chez les *Malus* et de la description historique de nombreuses espèces sauvages différentes sous un seul et même nom ('crab apple', 'wild apple'). L'utilisation de marqueurs moléculaire a toutefois permis de préciser la taxonomie de certaines des espèces de *Malus* (Robinson *et al.* 2001; Velasco *et al.* 2010; Cornille *et al.* 2012; Dunemann *et al.* 2012) La plupart des études se sont toutefois focalisées sur la contribution de deux espèces sauvages *M. sylvestris* et *M. sieversii*, endémique de l'Asie Centrale et de l'Europe respectivement, au pool génétique du pommier cultivé.

4.1.1 Origine du pommier cultivé

Le nom *M. x domestica* est le nom généralement accepté comme le nom scientifique approprié pour le pommier cultivé (Korban & Skirvin 1984). Le « x » faisant référence au caractère hybride du pommier domestique qui fut assez rapidement suspecté. Dès les premières prospections menée par Nikolay Vavilov dans les années 20, une forte diversité a été décrite chez un pommier sauvage d'Asie centrale *M. sieversii* (Janick *et al.* 1996). Ce pommier sauvage s'étend sur une large aire de répartition, des bords de la mer Caspienne aux montagnes du Tian Shan. Cette espèce possède l'ensemble de la diversité des formes, couleurs et goût présente aujourd'hui chez le pommier domestique (Forsline *et al.* 1994; Forsline 1995). La forte ressemblance entre les fruits de l'espèce sauvage *M. sieversii* et ceux du pommier domestique a alors poussé Nikolay Vavilov à suggérer que *M. sieversii* était le principal progéniteur du pommier cultivé. Bien que cette approche était à l'origine spéculative et basée sur peu de données, les analyses phylogénétiques de séquences d'ADN nucléaires et chloroplastiques ont ensuite permis de confirmer ce scénario (Robinson *et al.* 2001; Harris *et al.* 2002). Plus récemment, les apports apportés par la publication de la séquence du génome du pommier et plus particulièrement le séquençage de 23 gènes de ménage de 74 accessions de *Malus* a nouveau validé cette hypothèse (Velasco *et al.* 2010). Selon ces derniers auteurs, les deux fonds génétiques sont si proches qu'ils appartiennent à une seule et même espèce *M. pumila*, préconisant ainsi de suivre la nomenclature initialement proposée par Juniper & Mabberley (2006).

A la fin du néolithique (il y a environ 4 500 ans), les marchands et leurs animaux domestiques (essentiellement des chevaux et des chameaux) empruntaient les grandes routes commerciales reliant l'Europe à la Chine, connues sous le nom des routes de la Soie, qui traversaient l'aire de répartition de *M. sieversii*. Au cours de ces déplacements, les voyageurs et leurs animaux ont pu disséminer pommes et pépins de part et d'autre de cette route (vers la Chine à l'Est ou vers l'Europe à l'Ouest). La sortie de *M. sieversii* de son aire d'origine aurait pu permettre à d'autres espèces sauvages de s'hybrider avec *M. sieversii* et ainsi de contribuer au pool génétique du pommier cultivé. L'hybridation potentielle avec de nombreuses espèces de *Malus* telles que *M. orientalis*, *M. baccata* (dont les sous-espèces ou hybrides naturels *M. prunifolia*, *M. mandshurica*, *M. asiatica*) a été proposée par certains auteurs (Janick *et al.* 1996; Forsline *et al.* 2003; Cornille *et al.* 2012) et réfutée par d'autres (Juniper & Mabberley 2006).

La contribution du pommier sauvage Européen *M. sylvestris* au pool domestique du pommier était toute aussi débattue (Wagner & Weeden 2000; Coart *et al.* 2006). De récents travaux ont montré que cette espèce fut un important contributeur secondaire au pool domestique Européen (Cornille *et al.* 2012). Le pommier sauvage *M. sylvestris* présente une aire de répartition s'étendant de l'Europe de l'Est à l'Europe de l'Ouest. Cette espèce pousse de façon isolée en forêt formant parfois des petits bosquets et produit des fruits amers et astringents, utilisés bien avant l'introduction du pommier domestique par les Celtes pour la production de cidre. Les différences observées entre les études dans la contribution de *M. sylvestris* pourrait être assez dépendante des cultivars de *M. x domestica* pris en compte dans ces études (cultivars Européens ou Asiatiques).

La domestication du pommier a fortement bénéficié de la découverte de l'art du greffage, qualifié parfois de 'domestication instantanée' (Zohary & Hopf 2000), probablement en Mésopotamie il y a 3800 ans (Harris *et al.* 2002; Juniper & Mabberley 2006). Cette technique aurait en effet rendu possible le transfert des meilleurs génotypes des pommiers sauvages vers les jardins par clonage. Étonnamment, aucun goulet d'étranglement n'a pour autant été détecté au moment de la période de découverte du greffage suggérant que la diversité a été quand même maintenue (Cornille *et al.* 2012). Le pommier aurait ensuite été introduit en Europe et en Afrique du Nord par les Grecs et les Romains il y a environ 2000-2500 ans (Forsline *et al.* 2002 ; Juniper et Mabberley 2006) et l'Europe aurait ensuite servi de tête de pont pour la dissémination du pommier à travers le monde après la découverte des nouveaux continents (Morgan & Richards 1993).

4.1.2 *Biologie et génétique du pommier*

La reproduction du pommier est essentiellement allogame, et soumise à une incompatibilité de type gamétophytique (Lespinasse 1990). Chez certains cultivars, l'autofécondation est parfois possible mais le taux de réussite est faible et les descendants semblent posséder une faible vigueur (Janick *et al.* 1996). Ce mode de reproduction allogame est à l'origine de la forte variabilité et hétérozygotie observée chez les pommiers. En vergers commerciaux, la pollinisation est effectuée par des variétés uniquement pollinisatrices (dont le cv. 'Everest' obtenu à l'INRA d'Angers) ou par l'utilisation d'un panel de cultivars de pommiers ayant des allèles d'incompatibilités différents.

Le pommier cultivé tolère une large gamme de températures, de sols ou de latitudes permettant une culture dans la plupart des zones tempérées (Morgan & Richards 1993). La culture a ensuite été étendue à des climats plus froids ou plus chauds. Dans les régions plus froides, le pommier sauvage Sibérien *M. baccata* a été utilisé comme porte-greffe pour améliorer la tolérance du pommier aux conditions de températures des régions les plus

froides. *A contrario*, dans les régions les plus chaudes, le besoin en froid hivernal du pommier pour initier la floraison a été contourné dans certains pays par la défoliation manuelle ou chimique des feuilles après la récolte des fruits (Taylor *et al.* 1984; Koutinas *et al.* 2010). De plus, des variétés ayant un plus faible besoin en froid, adaptés aux pays aux hivers peu rigoureux, ont aussi été sélectionnées en croisant d'anciennes variétés méditerranéennes et des variétés modernes (Carter 2007).

Le pommier possède n=17 chromosomes. Le pommier semble avoir résulté d'une fusion des génomes des *Prinoïdeae* [une sous-famille comprenant des espèces domestiquées tels que le prunier, le cerisier ou le pêcher (n=8)] -et des *Spiraeoïdeae* [une sous famille comprenant des arbustes à fleurs tels que les sorbaires ou les spirées n=9] par allopolyploïdisation (Chevreau *et al.* 1985; Chevreau & Laurens 1987; Zhang *et al.* 1988; Maliepaard *et al.* 1998). Bien que la majorité des cultivars actuels de pommiers soient diploïdes, quelques variétés sont triploïdes comme 'Belle de Boskoop' ou 'Reinette du Canada' (Lespinasse 1990).

4.1.3 Importance économique

Le pommier est cultivé dans tous les pays tempérés des deux hémisphères (Morgan & Richards 1993; Juniper & Mabblerley 2006). Le pommier est, après la banane, le deuxième fruitier produit dans le monde, et le deuxième en Europe après la vigne (FAOSTAT 2012). En 2010, la production mondiale de pomme approchait 70 millions de tonnes par an dont près de la moitié (47,8%) uniquement par la Chine (FAOSTAT, 2012). Les Etats-Unis d'Amérique (6,1% de la production mondiale) et la Turquie (3,7%) sont respectivement deuxième et troisième producteurs mondiaux. La France est actuellement le septième producteur mondial avec 1,7 millions de tonnes produites par an (2,5%) et sa culture s'étalerait sur environ 40 000 hectares. Les vergers français sont encore majoritairement composés de 'Golden Delicious' (Juniper & Mabblerley 2006; EUROSTAT 2007).

4.2 La tavelure du pommier

4.2.1 Importance économique

La tavelure est la maladie du pommier la plus préjudiciable économiquement. Elle est présente dans toutes les zones de production du pommier mais son importance est moindre dans les régions plus arides (MacHardy 1996). Les attaques sont plus sévères lorsque le climat est humide et frais au printemps. Le développement de la maladie est influencé par de nombreux facteurs tels que l'état sanitaire du verger, sa topographie, la sensibilité des cultivars utilisés, la fréquence des périodes d'infection.

Les pertes économiques liées à la tavelure résultent principalement des infections du fruit, détériorant leur aspect visuel et les rendant impropres à la commercialisation, bien que potentiellement consommables. En effet, l'impact économique de cette maladie est surtout lié au fait qu'aucun symptôme de tavelure n'est toléré dans les circuits de distribution des pays occidentaux et à l'exportation et au fait que plus de 90% de la production mondiale est assurée par des variétés sensibles à la tavelure. En outre, les défoliations répétées provoquées par la maladie affecteraient aussi la vigueur, le rendement et la résistance aux froids hivernaux de l'arbre.

Outre ces pertes d'exploitation, le contrôle de la tavelure nécessite aussi un apport conséquent en fongicides, que ce soit en agriculture conventionnelle, raisonnée ou biologique

(Holb 2006). Le nombre de traitements fongicides contre la tavelure varie fortement en fonction des conditions météorologiques, de la pression inoculum et des pratiques culturales (par exemple en agriculture raisonnée : 6-16 ; biologique : 10-26 ; Ellis *et al.* 1998; Holb *et al.* 2003a ; Beresford et Manktelow 1994; MacHardy 1996; Holb *et al.* 2003b). Dans un contexte général de réduction des traitements phytosanitaires, d'autres méthodes de lutttes sont désormais employées contre la tavelure telles que la prophylaxie, la modification des méthodes culturales, la lutte biologique et la résistance variétale. Les coûts liés aux différents moyens de lutte contre cette maladie sont conséquents et doivent être intégrés au coût global de cette maladie, même s'ils restent nettement moindres que les coûts engendrés par la main d'œuvre.

4.2.2 Agent pathogène : *V. inaequalis*

4.2.2.1 Taxonomie

V. inaequalis est le champignon responsable de la tavelure du pommier. Ce champignon est classé de la façon suivante : division des Eucomycètes (thalle cloisonné filamenteux), subdivision des Ascomycètes (reproduction sexuée via des asques contenant des ascospores), classe des Dothideomycètes (asques à deux tuniques en série parallèle enfermés dans un périthèce s'ouvrant à maturité par un ostiole, pas de paraphyses vraies), ordre des Venturiales (périthèces globuleux ou peu allongés de couleur foncée, pourvus de pseudoparaphyses), famille des Venturiaceae (périthèces enfoncés dans les feuilles ou les fruits, ascospores bicellulaires) (Sivanesan 1977; Berbee *et al.* 2000; James *et al.* 2006; Winton *et al.* 2007; Zhang *et al.* 2011). La classe des Dothideomycetès comprend d'autres phytopathogènes majeurs comme *Phaesophaeria nodorum*, *Mycosphaerella graminicola* (septorioses du blé) et *Leptosphaeria maculans* (phoma du colza) (Winton *et al.* 2007) . Des approches de phylogénie basée sur le séquençage de l'ADN ribosomal 28S ont été récemment réalisées sur plusieurs espèces de la classe des Dothideomycetès (Zhang *et al.* 2009; Zhang *et al.* 2011) et suggère que *M. graminicola*, *L. maculans* et *V. inaequalis* appartiennent à trois ordres différents (Zhang *et al.* 2011, Figure 10).

Le genre *Venturia* compterait 52 espèces morphologiques (Sivanesan, 1977). La phylogénie du genre *Venturia* est mal-connue. L'étude la plus approfondie s'appuie sur le séquençage de l'ADN ribosomal (18S et 5,8S) pour évaluer les relations entre six espèces seulement (Schnabel *et al.* 1999). Cette étude a permis toutefois l'identification de trois clades principaux: des séquences de *V. inaequalis* (sur pommier), des séquences de *V. nashicola* et *V. pirina* (sur nashi et poirier Européen respectivement et, des séquences de *V. asperata*, *V. cerasi* et *V. carpophila* (sur pommier, cerisier et pêcher). Les familles végétales les plus représentées parmi les hôtes de *Venturia* spp sont les *Rosaceae* et les *Salicaceae*. Au sein de *V. inaequalis*, des formes spéciales pathogènes sur pommier et sur pyracantha ont été décrites (Le Cam *et al.* 2002) *V. asperata* est un champignon classiquement décrit comme saprotrophe sur la litière foliaire du pommier. Une récente étude montre que ce pathogène peut émerger sur pommier cultivé en absence de compétition avec *V. inaequalis*, c'est-à-dire lorsque la niche écologique est vide (Caffier *et al.* 2012).

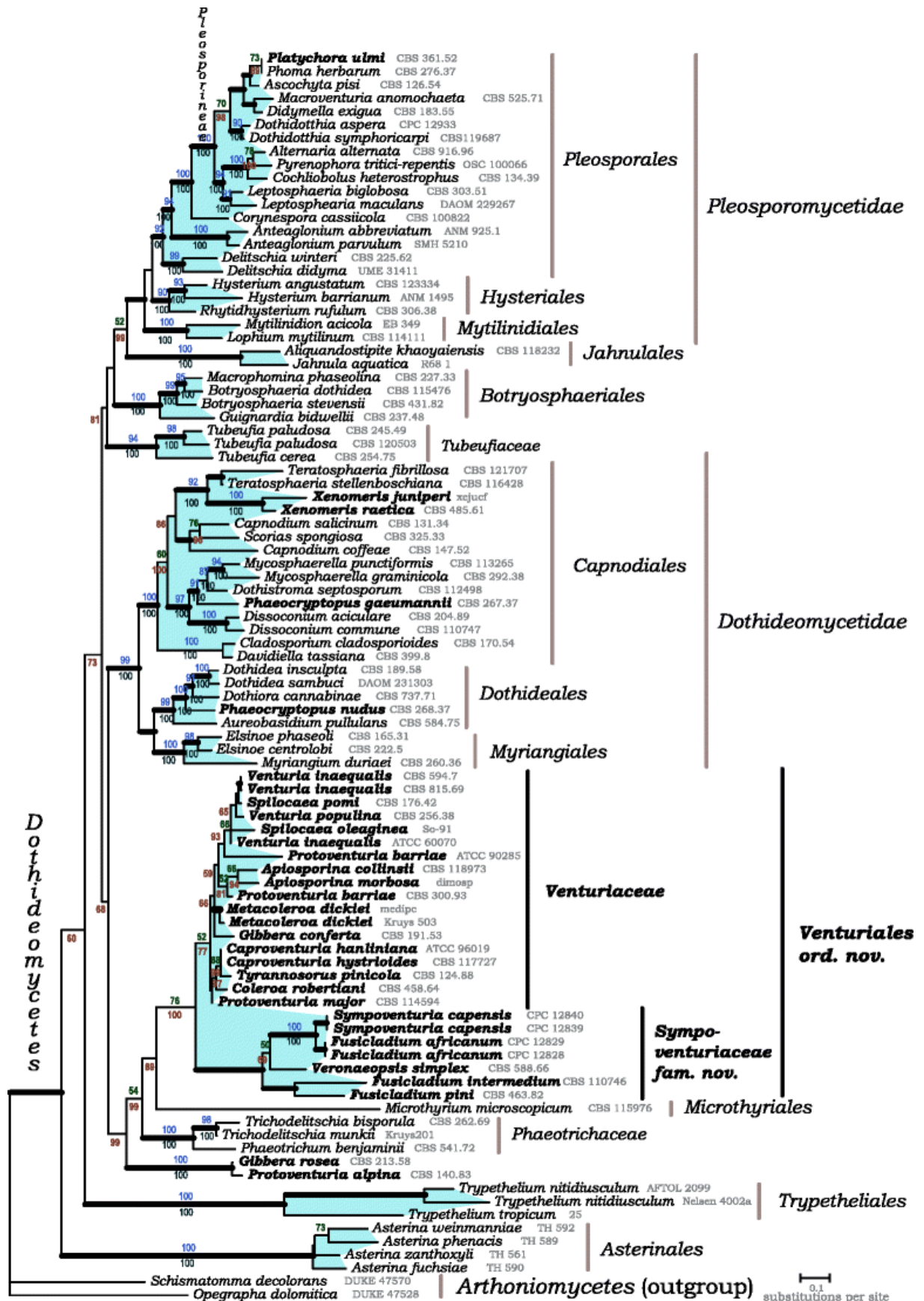


Figure 10: Phylogénie de la classe des Dothidéomycètes basée sur le séquençage du locus ribosomal 28S (Zhang et al. 2011).

4.2.2.2 Symptomatologie

Les symptômes dus à l'infection par *V. inaequalis* sont principalement visibles sur les feuilles et les fruits (Figure 11), mais ils apparaissent aussi sur les pétioles, les sépales, les fleurs, les jeunes pousses et les bourgeons.

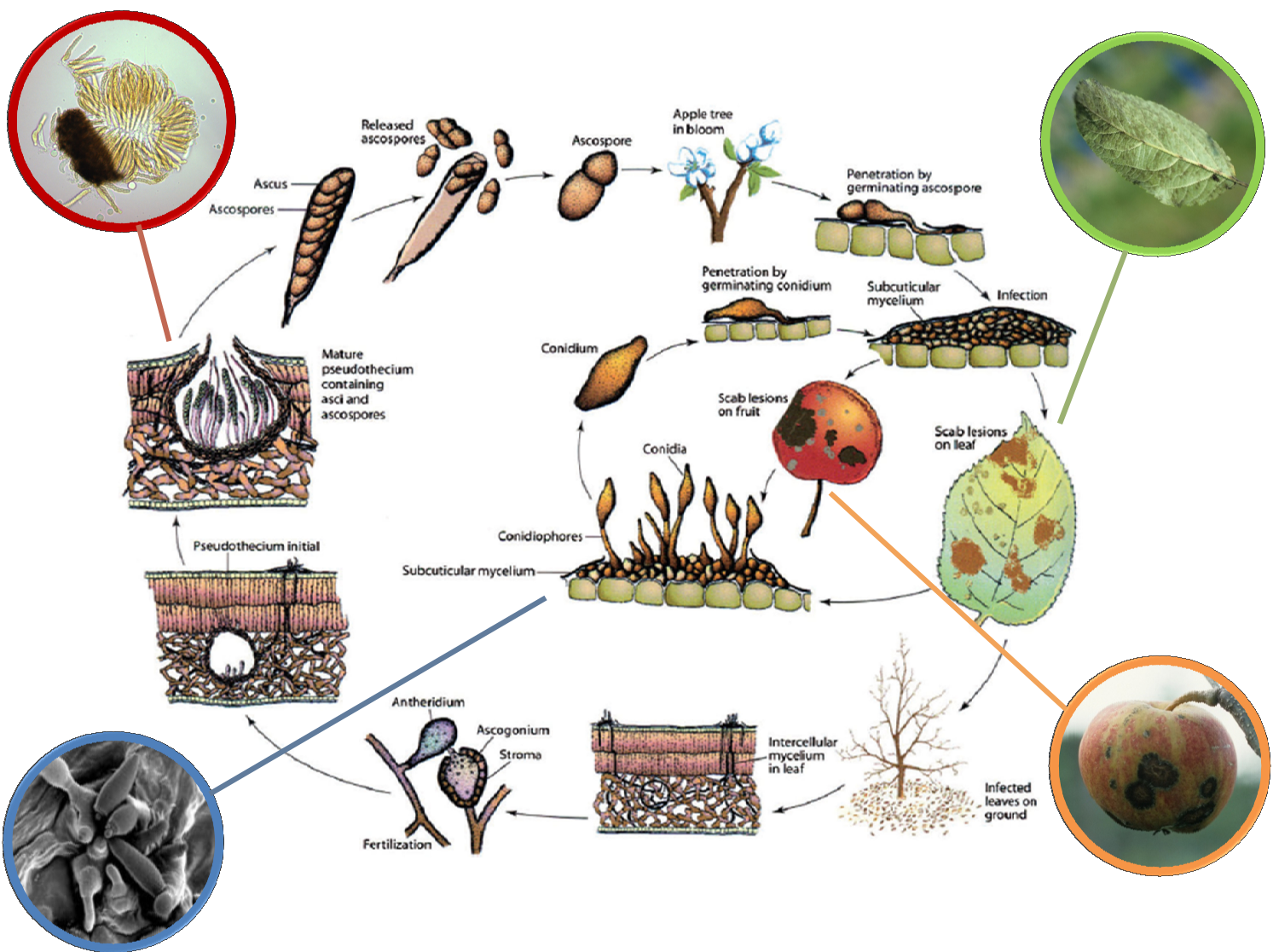


Figure 11 : Cycle biologique de *Venturia inaequalis* (modifié d'après Agrios, 2005). Le cycle de vie de *V. inaequalis* comprend une phase sexuée annuelle, unique et obligatoire, et de multiples phases de reproduction asexuées. Les différentes photographies illustrent la production de spores sexuées (contour rouge), asexuées (bleu) et les symptômes sur feuille (vert) et fruits (orange). Photos : photothèque INRA et Le Cam.

Sur feuilles, le champignon produit des lésions nécrotiques ou chloriques qui peuvent être isolées, dispersées sur toute la feuille ou recouvrir la surface entière. Les lésions se développent généralement plus rapidement sur les jeunes feuilles que sur les feuilles âgées à cause de l'existence d'une résistance ontogénique. Les tâches sont d'abord de petites tailles, circulaires et translucides. Elles s'agrandissent ensuite et deviennent de couleur brun foncé et prennent un aspect velouté par l'abondance de la conidiogénèse. Au cours du vieillissement de la lésion, les feuilles vont se dessécher ce qui induit des déchirures du limbe entraînant une défoliation plus ou moins sévère en fonction de la gravité de la maladie.

De la même façon que sur feuille, les fruits sont plus sensibles à un stade juvénile. Les lésions s'y développent plus rapidement que sur un fruit mûr. Quand un jeune fruit est infecté, il se déforme et l'infection du pédicelle peut conduire à la chute du fruit. En cas d'attaque très tardive, les lésions peuvent parfois être invisibles à l'œil nu à la récolte mais apparaître lors de la conservation des fruits.

4.2.2.3 Cycle de vie

V. inaequalis est un champignon hémibiotrophe. Son cycle de vie complet comprend deux phases : une phase parasitaire au cours de laquelle se déroulent plusieurs cycles de reproduction asexuée et une phase saprophytique sur la litière foliaire au cours de laquelle a lieu la reproduction sexuée (Figure 11). Le développement du champignon au cours de ces deux phases est très dépendant des conditions climatiques. En effet dans les régions à hiver peu rigoureux, le pommier ne perd pas toutes ses feuilles et pourrait ainsi assurer une partie de sa survie hivernale selon le mode clonal.

L'infection débute au printemps lorsque les conditions de température et d'humidité déclenchent la projection d'ascospores sur les jeunes organes de l'hôte. Les ascospores ne sont projetées qu'à une très faible distance (quelques millimètres), elles sont ensuite dispersées par le vent sur des distances relativement modestes de l'ordre de quelques mètres ou dizaines de mètres (Aylor 1998 ; Leroy *et al.* en révision, voir annexe ; MacHardy *et al.* 2001). Des simulations suggèrent que des événements de dispersion à plus longue distance peuvent se produire dans des conditions très favorables à la dispersion (pluviométrie, fort vent) et sous une forte pression de l'inoculum mais sans jamais dépasser quelques kilomètres (Aylor 1998). En présence d'eau libre, les ascospores germent, initient un tube germinatif et pénètrent sous la cuticule après la formation d'un appressorium (Smereka *et al.* 1987; Jha *et al.* 2009) Chez *V. inaequalis*, la sécrétion enzymatique de cutinases semble suffisante à la rupture de la cuticule sans *a priori* nécessiter de pénétration mécanique (Köller & Parker 1989; Köller *et al.* 1991). Après pénétration, l'hyphe d'infection se différencie en stroma sous-cuticulaire à partir de laquelle un mycélium puis des conidiophores vont être abondamment produits. Ces conidiophores, porteurs de conidies, vont alors rompre l'épiderme et ainsi libérer les conidies qui constitueront l'inoculum secondaire. Les conidies sont dispersées par le vent ou par la pluie sur de courtes distances (quelques mètres). Les capacités intrinsèques de migration des ascospores et conidies étant relativement faibles, les migrations à longue distance se produisent principalement par le transport de matériel infecté et ont donc une origine anthropique.

4.2.2.4 Structure des populations

La structure génétique des populations de *V. inaequalis* a été étudiée à plusieurs échelles : à l'échelle de l'arbre (Tenzer 1997), à l'échelle du verger (Tenzer & Gessler, 1997a ; Tenzer et Gessler, 1997b ; Guérin and Le Cam, 2004, Guérin et al., 2007 ; Leroy *et al.* en révision, voir annexe), à l'échelle du pays (Tenzer & Gessler 1997), à l'échelle Européenne (Tenzer *et al.* 1999; Tenzer & Gessler 1999), à l'échelle eurasienne (Gladieux *et al.* 2010) ou à l'échelle mondiale (Gladieux *et al.* 2008).

Les études effectuées par l'équipe suisse suggèrent que la diversité génétique et génotypique est élevée à toutes les échelles, confirmant l'efficacité d'un mode de reproduction mixte pour générer de la variabilité. A l'échelle de l'arbre, un suivi des populations attaquant un seul arbre à cinq dates différentes a été effectué pendant une période d'un mois (juin-juillet). L'analyse de la diversité génétique basée sur des marqueurs RAPD (Random Amplified Polymorphism DNA) a mis en évidence une augmentation de la diversité au cours du temps avec l'apparition de nouveaux haplotypes par migration entre les arbres (Tenzer & Gessler 1997). Cette étude a aussi mis en évidence une diversité haplotypique plus élevée à l'échelle du verger qu'à l'échelle de l'arbre, ainsi qu'une faible clonalité. A l'échelle Européenne, une étude de la structure des populations a été conduite sur des échantillons de 11 vergers provenant de cinq pays (France, Allemagne, Suisse, Hollande & Italie) en utilisant des marqueurs RAPD, une séquence ribosomale et 7 marqueurs microsatellites (Tenzer *et al.* 1999). Les auteurs mettent en évidence une forte diversité génétique et une faible différenciation entre les localités. Un signal significatif d'isolement par la distance à travers l'Europe a aussi été détecté suggérant une dispersion préférentielle sur des courtes distances.

Plus récemment, les apports de la génétique des populations ont aussi permis l'élucidation de l'origine probable de *V. inaequalis* (Gladieux *et al.* 2010) et des grandes routes de dispersion empruntée par le champignon au cours de son processus de dispersion mondial (Gladieux *et al.* 2008). Cette approche a aussi permis une meilleure caractérisation de la diversité génétique présente dans les vergers conservatoires de pommiers (Leroy *et al.* en révision, voir annexe).

4.2.2.4.1 Dans le compartiment sauvage et cultivé

Une étude de génétique des populations réalisée par Gladieux *et al.* (2010) et portant sur un échantillonnage Eurasiatique de 948 souches de *V. inaequalis* isolées de pommiers cultivés (*M. x domestica*) et des pommiers endémiques sauvages Asiatiques et Européens (*M. sieversii* et *M. sylvestris* respectivement) et génotypées à l'aide de 10 marqueurs microsatellites. L'utilisation des méthodes d'assignations a mis en évidence trois populations : (1) une population Européenne regroupant des souches isolées à la fois sur le pommier sauvage endémique Européen, *M. sylvestris* et sur les pommiers domestiques en Europe; (2) une population Asiatique regroupant les souches isolées du pommier endémique Asiatique, présent en plaine dans la périphérie des zones anthropisés d'Asie centrale ; enfin (3) une population isolée, restreinte géographiquement aux montagnes de l'Est du Kazakhstan. Cette dernière population présente un nombre d'allèles uniques plus important que les deux autres populations.

Les probabilités postérieures du temps de divergence entre la population isolée des montagnes du Kazakhstan et les deux autres populations sous un modèle d'isolement avec migration (IMa) suggère une divergence de ces deux pools génétiques il y a 2000 à 4000 ans, ce qui coïncide assez bien avec la période de domestication du pommier estimée (2000-7000 ans ; Juniper & Mabberley). Les auteurs suggèrent donc un scénario de coévolution par host-tracking, c'est-à-dire une évolution de l'agent pathogène au cours de la domestication de son hôte. La validation de cette hypothèse écologique est néanmoins très dépendante de la

variance dans le nombre de répétitions des microsatellites de la population ancestrale. De plus, le temps de divergence calculé se base sur les valeurs médianes d'intervalles de confiance relativement large (environ 700 à 11 500 ans) qui eux-mêmes sont dépendants d'un fort taux de mutation (5.10^{-4}). Chez les rares espèces fongiques dont le taux de mutation est connu, il semble que cette valeur soit supérieure (e.g. *Neurospora Crassa* $2.8.10^{-6}$ à $2.5.10^{-5}$ par génération Dettman & Taylor 2004 ; *Saccharomyces cerevisiae* $7.2.10^{-7}$ à $7.2.10^{-5}$ par génération Lynch *et al.* 2008). La détermination du taux de mutation aux locus microsatellites de *V. inaequalis* et/ou la comparaison de scénarios d'émergence par les méthodes de type Approximate Bayesian Computation (ABC, Tavaré *et al.* 1997 ; Beaumont *et al.* 2002) avec la prise en compte d'hypothèses alternatives de contact secondaire des populations permettra de valider ou d'écarter ce scénario évolutif (voir p. ex. Roux, 2010 ; Roux *et al.* 2011).

De plus, en réalisant des inoculations croisées de ces populations sur les différents hôtes (*M. sieversii*, *M. sylvestris* et *M. x domestica*), Le Vân *et al.* (2012) ont démontré que les souches isolées de la population provenant des montagnes Kazakhes étaient incapables d'infecter les cultivars sensibles 'Gala' et 'Top Red Delicious' très déployés dans les vergers (n=8 souches). Les auteurs suggèrent que, sous ce modèle d'host-tracking, la domestication du pommier s'est accompagnée d'une acquisition de virulence chez le pathogène. De plus, selon cette même étude, la propagation de *V. inaequalis* dans les agrosystèmes se serait aussi accompagnée d'une augmentation de l'agressivité globale des populations de *V. inaequalis*.

4.2.2.4.2 Dans les vergers cultivés

Une étude plus large de Gladieux *et al.* (2008) réalisée sur un échantillonnage mondial de 1273 souches originaires des 5 continents a démontré que la richesse allélique diminue graduellement avec la distance entre le centre d'origine du pommier (Asie Centrale) et le lieu de collecte des souches. Ce résultat supporte l'hypothèse de forts effets de fondations dans chacune des zones échantillonnées. Plus globalement, l'Europe aurait servi de « tête de pont » pour la dispersion de la maladie à travers le monde suite à la découverte des nouveaux continents. Les auteurs suggèrent que *V. inaequalis* est arrivé à la dernière étape de son processus invasif, avec une distribution sur l'ensemble de l'aire de répartition de son hôte (aujourd'hui mondiale), des populations bien établies présentant une forte diversité génétique, une reproduction sexuée régulière et une dynamique d'expansion.

4.2.2.4.3 Dans les vergers conservatoires

Une autre étude de génétique des populations de pathogène portant sur un échantillonnage de souches de *V. inaequalis* infectant une grande diversité de *Malus* (5 espèces & 3 hybrides) présents localement dans une parcelle expérimentale de Dresde en Allemagne a été initiée (Leroy *et al.* en révision, voir annexe). Bien que ces hôtes soient connus pour porter une large diversité de gènes de résistance et de caractères quantitatifs de résistance, cette étude n'a détecté que deux populations spécialisés de *V. inaequalis* : la première infectant les *Malus* porteurs d'un gène de résistance appelé *Rvi6* (*M. x floribunda*, certains *M. baccata*) et la seconde infectant les pommiers non-porteurs de ce gène de résistance *Rvi6*. A l'intérieur de chacune des deux populations, un isolement par la distance significatif a été détecté suggérant que la distance de dispersion intrinsèque du pathogène est suffisamment faible pour pouvoir retracer les différents événements de migration et de mutation sur un verger d'une longueur inférieure à 150 mètres.

4.3 Résistance génétique du pommier à *V. inaequalis*

4.3.1 Résistance naturelle : large gamme de gènes de résistance

Chez le pommier, dix-sept gènes majeurs conférant une résistance à la tavelure ont été identifiés au sein du genre *Malus* (Bus *et al.* 2011). Une partie d'entre eux ont été identifiés chez *M. x domestica* (e.g. *Rvi1*, *Rvi4*, *Rvi10*, *Rvi15*), d'autres gènes tels que *Rvi5*, *Rvi8*, *Rvi11* et *Rvi12* ont été identifiés chez *M. micromalus*, *M. sieversii*, *M. baccata jackii*, *M. baccata* respectivement. Enfin les gènes *Rvi6* (=Vf) et *Rvi7* ont été identifiés chez le clone 821 de *M. x floribunda*. Les 17 gènes majeurs ont été cartographiés sur la carte génétique du pommier composé de 17 groupes de liaisons (LGs). Neuf des dix-sept gènes cartographient sur 2 groupes de liaisons (LG1 et LG2) formant des clusters de gènes de résistance. Les autres gènes sont répartis sur sept autres groupes de liaisons du pommier (LG3, LG4, LG6, LG8, LG10, LG12, LG17). Actuellement, sur les 17 gènes majeurs décrits, seuls 7 sont décrits comme impliqués dans des relations de type Gène-pour-Gène (*Rvi1*, *Rvi2*, *Rvi5*, *Rvi6*, *Rvi7*, *Rvi8* & *Rvi9*). En plus de ces résistances majeures à déterminisme monogénique, il existe chez le pommier des résistances partielles à déterminisme monogénique ou polygénique (Calenge *et al.* 2004; Gessler *et al.* 2006; Soufflet-Freslon *et al.* 2008). Cette résistance s'exprime au verger par un niveau de sensibilité plus faible qui s'explique par une plus faible croissance végétative du champignon (croissance radiale, sporulation) sur les feuilles et les fruits par rapport à un hôte ne présentant pas cette résistance quantitative.

Coté pathogène, la première observation de différences de pathogénéicité chez *V. inaequalis* sur des cultivars de *M. x domestica* a été effectuée par Aderhold en 1899 (Aderhold 1899). Dix-neuf gènes de virulence ont été identifiés ultérieurement (Boone & Keitt 1957; Williams & Shay 1957; Bagga & Boone 1968).

L'étude des populations de *V. inaequalis* a longtemps reposé sur l'identification de races physiologiques. Ces races ne sont pas directement décrites en fonction de la présence de ces 19 gènes de virulence, mais elles ont été définies en fonction de leur pouvoir pathogène sur une gamme de génotypes de *Malus* choisis en fonction de leurs résistances majeures. En effet, aucun locus d'avirulence n'a pour l'heure été cloné chez *V. inaequalis* et le mécanisme d'action de ces gènes reste totalement inconnu. Bien qu'une banque BAC contenant la région d'avr*Rvi1* (=avrVg) ait été construite (Broggini *et al.* 2007), ce locus n'a toujours pas été formellement identifié. L'utilisation récente d'une approche de transcriptomique et la recherche de polymorphisme non-synonymes dans de nombreux gènes a toutefois permis de diminuer la liste de gènes candidats pour être avr*Rvi1* (Broggini, communication personnelle). Les apports des données de séquence des génomes de *V. inaequalis* devrait aussi contribuer à la découverte d'autres candidats. En outre, quatorze gènes d'avirulence ont été cartographiés sur la carte génétique de *V. inaequalis* dont 5 des 7 gènes d'avirulence connus pour interagir avec des gènes de résistance du pommier (Broggini *et al.* 2011).

4.3.2 *Rvi6*, une résistance majeure utilisée dans les programmes de sélection du pommier

4.3.2.1 Origine de *M. x floribunda*

Le pommier sauvage *M. x floribunda* (et particulièrement le clone 821) a fortement été utilisé dans les programmes d'amélioration car porteurs de deux gènes de résistance majeurs : *Rvi6* et *Rvi7*. Bien que cet hôte soit omniprésent dans les programmes de sélection du pommier (Crosby *et al.* 1992), son origine est mal référencée. D'après Juniper & Mabberley (2006), *M. x floribunda* serait un hybride interspécifique de deux espèces de pommiers

sauvages utilisés en ornement pour l'abondance et le parfum de leurs fleurs. Selon ces auteurs, cette hybridation aurait eu lieu au Japon, d'où son nom français « pommier du Japon ». *M. x floribunda* trouverait son origine d'un croisement réalisé avant 1862 entre *M. toringo* et *M. baccata*, les pommiers sauvages endémiques du Japon et de la Sibérie respectivement. La mise en évidence récente de l'allèle type de *M. x floribunda* au locus *Rvi6* chez certaines accessions de *M. baccata* semble en effet suggérer que *M. baccata* est l'un des parents (Leroy *et al.* en révision ; Dunemann *et al.* 2012, voir annexe).

4.3.2.2 Historique de la résistance *Rvi6*

Au début du XX^{ème} siècle à l'université de l'Illinois, Crandall avait réalisé de nombreux croisements entre des accessions de *Malus* sauvages et des cultivars de pommier dont un croisement *M. x floribunda* (clone 821) par 'Rome Beauty'. Ce dernier croisement fût, comme les autres, rapidement tombé dans l'oubli. En 1943, les conditions météorologiques du printemps anormalement humide et doux pour l'Illinois, ont induit une épidémie sévère de tavelure provoquant de graves défoliations des cultivars sensibles, Fred Hough remarque alors que la F2 de ce croisement ségrégeait selon un ratio 1 résistant : 1 sensible (*i.e.* résultant probablement d'un croisement *Rvi6rvi6* x *rvi6rvi6*) (Hough 1944; Janick 2006). Plus tard, les descendants hétérozygotes pour la résistance ont mis en évidence un ratio de type 3 résistants : 1 sensible (Hough *et al.* 1953), ratio attendu avec tout gène dominant. Cependant, le gène *Rvi6* ne semble pas être dans le sens strict du terme, un gène dominant. En effet, les plantes homozygotes *Rvi6* sont généralement plus résistantes que les hétérozygotes (Belfanti *et al.* 2004). Inoculés avec *V. inaequalis*, les pommiers porteurs de *Rvi6* montrent des réactions de résistance variables, se situant entre l'absence de symptômes et la présence de lésions chlorotiques ou nécrotiques sporulantes. Hough *et al.* (1988) propose que la variation dans les réactions de résistance des cultivars *Rvi6* soit due à la présence de gènes « modulateurs » de l'expression de *Rvi6* (Gessler *et al.* 2006).

Pour élucider le mode d'action et la fonction de *Rvi6*, un clonage positionnel a été entrepris et a permis l'identification d'un cluster de plusieurs gènes *Rvi6* candidats codant pour des protéines de type récepteur (Vinatzer *et al.* 1998; Patocchi *et al.* 1999a; Patocchi *et al.* 1999b). Les séquences des protéines en acides aminés de ces gènes montrent une bonne similitude avec les gènes de résistance à *Cladosporium fulvum* présents chez la tomate (gènes *Cf*) (Vinatzer *et al.* 2001). Ces gènes sont appelés *HcrVf* et sont caractérisés par un domaine extracellulaire de type LRR (Leucine Rich Repeat) et un domaine transmembranaire putatif similaire à celui des gènes *Cf* (Jones *et al.* 1994). Le cluster de gènes de résistance *Rvi6* obtenu par clonage positionnel contient au moins quatre gènes candidats de type *HcrVf* dont trois ont pu être transcrits (Xu & Korban 2002). La transformation du cultivar sensible 'Gala' par l'introgession du seul gène *HcrVf2* a mis en évidence que ce gène induisait une résistance à la tavelure dans les plants transformés (Belfanti *et al.* 2004).

La découverte de la résistance *Rvi6* par Hough (1944) a permis à Hough et Shay d'initier un vaste programme d'amélioration de la résistance à la tavelure dès 1945. Ce programme a permis d'obtenir jusqu'à ce jour, plus de 80 variétés commercialisées, faisant de *Rvi6* la résistance la plus introgressée dans les cultivars de pommier (Crosby *et al.* 1992; Gessler & Pertot 2012). L'absence de contournement de la résistance dans les vergers expérimentaux pendant plus de 40 ans explique la prépondérance de l'utilisation de la résistance *Rvi6*, longtemps considérée comme durable (Crosby *et al.* 1992). Bien que le pseudo-rétrocroisement¹¹ avec de nombreuses variétés ait permis d'améliorer sensiblement les qualités gustatives des pommes de cultivars *Rvi6*, ces variétés ne représentent qu'une part

¹¹ Le pommier étant auto-incompatible, l'élimination du fond génétique sauvage a été réalisée par une approche de pseudo-rétrocroisements (voir Gessler & Pertot 2012 pour plus de détails)

relativement marginale du marché mondial de la pomme qui est essentiellement composé de 12 cultivars, tous sensibles à la tavelure. Sans le recours à de grandes campagnes publicitaires, l'acceptation de nouvelles variétés se fait difficilement auprès des consommateurs, des grandes centrales d'achat et des producteurs. Le cultivar 'Cripps Pink', commercialisé sous le nom de Pink LadyTM, est un bon exemple d'un succès d'une nouvelle variété rapidement adopté par les consommateurs à la fin des années 1990 ; cependant, ce cultivar est lui aussi un des plus sensible à la tavelure. En 2007, les cultivars de pommiers résistants (essentiellement des cultivars *Rvi6*) ne représentaient qu'une très faible proportion de la surface de production des 27 états membres de l'Union Européenne (0.5% dont 0.24% Topaz et 0.1% pour Prima), chiffre qui est à opposer avec les surfaces des trois principaux cultivars sensibles 'Golden Delicious' (13%), 'Jonagold' (6.6%) ou 'Gala' (6%) (EUROSTAT, 2007).

4.3.2.3 Contournement de la résistance *Rvi6*

Les premières observations sporadiques d'infection des cultivars *Rvi6* ont été relevés à la fois sur des cultivars de *M. x domestica* et sur des individus de *M. x floribunda* en 1988 en Allemagne (Parisi *et al.* 1993), en 1989 en Angleterre (Roberts & Crute 1994). La population observée en Angleterre présentait déjà les deux signes sexuels de *V. inaequalis* (Tony Roberts, communication personnelle). En France, la première mise en évidence de souches virulentes envers le gène *Rvi6* a été rapporté en 1995 dans un verger de Normandie sur le cultivar de pommier à cidre 'Judeline', seulement six ans après sa plantation (Parisi *et al.* 2000), puis dans d'autres pays Européens (Hollande, Suisse, Belgique, Danemark). Durant la saison 2000, de nombreux foyers de tavelure de grande intensité ont été observés en Normandie sur le cultivar 'Judeline', correspondant au premier cas de contournement à grande échelle de la résistance *Rvi6*. Aujourd'hui, la résistance est largement contournée en Europe (p. ex. France, Danemark, Suède, Pologne) (Guérin et le Cam, 2004; Le Cam, Monika Michalecka, communications personnelles) mais son efficacité peut être partiellement préservée par un apport raisonné en fongicides et la mise en place de mélanges variétaux (*Rvi6* et non-*Rvi6* aux vergers) (Didelot *et al.* 2007). Dans l'hémisphère sud, une observation ponctuelle de symptômes de tavelure sur une variété possédant *Rvi6* a été rapportée (Bus *et al.* 2000), mais aucun événement de contournement à grande échelle. La surface occupée par les cultivars *Rvi6* dans l'hémisphère sud est encore plus faible que dans l'hémisphère nord.

Les analyses de génétique des populations effectuées dans les vergers cidricoles montrent que les populations virulentes capable d'infecter la variété Judeline *Rvi6* (*virRvi6*) partagent une origine commune et possède une diversité réduite comparé aux populations présentes sur les variétés non-*Rvi6* (*avrRvi6*) (Guérin et Le Cam, 2004 ; Guérin *et al.* 2007). L'existence d'une forte différenciation entre les populations *virRvi6* et *avrRvi6* (Guérin *et al.* 2007) et son apparent maintien au cours du temps (Gladieux *et al.* 2011) suggérait un flux de gènes relativement restreint entre les deux populations, malgré le fait que les cultivars *Rvi6* soient déployés dans des même vergers que les cultivars ne possédant pas cette source de résistance.

Les populations virulentes présentaient deux allèles majoritaires pour le marqueur microsatellite *Itc1g*, qui est décrit comme lié au locus d'*avrRvi6*. L'absence de ces deux allèles dans les populations *avrRvi6* suggérait que la sous-population virulente n'a pas émergé à partir des populations *avrRvi6* du verger étudié, mais résulterait plutôt d'une migration de souches à partir d'une source extérieure. Cette hypothèse étant renforcée par le fait que la population virulente n'est pas pathogène sur les variétés de pommier à cidre non *Rvi6* de ces vergers (Gladieux *et al.* 2011). Deux hypothèses différentes peuvent donc être proposées, soit les populations *virRvi6* sont apparues par mutation dans un autre agrosystème et ont ensuite migré, soit les populations étaient préexistantes sur un hôte alternatif à l'extérieur des

agrosystèmes.

4.3.2.3.1 Modèle de la « mutation vers la virulence »

Les rapides changements évolutifs aux locus d'avirulence sont attribués à une perte de fonction de l'allèle d'avirulence à l'intérieur des populations des agrosystèmes (encadré II). Ces changements peuvent prendre différentes formes (Figure 12): des mutations ponctuelles non-synonymes, des insertions d'éléments transposables ou des délétions partielles ou complètes du gène d'avirulence (Joosten *et al.* 1997; Orbach *et al.* 2000; Catanzariti *et al.* 2006; Gout *et al.* 2007; Stergiopoulos *et al.* 2007; Fudal *et al.* 2009). Les délétions peuvent aussi affecter les régions flanquantes des gènes, comme dans le cas de la délétion d'une région de 12 kb contenant le gène *SIX1* chez *Fusarium oxysporum* (Rep *et al.* 2004) ou d'une région de 260kb contenant *AvrLm1* chez *Leptosphaeria maculans* (Gout *et al.* 2007).

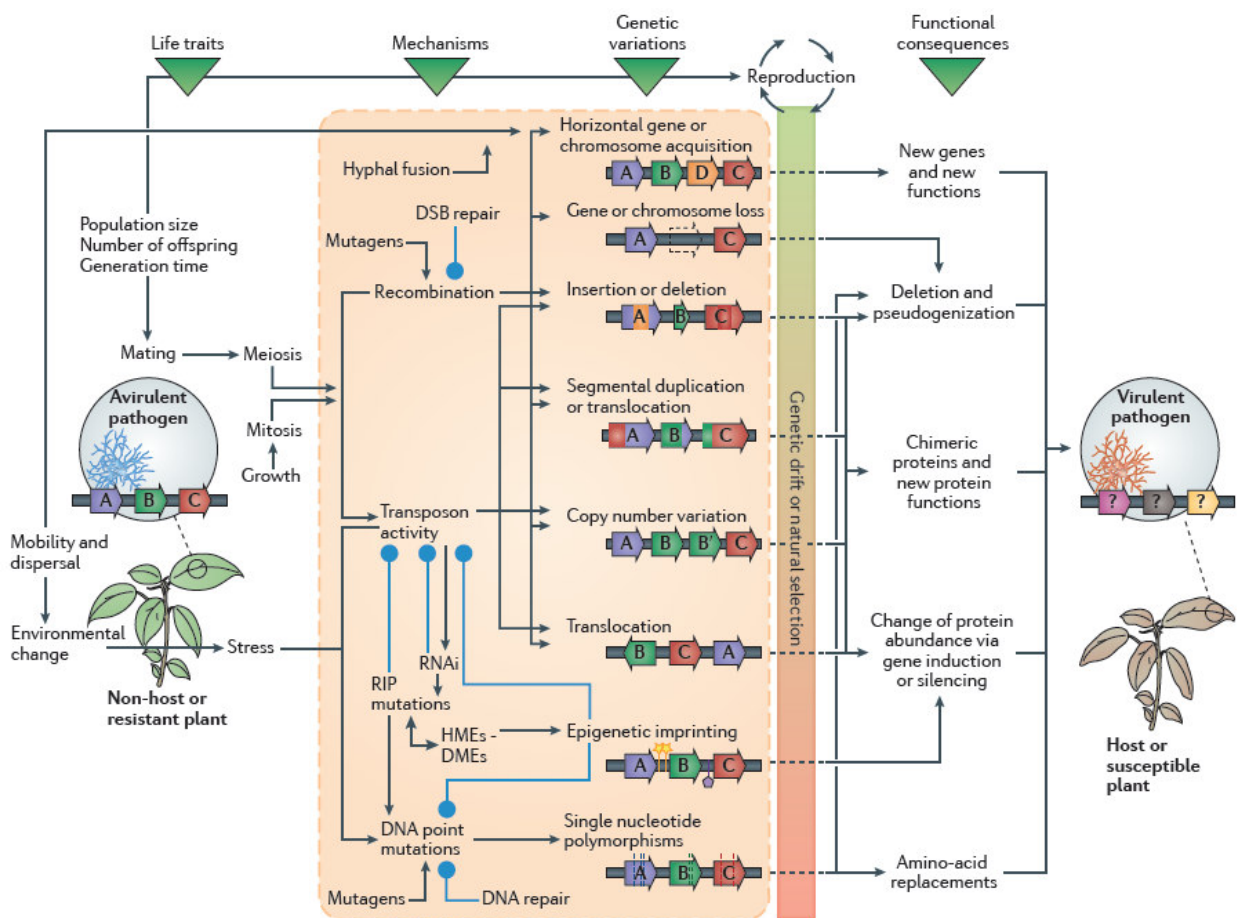


Figure 12 : scénarios évolutifs conduisant à la virulence chez les phytopathogènes filamenteux (d'après Raffaele *et al.* 2012). Les relations positives entre les mécanismes sont figurées en noir, les relations inhibitrices en bleu. Les acronymes correspondent à : DME, enzymes de méthylation de l'ADN ; DSB, réparation des dommages de l'ADN double brin ; HME, enzymes de méthylation des histones, RNAi, ARN interférence.

Les génomes des champignons phytopathogènes sont de grande taille et sont généralement décrits comme particulièrement plastiques (Ma *et al.* 2010; Raffaele & Kamoun 2012). Cette plasticité serait notamment conférée par la forte mobilité des éléments transposables, l'existence d'un système de contrôle de l'invasion des transposons induisant un très grand nombre de mutations ponctuelles (i.e. Repeat Insertion Mutation points, RIP), la présence de chromosomes dispensables (Figure 12). Pour l'ensemble de ces changements

permettant de générer des allèles de virulence, j'utiliserai par la suite l'expression de « mutation vers la virulence ». La mutation vers la virulence semble expliquer la plupart des cas d'émergence des populations virulentes. Sous ce modèle, Van der Plank (1963) a suggéré que la mutation vers la virulence est intrinsèquement favorable sur les hôtes résistants mais que celle-ci peut s'accompagner d'une perte de valeur sélective (coût de fitness) sur les hôtes sensibles (Parker & Gilbert, 2004). Ce coût de fitness est, en effet, nécessaire pour expliquer comment l'allèle fonctionnel avirulent peut être maintenu au cours du temps chez les pathogènes exposés à des hôtes porteurs du gène de résistance. De nombreuses exemples de coût de fitness ont été rapportés (Leach *et al.* 2001; Thrall & Burdon 2003; Huang *et al.* 2006; Montarry *et al.* 2010) même si certaines populations virulentes ne semblent parfois pas affectées (Leach *et al.* 2001; Peressotti *et al.* 2010). La présence ou non de coût de fitness n'aboutit pas du tout aux mêmes stratégies de déploiement des gènes de résistance (Sapoukhina *et al.* 2009).

Le modèle de mutation vers la virulence à partir de populations *avrRvi6* à l'intérieur des agrosystèmes a été proposé pour expliquer l'émergence des populations *virRvi6* au champ (Gladieux *et al.* 2011). Conformément au modèle de Van der Plank (1963), l'agressivité de 8 souches *virRvi6* est plus significativement plus faible que celui de 8 souches *avrRvi6* sur le cultivar non-*Rvi6* 'Gala' (Caffier *et al.* 2010). Une première hypothèse serait que la mutation vers la virulence s'accompagnerait d'un coût de fitness sur les cultivars non-*Rvi6*. Toutefois, les auteurs ne peuvent rejeter l'hypothèse alternative d'une meilleure adaptation des souches *avrRvi6* à 'Gala'. Même en considérant que ce coût de fitness est intrinsèque à l'allèle de virulence, les coûts estimés ne sont pas compatibles avec une exclusion totale des deux populations et le maintien de la différenciation observé par le suivi pluriannuel de vergers à cidre (Gladieux *et al.* 2011). Gladieux *et al.* (2011) expliquent la double exclusion observée dans les vergers à cidre par l'incapacité de la population *virRvi6* à infecter les variétés à cidre non-*Rvi6* et par conséquent à l'impossibilité pour les deux populations à se croiser. Ainsi, Gladieux *et al.* (2011) considèrent que l'adaptation au locus d'*avrRvi6* ne contribue probablement pas seule au maintien de la différenciation mais que plusieurs barrières prézygotiques pourraient être en déséquilibre de liaison complet avec le gène d'*avrRvi6* et contribuer à l'exclusion des deux populations selon un modèle d'un isolement reproductif *via* un 'caractère magique'. Selon les auteurs, la pléiotropie entre l'adaptation à l'hôte et l'isolement reproductif pourrait contribuer à ce déséquilibre de liaison apparent (Giraud *et al.* 2006; Gladieux *et al.* 2011).

4.3.2.3.2 Histoire évolutive des populations *virRvi6*

Le second scénario d'émergence possible est que les souches *virRvi6* étaient préexistantes sur un hôte alternatif sauvage et que les vergers ont été infectées suite à la migration des souches virulentes du compartiment non-cultivé vers les agrosystèmes (encadré IV). Une récente étude de génétique des populations menée sur des populations *avrRvi6* et *virRvi6* collectées dans différents vergers Européens et des accessions isolées de *M. x floribunda* (Lemaire *et al.* soumis, encadré IV) suggère que ce second scénario d'émergence est le plus vraisemblable. Les auteurs démontrent que les populations *avrRvi6* et *virRvi6* ont divergé en allopatrie plusieurs milliers d'années plus tôt que le déploiement de la résistance dans les agrosystèmes réfutant l'hypothèse d'une mutation récente vers la virulence (encadré II). Suite au déploiement des cultivars *Rvi6* dans les vergers, les populations *virRvi6* sur pommiers domestiques ont été fondées à partir de plusieurs événements de migration à partir de *M. x floribunda* ou d'autres *Malus* porteurs de *Rvi6*. Notons toutefois que la virulence par rapport à *Rvi6* semble absente des populations infectant les deux principaux progéniteurs du pommiers cultivé (*M. sieversii* et *M. sylvestris*) (Lê Van *et al.* 2011). Ces

migrations de souches du compartiment non-cultivé vers le compartiment domestique ont pu être favorisées de deux manières. Premièrement, une forte proximité entre des jardins ornementaux contenant des *M. x floribunda* ou d'autres *Malus* porteurs de *Rvi6* et les vergers de cultivars résistants pourrait avoir contribué à un contact secondaire entre les deux populations. Alternativement, les vergers expérimentaux utilisés pour les programmes de sélection ont aussi pu contribuer à cette remise en contact. En effet, dans ces vergers, les parents et les descendants sont généralement plantés en sympatrie dans le but d'évaluer les niveaux de résistance et ainsi de limiter la variabilité environnementale.

Ce scénario semble corrélér à la plupart des connaissances actuelles. En effet, notons que la population vir*Rvi6* observée sur *M. x floribunda* en 1989 était présente sur une île avant même le déploiement des variétés *Rvi6* et possédait déjà les deux signes sexuels (Roberts, communication personnelle). D'abord ce modèle permet d'expliquer l'observation concomitante des premières souches sur les cultivars *Rvi6* de *M. x domestica* et de *M. x floribunda* (Parisi *et al.* 1993) dans différentes régions d'Europe, (Allemagne, Espagne, Angleterre). De plus, il suggère que l'utilisation de *M. x floribunda* en ornement ou dans les vergers expérimentaux de sélection aurait pu faciliter la migration de souches de *M. x floribunda* vers le pommier *Rvi6*.

La description de cette émergence des populations vir*Rvi6* est une chance car elle permet à la fois de formuler des attendus de barrières exogènes mais aussi des barrières endogènes accumulées pendant la divergence et de comprendre la part de l'isolement reproductif conféré par ces deux types de barrières en situation de contact secondaire dans des agrosystèmes mixtes de pommiers (cultivars non-*Rvi6* et *Rvi6*) de pommiers.

4.4 Questions de recherche

L'objectif de cette thèse est de caractériser les conséquences de ce contact secondaire entre les souches avr*Rvi6* et vir*Rvi6*. Les résultats principaux de mon travail de thèse se retrouvent sous la forme de deux chapitres. Le premier chapitre sera consacré à la mise en évidence d'une zone ponctuelle d'hybridation et à la caractérisation des barrières génétiques d'isolement reproductif entre les populations divergentes vir*Rvi6* et avr*Rvi6*. La seconde partie s'intéressera aux conséquences de ce contact secondaire pour l'évolution de la virulence et de l'agressivité des populations de pathogènes et les conséquences épidémiologiques pour les hôtes.

La thèse vise à répondre aux trois questions de recherche ci-dessous :

Existe-t-il des zones de contact avec flux génique entre populations avirulentes et virulentes vis à vis de *Rvi6*?

Les barrières génétiques post-zygotiques endogènes contribuent-elles aussi à l'isolement reproductif des populations de pathogènes ? Quels types de barrières génétiques sont mis en évidence dans ces zones de contact ?

Comment sont affectés les hybrides : dépression hybride ou recombinaison contribuant à l'établissement de nouvelles combinaisons génétiques à forte valeur sélective ?

RESULTATS
CHAPITRE 1

1. Article 1

Finding ecological divergence in pathogens: a needle in a haystack?

Thibault Leroy, Valérie Caffier, Jean-Marc Celton, Pierre Gladieux,
Charles-Eric Durel, Christophe Lemaire* et Bruno Le Cam*

Although speciation is one of the main topics in biology, the contribution of exogenous selection (ecological selection) to the overall reproductive isolation is a still open debate. In pathogens, authors generally assume a predominance of adaptation to host over gene flow and neglected the alternative hypothesis of an accumulation of Dobzhansky-Muller Incompatibilities (DMI). Using a 969 AFLP loci dataset, we tested the alternative hypothesis of an accumulation of endogenous barriers between two specialized populations of a fungal pathogen infecting apples. This plant-pathogen interaction was previously described as a model to attend strong ecological barriers with a main contribution of a single gene acting as a magic trait. Contrary to this expectation, a secondary contact zone with gene flow between the two populations was detected indicating that exogenous barriers were not strong enough to isolate by itself the two populations. Second, comparing the two populations by genome scans, several non-neutral outlying loci were detected. Combining fine characterization of genetic hitchhiking in surrounding regions of the outliers with their genetic mapping, our study revealed for the first time occurrence of endogeneous barriers in pathogens. Subsequently, we discussed the contribution of endogenous (i.e. non-ecological) barriers to the overall reproductive isolation of pathogens populations. We argue that in pathogens, as well as for other organisms, determining the genetic basis of diversification is much more complex than commonly thought.

Reproductive isolation in pathogens is often viewed as a by-product of ecologically based divergent selection (1) focusing solely on ability to infect a given hosts. Ecological divergence is then expected to be driven by few exogenous genetic barriers corresponding to loci involved in adaptation to host. Even if this adaptive model is now popular, it still needs to be validated for a wider range of pathogens. Furthermore, the “null hypothesis” of speciation (2) has been neglected in pathogens. Under this model, the endogenous genetic barriers due

to Dobzhansky-Muller incompatibilities (DMI) cumulate in allopatry. Besides, recent theoretical developments suggest that these endogenous barriers (independent of environment) might initiate speciation as fast as the exogenous ones (3). In the light of these new developments, the contribution of ecology to the divergence of pathogen populations was probably overestimated and needs to be reevaluated (3, 4).

Hybridization zones occurring after secondary contacts offer a strong opportunity to characterize the nature of genetic barriers to gene flow (5). As a matter of fact, microorganisms, and particularly pathogens, have been quite neglected of the hybrid zone framework. Consequently most of the studies only focus on ecological barriers instead of scanning for all loci involved in reproductive isolation (1, 6). However, as microbial diversification negatively impacts health and agriculture, the understanding of these causes is a fundamental issue in the study of the dynamics of new emergences. Furthermore, pathogens are intrinsically good models for speciation with numerous events of recent diverging populations, and secondary contacts favored by human-mediated host translocation for agriculture and trading, which allow investigations on the early stages of speciation (7).

Genetics barriers are usually assessed by genome scan methods through the identification of loci deviating from neutral assumptions (so-called outliers) (8). Loci detected by genome scans, comparing ecologically divergent populations of pathogens, are often interpreted as markers of the exogenous barriers encoding for adaptation to host. When the selection is strong and the pathogen specialization is recent, a large footprint of local hitchhiking might be expected on the surrounding region of the targeted gene (9, 10). Consequently, a handful of genetic markers might be sufficient to detect these so-called genomic “islands” of divergence (11, 12). However, even when corrected for demographic fluctuations, outlier detection methods suffer from a high false positive rate [e.g. (13)] Moreover, ecological interpretation of outliers might be confusing because they may also map the endogenous barriers (4). Even if the endogenous barriers are expected to generate smaller footprints of ongoing local selection than the exogenous barriers (11), the formers are expected to be quickly accumulated during divergence and therefore may largely outnumber the exogenous barriers (2, 14). Here, using an empirical example in pathogen, we demonstrated that non-ecological barriers explained a large part of the reproductive isolation between two recently diverging populations of *Venturia inaequalis*.

Apple scab as a model for assessing genetic barriers

Although *V. inaequalis* is the foremost worldwide pathogen on apples and a model for ecological speciation in pathogens (1), the nature of causative barriers to gene flow between divergent populations has never been sought. A striking example is the maintenance of genetic differences between two *V. inaequalis* populations in sympatry - the “orchard native population” and an emerging population enables to overcome the *Rvi6* resistance gene present in new varieties. Overcoming of this resistance gene introgressed by breeders from the wild progenitor *M. x floribunda* was first attributed to an adaptive

mutation at a single exogenous barrier, the avirulence gene *-avrRvi6-* (15-17) as exemplified in numerous host pathogens interactions where infected susceptible varieties in the fields play the key role of virulence source. Because *V. inaequalis* mates within its host after successful infection and because pathogenicity-related loci prevent infection of non-host trees, we argued that adaptation to specific hosts may alone maintain both genetic differentiation between and adaptive allelic combinations within sympatric populations parasitizing different apple varieties, thus acting as a 'magic trait' (15). However, the recent assessment of the evolutionary history of the emerging population at the European scale rather suggests that the emerging population would actually find its origin on the wild where it preexisted on *Malus x floribunda*. After a divergence process which occurred several thousand years ago in allopatry, this virulent population recently invaded European orchards given that the new *Rvi6* varieties acted as gateways for its emergence in agroecosystems (Lemaire *et al.* in prep). Following this secondary contact, two scenarios were then possible, either there was gene flow between the two differentiated populations, either they remained differentiated.

Secondary contact zone and asymmetric gene introgression

Using a 969 AFLP loci dataset, native and emerging pathogen populations were compared in two different orchards, free from chemical sprays, which both contain *Rvi6* and non-*Rvi6* apple cultivars. In the first orchard, the two populations seemed poorly permeable to gene flow as already reported in numerous orchards (15-17) whereas in the second orchard, a secondary contact zone with gene flow were detected. Indeed, factorial correspondence analysis (CA) revealed three well-separated clusters. The first axis separated the two populations from site 1; but not the two populations from site 2 which formed a single independent group (A, Fig 1). In addition, based on a 25 'diagnostic' loci subset, individuals sampled on non-*Rvi6* hosts and *Rvi6* cultivars sampled in the site 1 expressed a non-overlapping range of hybrid indexes. Individuals from the native apple population of the site 1 had a lower hybrid index (from 0 to 0.17; mean = 0.06 ± 0.05) than individuals from the emerging population (from 0.42 to 1; mean = 0.92 ± 0.15) (B, Fig 1). At the opposite, populations from the site 2 had an overlapping range of hybrid indexes (for non-*Rvi6* hosts 0.43 ± 0.11 and for *Rvi6* hosts 0.55 ± 0.10) which revealed that both genomes are hybrids resulting from a contact zone with gene flow between native and emerging populations (C, Fig 1). In addition, mean between-population genetic differentiation over the 969 loci for population from site 1 is higher ($\theta=0.271^{***}$) than in the site 2 ($\theta=0.046^{***}$) strengthened the role of the recombination that mixed the two populations within the second site.

Conversely, populations from the site 2 exhibits an overlapping range of hybrid indexes (for non-*Rvi6* hosts 0.43 ± 0.11 and for *Rvi6* hosts 0.55 ± 0.10) which revealed that both genomes are admixed as a result of the secondary contact with gene flow between native and *M. x floribunda*-like populations (C, Fig 1). In addition, mean between-population genetic differentiation over the 969 loci for population from site 1 ($\theta=0.271^{***}$) and for site 2 ($\theta=0.046^{***}$) highlights the role of the recombination that mixed the two populations

within the second site.

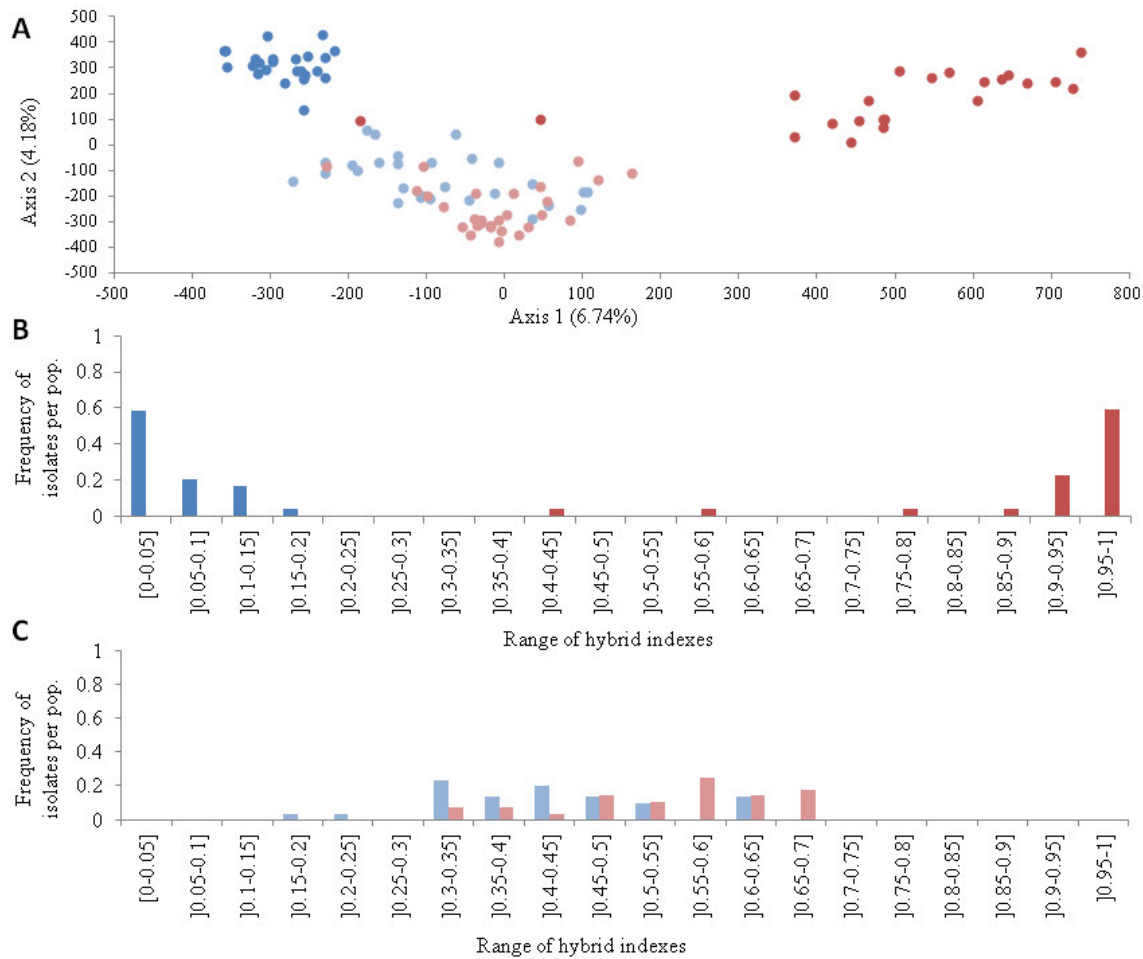


Fig. 1: Permeability to gene flow between populations infecting non-*Rvi6* and *Rvi6* hosts in sympatry. In the first site, genomes of *V. inaequalis* populations infecting non-*Rvi6* populations (blue) and population infecting *Rvi6* hosts (red) were almost impermeable to gene flow while in the second site, populations infecting non-*Rvi6* hosts (light blue) and *Rvi6* hosts (pink) were highly permeable to gene flow (A) Output of the 2-axis of the factorial correspondence analysis computed over the 969 loci dataset revealing three well-separated clusters. The first axis (6.74% of the total inertia) separated the two populations from site 1; but not the two populations from site 2 which formed a single independent group. Note the highest proximity between the population infecting non-*Rvi6* hosts from the site 1 and populations from the site 2. (B) Classes of hybrid index for the two populations from the first site based on a 25 diagnostic loci (see supplementary Material and Methods). Individual from the native apple orchards population of the site 1 had a lower hybrid index (from 0 to 0.17; mean = 0.06 ± 0.05) than individuals from *M. x floribunda*-like population (from 0.42 to 1; mean = 0.92 ± 0.15). (C) Classes of hybrid index for the two populations from the second site. Overlapping range of hybrid indexes (for non-*Rvi6* hosts 0.43 ± 0.11 and for *Rvi6* hosts 0.55 ± 0.10) revealed a secondary contact zone with gene flow between these two populations.

We showed that the *avrRvi6* barrier is not sufficient enough to isolate by itself the two populations. As a matter of fact, *avrRvi6* was not acting as a magic trait as previously suggested (15). In fact, strains sampled in the site 2 were phenotyped for the virulent (*virRvi6*) or avirulent trait (*avrRvi6*) towards the resistance gene *Rvi6*. To date, strain origins (a *Rvi6* or a non-*Rvi6* host) were fully correlated with the phenotype avirulence/virulence at the *Rvi6* locus assessed through pathogenicity tests (15-17). For the stains sampled in the second

site, even if all strains sampled on *Rvi6* expressed a *virRvi6* phenotype, strains sampled on non-*Rvi6* hosts not exclusively expressed a *avrRvi6* phenotype and a high proportion of strains (47%) are able to produce symptoms on *Rvi6* hosts in our greenhouse inoculation test. With this observation in mind and given the particular lifecycle of the pathogen (*i.e.* sex is only possible between strains infecting the same leaf); recombination between the two populations was only possible if members of both populations successfully infect a same leaf of non-*Rvi6* host and mate together. Consequently, in this orchard, *M. x floribunda*-like population can be regarded as a recent invader of the non-*Rvi6* hosts. In this context, gene flow is expected to be strongly asymmetric with a massive introgression of the native population into the emerging one (18). Our empirical report follows this expectation. Indeed, mean value of hybrid indexes based on the phenotype (native orchard population 0.37 ± 0.09 ; *M. x floribunda*-like 0.53 ± 0.11) is biased toward a greater introgression from the native to the invasive population (47% of *avrRvi6* alleles) rather than from the invasive to the native one (36% of *virRvi6* alleles, supplementary data Fig 1).

Contribution of endogenous barriers to speciation

Genome scans were performed on the hybrid population from the site 2 genotyped at 969 AFLP loci. First, such a scan was performed by comparing two populations given the habitats (*Rvi6* or non-*Rvi6* hosts). Twenty-four outliers exhibited abnormally high levels of F_{ST} (A, Fig 2; with θ from 0.129 to 0.399, mean θ over outliers: 0.274^{***}). Additionally, a second genome scan, given the phenotypes of the individuals inferred from pathogenicity tests (avirulent vs. virulent toward the *Rvi6* gene), was performed to detect markers in strong linkage disequilibrium with the expected exogenous barrier (*avrRvi6* or *virRvi6* allele; B, Fig 2). This second genome scan revealed 24 outliers (θ from 0.231 to 0.637; mean θ over outliers: 0.461^{***}), including eight loci already detected by the first genome scan (Fig 2, red circles). Given the relatively high number of outliers detected (4.1% of the 969 loci detected); the simple hypothesis that outliers only mapped a single exogenous barriers acting as a 'magic trait' seems unlikely.

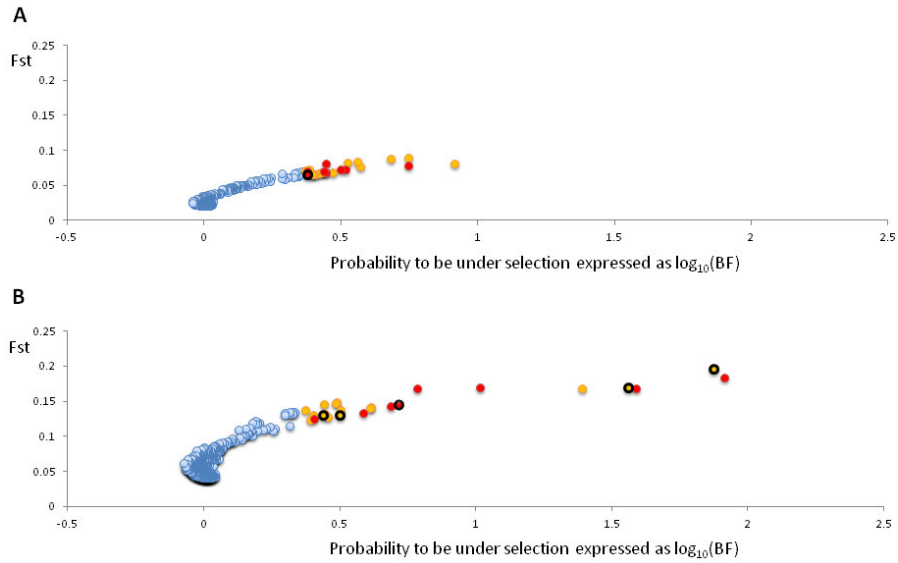


Fig. 2: Genome scans for the detection of AFLP outlier loci potentially subject to selection. (A) Outliers revealed by a scan comparing populations based on the habitats (non-Rvi6 vs. Rvi6 habitats). (B) Scan comparing populations given the phenotypes inferred from pathogenicity tests (*avrRvi6* vs. *virRvi6*). A cutoff of $p=0.7$, close to the “substantial” threshold included on the BAYESCAN interface, was used to isolate outliers from the dataset as previously proposed (e.g. Buckley et al., 2012). Neutral loci (light blue) express lower F_{ST} values than the 32 loci detected by one scan (orange) or the 8 loci detected by the two independent scans (red). Outliers used for the genetic map (see Fig 3) are surrounded by a black outline.

To maximize our power to detect the exogenous barriers, we subsequently studied outliers in strong linkage disequilibrium with the phenotype (B, Fig 2). Interestingly, even if the strategy is biased toward an easier detection of the exogenous barriers than the endogenous ones, several arguments supported that several outliers only mapped postzygotic endogenous barriers.

First, F_{ST} values were assessed for several SNPs distributed in the surrounding region of four outliers (A, Fig. 3). For each of the four investigated regions, SNP polymorphisms exhibited high F_{ST} values at least for those closest to the AFLP outlier loci. F_{ST} values are consistent with previously detected outlier statuses thus validating that these four outliers are true positives. Furthermore, the genetic hitchhiking effect did not extend far beyond tens of kilobases on both sides of the outlier suggesting that many recombination events occurred, thus partially disrupted the allelic association between surrounding neutral loci and the gene that is positively selected. Several authors suggested that the exogenous barriers can generate larger “islands” of differentiation than the endogenous barriers (11). If such a hypothesis is true, sizes of genomic regions affected by genetic hitchhiking is more consistent with the endogenous barriers than the exogenous barriers.

Second, a genetic mapping strategy was performed by genotyping the 24 outliers on a F1 progeny of a cross between an *avrRvi6* and a *virRvi6* strain. Although the two parental strains were poorly polymorphic at these outlying loci, 5 outliers were successfully mapped. Outliers were mapped on several linkage groups (LG) but none of these outliers were located near the *avrRvi6* gene assessed by the phenotype (B, Fig 3). Furthermore, two outliers

significantly deviate from expected Mendelian segregation ratio of 1:1. Given that this crossing experiment was conducted *in vitro*, ecological selection cannot be responsible for such a non-mendelian segregation of markers and therefore active selection against hybrids (*i.e.* endogenous barriers) is the only selection pressures able to lead to these distortions. As a consequence, an *in vitro* crossing experiment can be helpful to dissociate endogenous and exogenous barriers to gene flow. Here, given that *V. inaequalis* is a haploid organism; segregation distortion can be detected if, and only if, selection induces highly contrasted fitnesses in recombinants [high mortality or sterility for one recombinant (*e.g.* $W_{Ab} \approx 0$) but not the other ($W_{aB} \approx 1$), for more details see supplementary data Fig 2]. Despite this limitation, these two outliers are the signature of a strong and active selection against hybrids (B, Fig 3).

Third, several SNPs around the outliers were further located into the genetic map and mapped together on the same linkage group (C, Fig 3). Intuitively, in this situation, it makes sense to support a hypothesis of a single exogenous barrier. However, such a pattern can also be caused by strong selection against hybrids at several independent endogenous barriers. Indeed, deleterious epistatic interactions are expected to maintain parental allelic combinations at these independent speciation genes, thus leading to strong linkage disequilibrium between them even if outliers map on different chromosomes, therefore those outliers are expected to co-locate on the same single group. Here, given that SNPs are physically located on several scaffolds throughout the fungal genome, the hypothesis of a single exogenous barrier can be rejected in favor of the alternative hypothesis of various endogenous barriers. However, given that genome scans are still predominantly performed on anonymous markers such as AFLP, the exclusion of this alternative hypothesis is challenging.

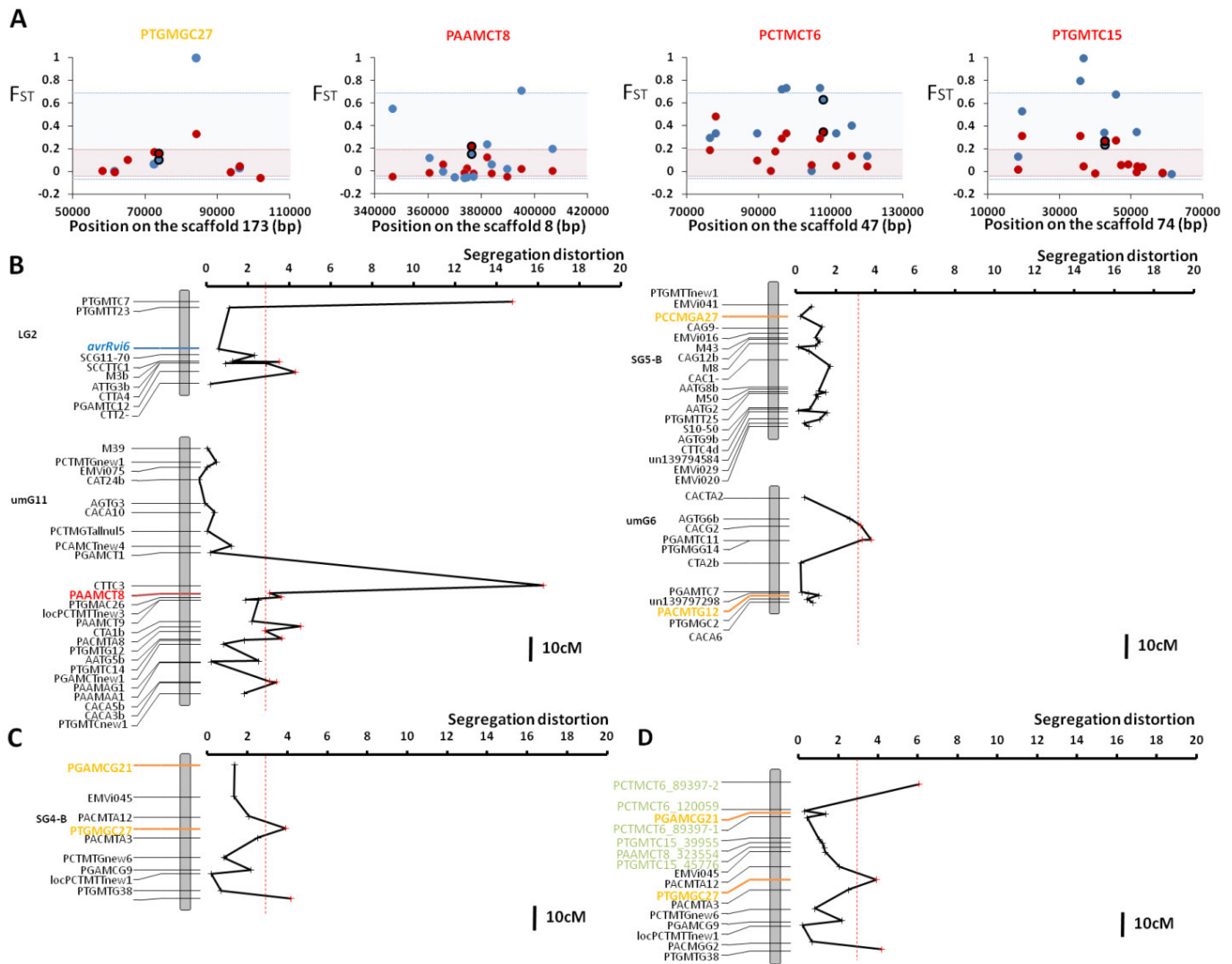


Fig 3: Independent evidences for the endogenous nature of the reproductive barriers. (A) Fine characterizations of F_{ST} in the surrounding region of several outliers. Patterns revealed very restricted footprints of the genetic hitchhiking over the genome. For each pattern, several SNPs express higher F_{ST} values comparing populations infecting non-Rvi6 and Rvi6 hosts in the site 1 (blue points) and in the site 2 (red) than expected with for 95% of neutral loci from the site 1 (light blue) and the site 2 (pink). Hitchhiking not extended far from tens of kb around the outlier (labeled with a black outline). Positions on the scaffolds of the 'B04' genome are reported (Le Cam, personal communication). (B) Genetic map of 4 LGs containing outliers (orange or red depending on the status Fig 2.) and associated segregation distortion along the LG. Segregation distortion is reported as the natural logarithm of $1/p$ where p represents the p-value associated to the χ^2 test. Note that none of the outliers mapped close to the AvrRvi6 locus (blue) assessed by the phenotype and also the strong segregation distortion observed in the region of PAAMCT8. (C) A particular focus on the fifth LGs where two outliers mapped together in the same linkage group. Note that PTGMGC27 is also significantly distorted. (D) Genetic map of the fifth LG adding several SNPs in the surrounding region of three AFLP outliers (PCTMCT6, PTGMTC15 and PAAMCT8). Note that outliers mapped together even if those loci located on several and distinct regions throughout the fungal genome.

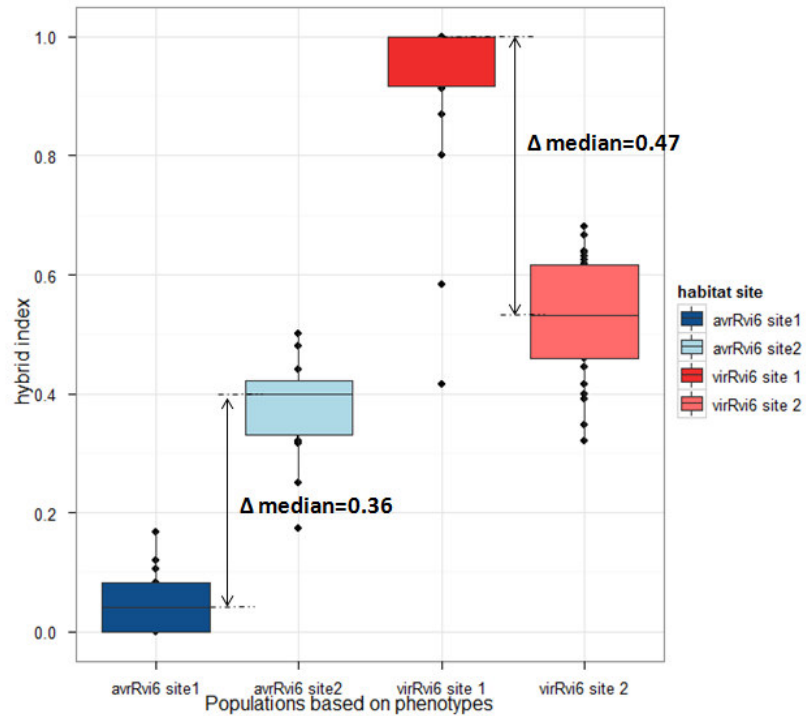
Looking at the 'holy grail' of ecological speciation: dreams and disillusion

Ecological speciation is a reemerging topic for evolutionary biologists and recent improvements on the F_{ST} outlier methods lead to a resurging of empirical studies looking for the genetic basis of adaptation and the contribution of ecological speciation. A burgeoning literature on genome scan pinpoints candidate loci for adaptation and speciation genes [e.g.

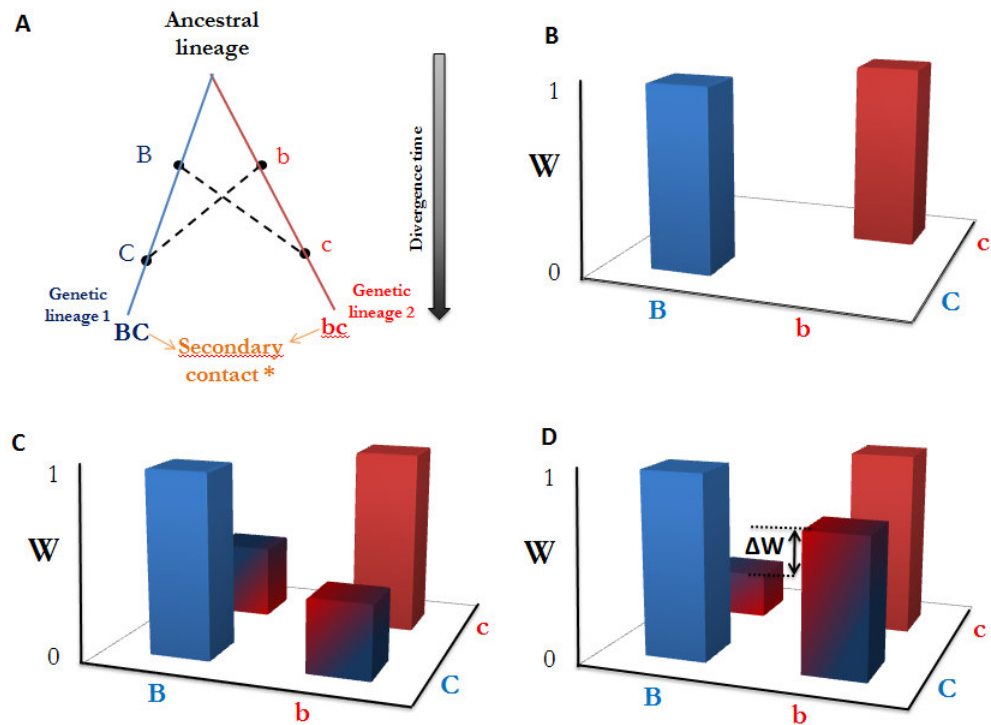
(19, 20)]. However a large part of outliers might be explained by others factors such as endogenous selection [(4) but see also (13)]. Empirically, we provided a strong evidence of the unsuspected contribution of endogenous barriers to the reproductive isolation between two recently diverging population of a plant pathogen. Indeed, pathogens are often introduced as so excellent models for ecological speciation that alternative hypothesis of endogenous barriers are often fully neglected. Taking into account the “null hypothesis” of speciation through an accumulation of incompatibilities, our genome scan study showed that exogenous barriers are neither the single nor necessary the main contributor to the pathogen diversification. Given the strength of the host selective pressures (1, 6) and the short time of divergence generally assumed for pathogens (7), our finding suggests that non-ecological process can be even more common for organisms less subject to ecological selection than pathogens. Additionally, dissociating endogenous and exogenous barriers seems much more challenging than previously thought, even using both a natural population approach and an *in vitro* crossing experiment strategy. Consequently, a large part of current evidences for ecological speciation suffers of uncompleted rejection of non-ecological hypotheses and we encourage authors to rigorously test non-ecological hypotheses using integrative approach combining both natural populations and crossing experiments before invoking the putative role of ecological speciation.

1. T. Giraud, P. Gladieux, S. Gavrillets, *Trends Ecol Evol* **25**, 387 (Jul, 2010).
2. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer Associates, Sunderland, Massachusetts, 2004), pp. 545 p.
3. N. H. Barton, M. A. de Cara, *Evolution* **63**, 1171 (May, 2009).
4. N. Bierne, J. Welch, E. Loire, F. Bonhomme, P. David, *Molecular Ecology* **20**, 2044 (May, 2011).
5. N. H. Barton, G. M. Hewitt, *Annual Review of Ecology and Systematics* **16**, 113 (1985).
6. T. Giraud, *Trends Parasitol* **22**, 151 (Apr, 2006).
7. E. H. Stukenbrock, B. A. McDonald, *Annual Review of Phytopathology* **46**, 75 (2008).
8. R. C. Lewontin, J. Krakauer, *Genetics* **74**, 175 (1973).
9. N. H. Barton, M. Shpak, *Genet Res* **75**, 179 (Apr, 2000).
10. B. Charlesworth, M. Nordborg, D. Charlesworth, *Genetics Research* **70**, 155 (Oct, 1997).
11. P. Nosil, D. J. Funk, D. Ortiz-Barrientos, *Mol Ecol* **18**, 375 (Feb, 2009).
12. T. L. Turner, M. W. Hahn, S. V. Nuzhdin, *PLoS Biol* **3**, e285 (Sep, 2005).
13. M. Nei, Y. Suzuki, M. Nozawa, *Annu Rev Genomics Hum Genet* **11**, 265 (Sep 22, 2010).
14. D. R. Matute, I. A. Butler, D. A. Turissini, J. A. Coyne, *Science* **329**, 1518 (Sep 17, 2011).
15. P. Gladieux *et al.*, *Mol Ecol* **20**, 4521 (Nov, 2011).
16. F. Guérin, P. Gladieux, B. Le Cam, *Fungal Genetics and Biology* **44**, 284 (2007).
17. F. Guerin, B. Le Cam, *Phytopathology* **94**, 364 (2004).
18. M. Currat, M. Ruedi, R. J. Petit, L. Excoffier, *Evolution* **62**, 1908 (Aug, 2008).
19. A. Bonin, P. Taberlet, C. Miaud, F. Pompanon, *Mol Biol Evol* **23**, 773 (Apr, 2006).
20. C. L. Meyer, R. Vitalis, P. Saumitou-Laprade, V. Castric, *Mol Ecol* **18**, 2050 (May, 2009).

Supplementary Data



Supplementary data Fig 1: Box plot representing the range of hybrid indexes for avrRvi6 (blue) and virRvi6 populations (red) from the site 1 and the site 2 (avrRvi6 in light blue, virRvi6 in pink). Note the differential introgression of alleles between the two genetic backgrounds with a preferential introgression of avrRvi6 alleles in the virRvi6 background (Δ medians between virRvi6 populations = 0.47 compared to Δ medians between avrRvi6 populations = 0.36).



Supplementary data Fig 2: Model for Dobzhansky-Muller incompatibilities (DMI) in haploids. (A) Two genetic lineages diverged in allopatry from an ancestral lineage, they independently accumulated mutations that will be fixed by drift or natural selection. Here, to simplify, only two new alleles arose and subsequently fixed in each derived genetic lineages. These alleles are incompatible with the other new alleles in the genetic background (dotted lines). These genetic barriers might be revealed if populations outbreed in a secondary contact zone. (B) Fitnesses of parental lineages before a secondary contact. (C) An example of possible hybrid depression due to DMIs in a situation of secondary contact. (D) Another example of hybrid depression where DMIs do not induce the same hybrid depression depending on which incompatible alleles is introgressed in a given genetic background (here $s_{Bc} > s_{bC}$). Note that in this latter case, a differential fitness between hybrids (ΔW) might contribute to an over-representation of some alleles (here alleles b and C) which can be detected as a significant segregation distortion using a genetic map approach.

Supplementary Data Table 1: Sampling origin, phenotype at the *avrRvi6* locus and hybrid index of *V. inaequalis* strains used for this study.

Sampling site	Sampling year	Habitat	Phenotype at the <i>avrRvi6</i> locus	Strain	Hybrid index	
Site 1 (La Folletière, France)	2004	non- <i>Rvi6</i>	<i>avrRvi6</i>	F367-04	0.045	
				F370-04	0.080	
				F371-04	0.000	
				F372-04	0.000	
				F373-04	0.000	
				F374-04	0.120	
				F375-04	0.000	
				F376-04	0.120	
				F377-04	0.105	
				F378-04	0.040	
				F379-04	0.000	
				F380-04	0.080	
				F381-04	0.083	
				F382-04	0.083	
				F383-04	0.042	
				F384-04	0.042	
				F385-04	0.167	
				F386-04	0.083	
				F388-04	0.040	
				F389-04	0.042	
		F390-04	0.040			
		F391-04	0.120			
		F392-04	0.000			
		F393-04	0.000			
		<i>Rvi6</i>		<i>virRvi6</i>	F101-04	0.933
					F103-04	1.000
					F104-04	1.000
					F105-04	1.000
					F111-04	0.947
					F112-04	1.000
					F113-04	0.800
					F114-04	1.000
					F116-04	1.000
F117-04	1.000					
F118-04	0.583					
F120-04	1.000					
F123-04	0.870					
F124-04	0.920					
F125-04	1.000					
F129-04	0.960					
F130-04	1.000					
F131-04	0.917					
F135-04	1.000					
F136-04	0.417					
F139-04	1.000					
F140-04	0.913					
Site 2 (Beaucouzé, France)	2009	non- <i>Rvi6</i>	<i>avrRvi6</i>	09BCZ001	0.440	
				09BCZ008	0.417	
				09BCZ009	0.417	
				09BCZ012	0.400	
				09BCZ014	0.333	
				09BCZ015	0.500	

Supplementary Data Table 1 (continued)

Site 2 (Beaucouzé, France)	2009	non-Rvi6	<i>avrRvi6</i>	09BCZ016	0.480
				09BCZ017	0.333
				09BCZ021	0.440
				09BCZ026	0.174
				09BCZ029	0.316
				09BCZ031	0.333
				09BCZ032	0.417
				09BCZ043	0.320
				09BCZ048	0.250
				09BCZ049	0.400
				09BCZ002	0.480
				09BCZ003	0.400
		09BCZ005	0.625		
		09BCZ010	0.522		
		09BCZ013	0.417		
		09BCZ020	0.640		
		<i>Rvi6</i>	<i>virRvi6</i>	09BCZ022	0.348
				09BCZ023	0.520
				09BCZ034	0.476
				09BCZ037	0.522
				09BCZ039	0.320
				09BCZ041	0.609
				09BCZ046	0.636
				09BCZ047	0.391
09BCZ064	0.680				
09BCZ066	0.320				
09BCZ067	0.400				
09BCZ068	0.583				
09BCZ074	0.632				
09BCZ075	0.667				
09BCZ076	0.444				
09BCZ077	0.563				
09BCZ078	0.478				
09BCZ083	0.480				
09BCZ084	0.458				
09BCZ085	0.600				
09BCZ086	0.542				
09BCZ089	0.682				
09BCZ090	0.391				
09BCZ092	0.560				
09BCZ094	0.583				
09BCZ095	0.680				
09BCZ097	0.583				
09BCZ098	0.619				
09BCZ110	0.524				
09BCZ111	0.522				
09BCZ112	0.579				
09BCZ113	0.611				
09BCZ115	0.625				
09BCZ117	0.348				
09BCZ118	0.458				
09BCZ120	0.680				

Supplementary Data Table 2: Selective bases used for AFLP combinations used in this study. The numbers of AFLP polymorphisms genotyped on the natural populations and on the progeny are reported.

2-bp added for selective primers		number of loci genotyped	
Pst restriction site	Msp restriction site	natural population	<i>In vitro</i> progeny
'CT'	'AA'	13	
	'AC'	33	
	'CC'	7	
	'CT'	13	3
	'GA'	15	
	'GT'	18	5
	'TA'	6	
	'TC'	15	
	'TG'	22	5
'AC'	'TT'	14	4
	'AA'	12	4
	'GC'	7	
	'GG'	7	2
	'TA'	12	9
	'TC'	14	
	'TG'	21	10
'GA'	'TT'	14	
	'AC'	20	
	'AG'	18	
	'CG'	26	15
	'CT'	15	8
	'GT'	14	
	'TA'	8	
'TG'	'TC'	16	7
	'TG'	13	
	'AC'	28	10
	'AG'	24	
	'AT'	10	
	'CC'	5	
	'CG'	25	
	'CT'	20	
	'GA'	21	
	'GC'	28	10
	'GG'	20	3
'GT'	26		
'AA'	'TC'	19	6
	'TG'	25	9
	'TT'	31	7
	'AA'	17	4
	'AC'	17	
	'AG'	17	5
	'AT'	23	
	'CA'	20	
	'CC'	32	
	'CA'	17	
	'CT'	14	11
	'GA'	19	
'CC'	'GC'	19	
	'GG'	21	
	'TA'	28	
	'TC'	16	
	'TG'	10	
	'GA'	23	10
Overall	'GG'	7	
	'GT'	14	
		969	147

Supplementary data Table 3: Sequence of the primers used for PCR and KASPar assays.

Locus name	Assays	Allele	Primers	
			Forward	Reverse
PAAMCT8	Classic PCR	Allele 1	AAGGCGAAGGCGAAGGCG	GACGTTGGAATACACATTC
PAAMCT8	Classic PCR	Allele 2	CGAAGACGAAGGCGAAGAAT	ATTCCAAGCGCTTTCACC
PCTMCT6	Classic PCR	Allele 1	CAAATGAAGCTGCAGTCT	TGATACCATTCCATTCCAG
PCTMCT6	Classic PCR	Allele 2	GCAATGAAGCTGCAGTTC	CAGGGACACTGTTCAAAGA
PTGMGC27	Classic PCR	Allele 1	GCTTGTGAGTTCTGACCGAT	TTGCCATCCGATACACT
PTGMGC27	Classic PCR	Allele 2	GGCCCTGGGACTGCAGTGT	ATGAAATACGAGAATTCGGAGGTC
PTGMTC15	Classic PCR	Allele 1	GGAAACCAATCTTTGAACC	CGAAAATCTCGCACTCCG
PTGMTC15	Classic PCR	Allele 2	TCACCTAGATCACGGAAAAC	CGCGAAAACCTCGCACTCTA
PAAMCT8_293421	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTACTCGTCCGACGCACTTTCG	GTGCAGCGTTGGAGTCTGTGA
PAAMCT8_293421	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTAACGCTCCGACGCACTTTCA	GTGCAGCGTTGGAGTCTGTGA
PAAMCT8_323554	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGGCAAAATTCGATCCGACACAAT	GTCAGTCCGAGCTGGGATTTTCT
PAAMCT8_323554	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTGGCAAAATTCGATCCGACACAAC	GTCAGTCCGAGCTGGGATTTTCT
PAAMCT8_346759	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGCAACTTGGAAAGGATGGCTACACC	ACGAAAGGGACTCACCTCATCTGT
PAAMCT8_346759	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTAACGCTTGGAAAGGATGGCTACACA	ACGAAAGGGACTCACCTCATCTGT
PAAMCT8_360466	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGAACGGATAGGCTGGAGAGATCC	ATATCTCAGTCTCCAGCCGCATA
PAAMCT8_360466	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTAACGCGATAGGCTGGAGAGATCT	ATATCTCAGTCTCCAGCCGCATA
PAAMCT8_365625	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCCGAGATGCATACCATCTAGAAC	AGTGTCTGGAATTCGCCATTTCTT
PAAMCT8_365625	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTCCGAGATGCATACCATCTAGAAG	AGTGTCTGGAATTCGCCATTTCTT
PAAMCT8_369966	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGTCTTATGTCGCCCCATCCG	ATGTGCTTCCAAGTTGATCTTCAC
PAAMCT8_369966	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTCTTATGTCGCCCCATCCA	ATGTGCTTCCAAGTTGATCTTCAC
PAAMCT8_373567	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGTGTCTGTGGCATCTCGACC	GTTCTTCCGAGCGGTGACTCT
PAAMCT8_373567	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTGTGTCTGTGGCATCTCGACT	GTTCTTCCGAGCGGTGACTCT
PAAMCT8_374518	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTTCCGCTGTGGCTCCGCCGTA	CAITTTGCTCAGACCGGAATCG
PAAMCT8_374518	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTCCGCTGTGGCTCCGCCGTTG	CAITTTGCTCAGACCGGAATCG
PAAMCT8_376364	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCAGAATATCCCGGAGGAATC	GAAGGGTGAAGCGCTTGGAAAT
PAAMCT8_376364	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTCAGAATATCCCGGAGGAAT	GAAGGGTGAAGCGCTTGGAAAT
PAAMCT8_377069	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTATCGGCAAACTTGTTTTGTTCCTC	AAGTGTGCCTTCTTCTTGTTCAGG
PAAMCT8_377069	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTATCGGCAAACTTGTTTTGTTCCTG	AAGTGTGCCTTCTTCTTGTTCAGG
PAAMCT8_382134	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGCGATTCAGTGGTGTGAGTAAG	GGAAGGGAATATTTCAITATTCGTTGA
PAAMCT8_382134	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTCCGATATGTTGGTGTGAGTAAA	GGAAGGGAATATTTCAITATTCGTTGA
PAAMCT8_383824	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGTTCCGAATCTACTGGTCCGAGGC	TAITCAATAAACAGGCCGATGGAC
PAAMCT8_383824	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTGTTCCGAATCTACTGGTCCGAGGT	TAITCAATAAACAGGCCGATGGAC
PAAMCT8_389703	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGTTCCGATAGCAGCTCGGTCA	ACTCGATCTTGTCCAGCAACAA
PAAMCT8_389703	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTGTTCCGATAGCAGCTCGGTGCG	ACTCGATCTTGTCCAGCAACAA
PAAMCT8_395083	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCCATCGACGCTCGCGGTGTC	AGATTGGGAGGGGCACTTCG
PAAMCT8_395083	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTCCATCGACGCTCGCGGTGTA	AGATTGGGAGGGGCACTTCG
PAAMCT8_406684	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCTTCTTAGGTCTTCAGGCGTCAA	TTTGATAAGGATACAAGAGACAAGA
PAAMCT8_406684	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTCTTCTTAGGTCTTCAGGCGTCCAG	TTTGATAAGGATACAAGAGACAAGA
PCTMCT6_48867	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTAATTCGGCTGCATCACTCTCGTTGAC	GCTCAATGCTCCGCTGTGG
PCTMCT6_48867	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTATTCGGCTGCATCACTCTCGTTGAT	GCTCAATGCTCCGCTGTGG
PCTMCT6_52681	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTAGAGTATCAITTCGTTGACCCCAAC	GCGACGGACTGTTGGAATAGTAGT
PCTMCT6_52681	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTAGAGTATCAITTCGTTGACCCCAAA	GCGACGGACTGTTGGAATAGTAGT
PCTMCT6_60910	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCAATAAGTTGAAAGAGTGAGTTCATATCT	CACCTTCAGCAACAGCTGTTCA
PCTMCT6_60910	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTCAATAAGTTGAAAGAGTGAGTTCATATCTG	CACCTTCAGCAACAGCTGTTCA
PCTMCT6_68130	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTTTCGATTTCTGGTCTCGACATCGAT	TGGGTGTATCATATCATGATGTCAG
PCTMCT6_68130	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTTCGATTTCTGGTCTCGACATCGAG	TGGGTGTATCATATCATGATGTCAG
PCTMCT6_76311	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCAATTCAGCAAAATCGAAAACGGAAAC	CTGTATGCGTTGAACATAAGCCTTC
PCTMCT6_76311	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTCAATTCAGCAAAATCGAAAACGGAAAT	CTGTATGCGTTGAACATAAGCCTTC
PCTMCT6_78024	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTTTCAGTCAITGCTGGGATCGTGT	AACAGGTAGGGTGTGCTGCAATCTT
PCTMCT6_78024	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTTCAGTCAITGCTGGGATCGTGC	AACAGGTAGGGTGTGCTGCAATCTT
PCTMCT6_79771	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTTAATCCCGATCCAATCGTGTGCAA	TCTTGAAGAGGTGTTTCGAGAAGGA
PCTMCT6_79771	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTAATCCCGATCCAATCGTGTGCGAG	TCTTGAAGAGGTGTTTCGAGAAGGA
PCTMCT6_86822	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGCTTCAITGCGGCAACAACAAT	GATGGCTATCATTTGCTGAATACC
PCTMCT6_86822	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTGCTTCAITGCGGCAACAACAAC	GATGGCTATCATTTGCTGAATACC
PCTMCT6_89397	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGCCGACGTGGTTTTTGAAGAAGTAA	AAAACCTCGGCTCGAAAACCATC
PCTMCT6_89397	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTGCCGACGTGGTTTTTGAAGAAGTAT	AAAACCTCGGCTCGAAAACCATC
PCTMCT6_93202	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTACCCGACGCTAGCATCTTCG	GAAGTGGTCCGATTAACACTGTTG
PCTMCT6_93202	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTACCCGACGCTAGCATCTTCCTCA	GAAGTGGTCCGATTAACACTGTTG
PCTMCT6_94510	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCCCAATTTCTAGTGTGTGCTTCTCG	GGAGAAGTTAGCGGGAGCAAACC
PCTMCT6_94510	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTCCCAATTTCTAGTGTGTGCTTCTCA	GGAGAAGTTAGCGGGAGCAAACC
PCTMCT6_96195	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGTTGAGTGCAGTGCACCCCGCC	GTTGGTGTGGGGCAGCAAAG
PCTMCT6_96195	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTGAGTGCAGTGCACCCCGCT	GTTGGTGTGGGGCAGCAAAG
PCTMCT6_97458	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGATGACCAAGTCCCTCTCCA	GCAGAGTCAAGTTCAGGCAGATGAA
PCTMCT6_97458	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTGATGACCAAGTCCCTCTCCG	GCAGAGTCAAGTTCAGGCAGATGAA
PCTMCT6_101651	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGGGGTCTTGGCAATATAGGAATA	TTGCTACAGACGCAAAATCTCCA
PCTMCT6_101651	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTGGGGTCTTGGCAATATAGGAAT	TTGCTACAGACGCAAAATCTCCA
PCTMCT6_102238	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCAAAAGAACCACTACTCTCCACTT	TGGCGCTCAGCGGACTCGTAT
PCTMCT6_102238	KASPar	Allele 2	GAAGGTCCGAGTCAACGGATTTAAAGAACCACTACTCTCCACTC	TGGCGCTCAGCGGACTCGTAT

Supplementary Data Table 3 (continued)

PCTMCT6_104554	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGACCTAACCTCTTGTTAGCGATCAA	GCTCAAGGTTCTGCTCTCCAGAT
PCTMCT6_104554	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCAACCTAACCTCTTGTAGCGATCAT	GCTCAAGGTTCTGCTCTCCAGAT
PCTMCT6_106805	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCAACGACCGAATGTAGGTTCTT	TATAATGTCCAATTAGTCGACGGGCTTTT
PCTMCT6_106805	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCAACGACCGAATGTAGGTTCTG	TATAATGTCCAATTAGTCGACGGGCTTTT
PCTMCT6_107833	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGGCACTGGAGCTGCAG	CTCGAGAAAGCACCCTGCTT
PCTMCT6_107833	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGGCTGGCACTGGAGCTGCAA	CTCGAGAAAGCACCCTGCTT
PCTMCT6_109834	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCCAATCTCATCACTCACCTT	TCCGAGGATCGATGGACATCAITT
PCTMCT6_109834	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCCAATCTCATCACTCACCCA	TCCGAGGATCGATGGACATCAITT
PCTMCT6_111416	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCTCTTTTTCAGCTCTGTTCTC	GGAGCACCATTATTCAGACGCGAA
PCTMCT6_111416	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGCTCTTTTTCAGCTCTGTTCTT	GGAGCACCATTATTCAGACGCGAA
PCTMCT6_112575	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGCATTTGATTTTCGCTCTTTCAT	CAACGGGCCAAAAGAAATCGGTACAT
PCTMCT6_112575	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCACTTGAITTTCCGCTCTTTCAC	CAACGGGCCAAAAGAAATCGGTACAT
PCTMCT6_115627	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCAGCTCCGATATCCAAGTGGT	GACGAGGATGAGATGCTGGTGAT
PCTMCT6_115627	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTAGCTCTGCGATATCCAAGTGGC	GACGAGGATGAGATGCTGGTGAT
PCTMCT6_116224	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCAGTGAACGACGAAAGAGTGGC	TTGCGGTGATCTCTGTAACCAT
PCTMCT6_116224	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCACTGTAACGACGAAAGAGTGGT	TTGCGGTGATCTCTGTAACCAT
PCTMCT6_120059	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCGGTATACCCCAATTGGTCAAGGT	AAGACTCGCTCGTCCAGTGTGAT
PCTMCT6_120059	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGGATACCCCAATTGGTCAAGGC	AAGACTCGCTCGTCCAGTGTGAT
PTGMGC27_49181	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTAACTTATGACAATCCGACAGTGTATC	GAACGCTGCTGATGTCAGGTG
PTGMGC27_49181	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTAACTTATGACAATCCGACAGTGTATT	GAACGCTGCTGATGTCAGGTG
PTGMGC27_58057	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGAGTACGGAGTCTTGTATCTCGTT	GGATATCTCTTGTCTTCAITTCGAG
PTGMGC27_58057	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGAGTACGGAGTCTTGTATCTCGTC	GGATATCTCTTGTCTTCAITTCGAG
PTGMGC27_61448	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGAGGAGAGCGAGTGGAAAGAT	AGCAGCACAAAGTGTITAGCACTGG
PTGMGC27_61448	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGAGGAGAGCGAGTGGAAAGAC	AGCAGCACAAAGTGTITAGCACTGG
PTGMGC27_62740	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGCTAGGAAAGGGTGGAGAAGAT	GTCCTCTTCAAGTATCTCCGCTA
PTGMGC27_62740	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCTGAGGAAAGGGTGGAGAAGAC	GTCCTCTTCAAGTATCTCCGCTA
PTGMGC27_65194	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGAGGACAACACGACTAGGAA	GTCTTCTCCGGTCTCTCGTT
PTGMGC27_65194	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCTGAGGACAACACGACTAGGAG	GTCTTCTCCGGTCTCTCGTT
PTGMGC27_65694	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGGGATGCGAATATGGATGGGA	AAACCTCTACCACTATCCAAACA
PTGMGC27_65694	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGGGATGCGAATATGGATGGGG	AAACCTCTACCACTATCCAAACA
PTGMGC27_68979	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCAGAAGTGAAGAAGAAGTAGTGAAAGA	GAACCTCATTCATGTCACCCCAACA
PTGMGC27_68979	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGAAGAAGAAGTAGTGAAAGAAGG	GAACCTCATTCATGTCACCCCAACA
PTGMGC27_72428	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTAATACCCTTCAAACGGCCCAA	TGTGCAAGTCTTAGTACACAATGGA
PTGMGC27_72428	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTAAATACCCTTCAAACGGCCCAAG	TGTGCAAGTCTTAGTACACAATGGA
PTGMGC27_72684	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTAGGAGAGTGTGTTGGAATGGTTAT	AAGAGAGCTTCTCTCCAGCAGTT
PTGMGC27_72684	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGAGGAGAGTGTGTTGGAATGGTTAT	AAGAGAGCTTCTCTCCAGCAGTT
PTGMGC27_77231	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTAGGACTGACAAGCTGGCTTGTAT	GCAGACATGGTCCATCCGGTA
PTGMGC27_77231	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGACTGACAAGCTGGCTTGTAT	GCAGACATGGTCCATCCGGTA
PTGMGC27_80246	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCCTATTGATGCTCAGGATAATTG	AAGATTCTCATTCATGGGGCAAC
PTGMGC27_80246	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCTCATTGATGCTCAGGATAATTG	AAGATTCTCATTCATGGGGCAAC
PTGMGC27_80338	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGGATTGAGGAGCTTCAAACGCT	ATGAGAATCTTCCAGATGCTGCTAA
PTGMGC27_80338	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGAGGAGCTTCAAACGCC	ATGAGAATCTTCCAGATGCTGCTAA
PTGMGC27_84166	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGCTTCCAGCCCTCTGGC	ACTCTCACCAACCGCTATTCT
PTGMGC27_84166	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGCTTCCAGCCCTCTGGG	ACTCTCACCAACCGCTATTCT
PTGMGC27_85019	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGAAACACGACGAGTCTGCACAG	GCAATCGTCCGATGTTCCCGTAT
PTGMGC27_85019	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGAAACACGACGAGTCTGCACAA	GCAATCGTCCGATGTTCCCGTAT
PTGMGC27_93521	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCGGTGCTAGTCTCAAAGGAGGGCT	ACTTCACTTCACTTCTTCACTCCA
PTGMGC27_93521	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGCTGCTAGTCTCAAAGGAGGGCG	ACTTCACTTCACTTCTTCACTCCA
PTGMGC27_96082	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGCTCCAGTGAAGAAGTCAAGTC	CTTCGTCGACCAACCGACCTC
PTGMGC27_96082	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGCTGCTCCAGTGAAGAAGTCAAGTT	CTTCGTCGACCAACCGACCTC
PTGMGC27_101923	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGCGAATGAGCAATTTGCATTGCT	ACAACCTGTTATATTTGCCCTGCT
PTGMGC27_101923	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGCGAATGAGCAATTTGCATTGCC	ACAACCTGTTATATTTGCCCTGCT
PTGMTC15_18289	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGTTGAGGAGATATGAAGTCGGTTGTG	AAACAGAGGAGTGCTCAGCAAAT
PTGMTC15_18289	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGTTGAGGAGATATGAAGTCGGTTGTC	AAACAGAGGAGTGCTCAGCAAAT
PTGMTC15_19404	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTTCCATTACAACCTCTCCACTTTT	CGTAATGTTAAGAAGCTGCGGTGA
PTGMTC15_19404	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCCATTACAACCTCTCCACTTTT	CGTAATGTTAAGAAGCTGCGGTGA
PTGMTC15_26147	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGGTGGGAAGAAGACAGAGTTACGA	TCTTCAGCGGTTAGCGGACATT
PTGMTC15_26147	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGGTGGGAAGAAGACAGAGTTACGG	TCTTCAGCGGTTAGCGGACATT
PTGMTC15_28775	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCCACCCGCTTTCAGCAATCATG	GCTGAAGAGTCCACTTTTGTGTCCAT
PTGMTC15_28775	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCCACCCGCTTTCAGCAATCATA	GCTGAAGAGTCCACTTTTGTGTCCAT
PTGMTC15_35775	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGCAAAAAGCTCCATACCATCCG	GTGAAACTACCACATGAACGTGCGAA
PTGMTC15_35775	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTAGCAAAAAGCTCCATACCATCCA	GTGAAACTACCACATGAACGTGCGAA
PTGMTC15_36689	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTAAGCTATTGGCCATCTTGGTGTC	CCAATCCAACCTAAAACCTCAACTGC
PTGMTC15_36689	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTAAAGCTATTGGCCATCTTGGTGTC	CCAATCCAACCTAAAACCTCAACTGC
PTGMTC15_39955	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGCTGCTGTTGTTGATTCTCTCAC	CTTCGGACACTCGAGCCGAA
PTGMTC15_39955	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCTGCTGCTGTTGTTGATTCTCTCAT	CTTCGGACACTCGAGCCGAA
PTGMTC15_42435	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTATCAGGTTCAAAGATTGGTTTCTG	GCCGACAGAAGGAGTCAAGGACA
PTGMTC15_42435	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGATCAGGTTCAAAGATTGGTTTCTTA	GCCGACAGAAGGAGTCAAGGACA
PTGMTC15_43825	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGTAACCTACCCCGGAGCG	GGGGTGGATGGCTTCTGTAGAA
PTGMTC15_43825	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTAAAGTAACTACCCCGGAGGACA	GGGGTGGATGGCTTCTGTAGAA
PTGMTC15_45776	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTACCAAGTAACTCCGAGCTCTCT	CGTGCTGTATAGCCCACTGTGAA
PTGMTC15_45776	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCCAGTAACTCCGAGCTCTCTG	CGTGCTGTATAGCCCACTGTGAA

Supplementary Data Table 3 (continued)

PTGMTC15_44355	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCCGAAAACAACGCAAGGACCG	CTGGTCATGAATCAGTCTGGCACTT
PTGMTC15_44355	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCCGAAAACAACGCAAGGACCA	CTGGTCATGAATCAGTCTGGCACTT
PTGMTC15_47122	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGACCTGTTGTAGTCAAGAGCTCTT	TTTAGCGGATTGTCAGAAGGTTGTTGAT
PTGMTC15_47122	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGACCTGTTGTAGTCAAGAGCTCTA	TTTAGCGGATTGTCAGAAGGTTGTTGAT
PTGMTC15_49204	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCGGTTACGAGAGAGCAGAGTC	ACACATGAGCAGAGTCACTAAAAGAGAT
PTGMTC15_49204	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCCGGTACGAGAGAGCAGAGTG	ACACATGAGCAGAGTCACTAAAAGAGAT
PTGMTC15_51409	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCCCGCTCAATGACATCGCTAAC	GCGCATCCAAGTTGGAGGTTTATGTT
PTGMTC15_51409	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGCCCGCTCAATGACATCGCTAAT	GCGCATCCAAGTTGGAGGTTTATGTT
PTGMTC15_51727	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTCAAAAACACTGCCAATCCAATCTCGC	GACTGGATATCAGGTACCGTTGGAA
PTGMTC15_51727	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTCAAAAACACTGCCAATCCAATCTCGT	GACTGGATATCAGGTACCGTTGGAA
PTGMTC15_53052	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGATCTGACGCGGAGACAAGAAC	TTCTTCAACCTGCTCGTTTTCTGTTCAAA
PTGMTC15_53052	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGATCTGACGCGGAGACAAGAAA	TTCTTCAACCTGCTCGTTTTCTGTTCAAA
PTGMTC15_58670	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTTAGTAACTCGGATAGCCGTA	AGAGACCTTACGAGATAACG
PTGMTC15_58670	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTAGTAACTCGGATAGCCGTG	AGAGACCTTACGAGATAACG
PTGMTC15_61339	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTTTATCGGTAAGTAGGTAAGATATATTCTTAG	ATAAAGAAGAAAATAAAGAGGCGGAAA
PTGMTC15_61339	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTTTATCGGTAAGTAGGTAAGATATATTCTTAA	ATAAAGAAGAAAATAAAGAGGCGGAAA
PTGMTC15_62357	KASPar	Allele 1	GAAGGTGACCAAGTTCATGCTGGAGAATATAAGATCTATATAGGTAAGCGTA	TCGGACAGGGATTCCAGGTTTATAG
PTGMTC15_62357	KASPar	Allele 2	GAAGGTCGGAGTCAACGGATTGGAGAATATAAGATCTATATAGGTAAGCGTG	TCGGACAGGGATTCCAGGTTTATAG

Supplementary Materials & Methods

Fungal sampling and AFLP genotyping

Fungal individuals were sampled in two different orchards located in the northwest of France. Both orchards contain *Rvi6* and non-*Rvi6* *M. x domestica* cultivars in sympatry (Supplementary data Table 1). The site 1 is a cider orchard located at La Folletière (Normandy, France). In this site, 24 strains were isolated from the cv. *Rvi6* ‘Judeline’ and 22 from the non-*Rvi6* cv. ‘Antoinette’. The site 2 is an experimental breeding orchard located in Angers (Pays de la Loire, France) which is about 250 km distant from the site 1. In this orchard, a progeny between cv. ‘Prima’ (*Rvi6*) and ‘TN10-8’ (a non-*Rvi6* apple accession) was randomly planted. In this second site, 30 strains were sampled on *Rvi6* and 28 on non-*Rvi6* hosts. We obtained a total of 104 single-spore isolates.

Fungal DNA was extracted from mycelial cultures using a phenol/chloroform protocol as previously described [1]. Individuals were genotyped with 55 AFLP primer-enzyme combinations (Supplementary data Table 2). AFLP protocol used was previously published (Leroy *et al.*, in revision) with minor changes including the use of primers with two selective bases and either GS500-TAMRA or GS500-LIZ as fluorescently labeled size standards (Applied Biosystems, Foster City, California, USA). Only AFLP fragments size comprised between 70 and 395 bp were scored under Genemapper 4.0 software (Applied Biosystems, Foster City, California, USA). The presence or absence of a polymorphic fragment was binary coded. Subsequently each AFLP combination was assessed independently for error rates between replicates under the AFLPscore package [2] available within the R statistical software environment [3]. We then removed loci expressing low allele frequencies (< 5%) and loci in full linkage disequilibrium. The ‘high-quality dataset’ contains 969 AFLP loci and is available through the DRYAD system under accession <http://XXXXXXXXXX>.

Outliers location in the *V. inaequalis* genome and SNP genotyping

Genomic location of outliers

Two methods were used to find genomic locations of outliers. First we used a polyacrylamide gel electrophoresis approach to isolate AFLP fragments. For each AFLP fragment, migration was performed with 0.4 μ L of each PCR product. Amplified fragments were separated on 10% denaturing polyacrylamide gels by electrophoresis at 35 watts for 5 or 6 hours and silver staining was performed using the Creste’s protocol [4]. For each band, two or three positive and negative samples were spotted into adjacent wells to easily detect the expected polymorphism. The band was then isolated from the gel and deposited into a well containing 10-20 μ L of sterilized water. DNA was separated from acrylamid gel by heating samples at 60°C for at least one hour. DNA was then used as matrix for a new PCR using the selective primers of the combination. Before sequencing, a validation of the polymorphism was done by manually scoring the extracted band against a fluorescently labeled size standard. Second we used an AFLP *in silico* approach using ISIF (‘*In Silico* Fingerprinting’ program; [5]) to detect putative AFLP fragment in genomes. ISIF used one or more full genome sequences as an input to simulate an AFLP. Here *in silico* AFLP was performed on 3

different *avrRvi6* genomes sequences (EUB04, INRA, France; WS1, ARC, South Africa; MNH120, PFR, New Zeland). We kept for further investigation all AFLP fragments with an expected AFLP fragment size which is less than 2-bp distant of the observed AFLP fragment size obtained with capillary electrophoresis.

To validate the genomic location of the fragment, PCR were performed using new primers designed on the flanking regions of the fragment. Sequencing of these amplicons was performed in order to determine the mutation site leading to the AFLP polymorphism. For each, we sequenced 3 fragments from DNA that amplify the AFLP marker and 3 other fragments from DNA that did not amplify. Then new specific primers (supplementary table 3) were designed to become diagnostic of the two different alleles. New primers were then used to validate the location by re-genotyping a subsample of strains.

Genotyping of additional SNPs in genomic regions of outliers

KASPar primers (KBioscience, Herts, UK) were designed to genotype single nuclear polymorphisms (SNPs) in the genetic regions where outliers were mapped. The supplementary Data Table 3 lists the primers used for the KASPar assays. KASPar genotyping was performed at the Gentyane Platform of INRA and data was analyzed with the Fluidigm SNP Genotyping Analysis (Fluidigm Systems, South San Francisco) using an NTC normalization with a confidence interval of 65%.

Fungal Phenotyping

Monoconidial suspensions of each strain from the site 2 were adjusted at a concentration of 2.5×10^5 conidia per milliliter. The mixture was then sprayed on a *Rvi6* cultivar (Prima) grafted on 'MM106' rootstock. Strains able to produce sporulating lesions on cv. Prima were considered virulent. For the samples collected in the site 1, greenhouse inoculation tests were not performed but phenotype was directly assessed through the habitat assuming exclusive habitats for *avrRvi6* and *virRvi6* populations in this orchard (see [6] for details). Avirulent and virulent strains are shown in the supplementary Data Table 1.

Data analysis

Correspondence analysis and hybrid index

Factorial correspondence analysis (CA) and Weir and Cockerham (θ), an analogous of Wright's F_{st} , were computed using Genetix software v. 4.05.02 [7]. Two CA were performed for this study. First, a CA was performed using populations from the site 1 and the site 2 based on the 969 loci dataset to express genetic differences between samples. Second, given that CA authorizes the sorting out of the contribution of each allele to these differences, we performed another CA using populations from the site 1 to range alleles following their contribution to the first axis (*i.e.* an avirulent or virulent genome) [8]. Alleles contributing to an *avrRvi6* genome in the site 1 were then encoded as a 0 and as 1 for a *virRvi6* genome for the whole dataset. For each individual, hybrid index was estimated as a number of *virRvi6* alleles divided by the number of called alleles based on a subset of 25 'diagnostic' loci. Here, a 'diagnostic' locus is a marker expressing high differences in allele frequencies between the two populations of the site 1 ($\Delta > 0.8$). Using this proxy, we assume that a hybrid index of 0

represent a fully *avrRvi6* gene and a hybrid index of 1, a fully *virRvi6* genome.

Detection of outlying loci

Outlier detection was performed using the BayeScan software [9]. This Bayesian algorithm is an extension of Beaumont & Balding's method [10]. Using an estimation of locus-population specific F_{ST} , it determinates a posterior probability of each locus to be under selection. We used a posterior probability >0.7 (e.g. [11]), close to the "substantial" threshold included in the interface to isolate outliers from the dataset.

Genetic map

Genetic linkage map was performed in a segregating F1 population composed of 103 hybrids produced by a cross between an *avrRvi6* (301) and a *virRvi6* (1066) strain [12]. F1 individuals were genotyped at 441 genomic loci including 116 AFLP loci also genotyped on natural populations (combinations used are listed in Supplementary data Table 2). Phenotyping on *Rvi6* and non-*Rvi6* cultivars was performed to map the *avrRvi6* locus. Linkage map was constructed using JoinMap v.4 (Kyazma, Wageningen, The Netherlands). Markers were assigned to linkage groups using the independence LOD parameter with a minimum LOD score of 4.0. Locus order and map distances were calculated using the regression mapping approach with the following parameters: a Kosambi's mapping function with a recombination frequency >0.45, LOD>1.0, a goodness-of-fit jump threshold for removal loci = 5.0, a number of added loci after which to perform a ripple = 1 and a third round. The segregation fit of each locus to a 1:1 ratio was examined and segregation distortions between different outlying loci were accessed using a chi-square test. The segregation fit of each locus to a 1:1 ratio was examined and segregation distortions between different outlying loci were accessed using a chi-square test. Segregation distortions were then plot throughout the linkage groups using the value of the natural logarithm of $\frac{1}{p}$. Graphical representation of the linkage groups was performed under Map Disto v. 1.7.1.5 [13]. Additionally, a second genetic linkage map was performed by adding 15 additional polymorphic KASPar SNPs to the 441 loci dataset.

1. Le Cam B, Devaux M, Parisi L (2001) Specific polymerase chain reaction identification of *Venturia nashicola* using internally transcribed spacer region in the ribosomal DNA. *Phytopathology* 91: 900-904.
2. Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Mol Ecol Resour* 8: 725-735.
3. R Development Core Team. R (2011) A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
4. Creste S, Tulmann Neto A, Figueira A (2001) Detection of Single Sequence Repeat Polymorphisms in Denaturing Polyacrylamide Sequencing Gels by Silver Staining. *Plant Molecular Biology Reporter* 19: 299-306.
5. Paris M, Despres L (2010) In silico fingerprinting (ISIF): a user-friendly in silico AFLP program. *Methods Mol Biol* 888: 55-64.
6. Gladieux P, Guerin F, Giraud T, Caffier V, Lemaire C, et al. (2011) Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Mol Ecol* 20: 4521-4532.
7. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. 4.05 ed. Montpellier: Laboratoire Génome, Populations,

Interactions, CNRS UMR 5000, Université de Montpellier II.

8. Daguin C, Bonhomme F, Borsa P (2001) The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism at locus *mac-1*. *Heredity* (Edinb) 86: 342-354.
9. Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics* 180: 977-993.
10. Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* 13: 969-980.
11. Buckley J, Butlin RK, Bridle JR (2012) Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Mol Ecol* 21: 267-280.
12. Benaouf G, Parisi L (2000) Genetics of Host-Pathogen Relationships Between *Venturia inaequalis* Races 6 and 7 and *Malus* Species. *Phytopathology* 90: 236-242.
13. Lorieux M (2012) MapDisto: fast and efficient computation of genetic linkage maps. *Molecular Breeding* 30: 1231-1235.

2. Résultats complémentaires

Dans l'article précédent, nous avons montré par une approche en population naturelle et une autre de cartographie génétique que les barrières d'isolement reproductif détectées étaient vraisemblablement d'origine post-zygotique endogène. Des arguments supplémentaires supportent cette hypothèse et seront présentés et discutés sous ce chapitre. Les résultats supplémentaires concernent des points particuliers répartis sur l'ensemble de l'article précédent. Ils ne seront pas présentés sous forme d'un seul article mais seront discutés un par un.

2.1 Conséquences des DMI sur cartes génétiques : une approche par simulation

2.1.1 Description de l'approche utilisée

Une approche par simulation a été développée pour mieux appréhender les conséquences de l'expression des incompatibilités sur les cartes génétiques. Les simulations ont été réalisées sous QuantiNEMO (1) supposant la divergence de deux populations diploïdes. Considérons à $t=0$ que deux populations ont divergé en allopatrie et que 3 nouveaux allèles ont été accumulés indépendamment dans chacune des populations à trois locus différents (Figure 1).

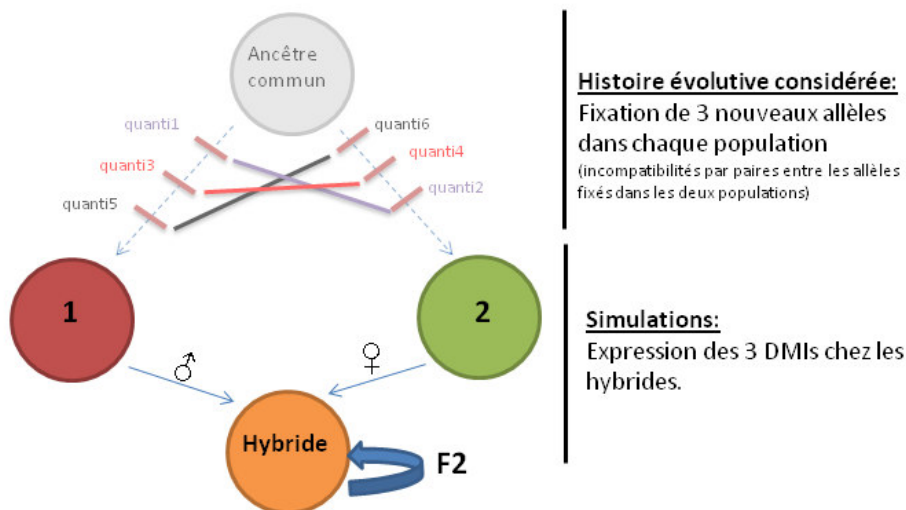


Figure 1 : Représentation synthétique de l'approche développée pour formaliser les attendus. Deux populations divergentes (1 et 2) présentent des allèles fixés différents aux 6 locus « quanti ». Les trois incompatibilités entre paires de locus sont figurées par des traits de couleurs entre les deux fonds génétiques (quanti1-quanti2 ; quanti3-quanti4 ; quanti5-quanti6). Le croisement entre les deux populations est rendu obligatoire par la migration vers un habitat « hybride » d'un seul type sexuel à partir de chaque compartiment.

Les 6 locus dits « quanti » forment trois incompatibilités génétiques de type DMI entre paires de locus chez les hybrides. Les 6 allèles « quanti » sont simulés sur 6 chromosomes

différents (en position 75 cM à chaque fois). La longueur totale des chromosomes est de 150 cM et 50 marqueurs sont répartis sur l'ensemble des chromosomes (marqueurs distants de 3 cM). Le croisement hétérogamétique est obligatoire dans un troisième habitat dit « hybride » car la migration n'est possible que pour un des deux sexes de chaque population divergente. Le croisement hétérogamétique oblige donc la recombinaison des deux pools génétiques divergents. Dans cette simulation, aucun effet de dosage allélique n'est considéré, au contraire, les allèles exprimant les incompatibilités sont considérés dominants. En effet, on suppose ici que l'expression de ces incompatibilités est particulièrement délétère pendant la phase haploïde du cycle.

2.1.2 Résultats et discussion

Les cartes génétiques ont été reconstruites sous JoinMap en utilisant les mêmes paramètres que décrits dans l'article précédent (Figure 2 pour un exemple).

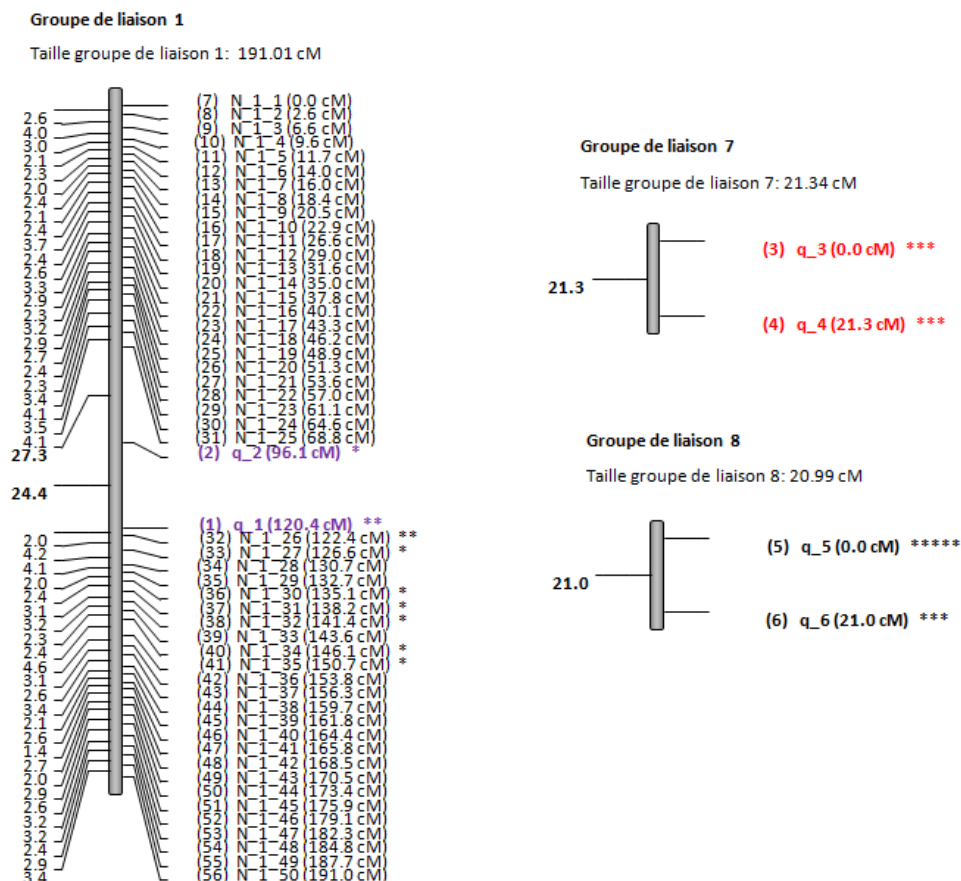


Figure 2 : Exemple de cartographie des barrières génétiques sous le modèle simulé. La simulation est réalisée sous les conditions décrites (cf § 2.1.1). Seuls les groupes de liaisons porteurs des locus « quanti » sont présentés. Les autres groupes de liaisons ne comportent que les marqueurs neutres sur une distance très proche de celle simulée (150 cM). La simulation a été effectuée en fixant la valeur sélective des hybrides (W) à de 1, 0,61, 0,14 et 0,01 pour les individus porteurs de 0, 1, 2 et 3 incompatibilités respectivement. Pour cette simulation, les individus de la F2 ayant une valeur sélective inférieure à 0.5 ont été considérées non-viables. L'intensité de la distorsion de ségrégation est représentée par des étoiles associée à la significativité du test de χ^2 : * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 1.10^{-4}$, ***** $p < 1.10^{-5}$.

Le premier résultat de cette simulation est que l'expression des incompatibilités affecte très sensiblement la carte génétique. Bien que les locus « quanti » soient tous indépendants, le déséquilibre de liaison généré par ces incompatibilités génétiques tend à cartographier ensemble les locus responsables d'une même incompatibilité sur un même groupe de liaison. Au moins un des deux caractères quantitatifs (locus « quanti ») se retrouve complètement isolé de son environnement génomique simulé (généralement les deux cf figure 2). En conséquence, la localisation des locus génétiquement proches ne renseigne pas nécessairement de la proximité physique entre un locus et une barrière génétique.

La deuxième information majeure est que les locus apparaissent très fortement distordus, cette distorsion peut d'ailleurs être observée sur son environnement génomique (les marqueurs les plus proches, p.ex. se référer au groupe de liaison 1 Figure 2). Toutefois, si la contre-sélection des hybrides à cause des incompatibilités se traduit par un très fort signal de distorsion, les locus responsables de ces incompatibilités vont tendre à être complètement isolés de leur environnement génomique. Ce résultat est conforme aux observations de l'article précédent où certains outliers (et SNP à proximité de ces derniers) co-localisent sur un même groupe de liaison bien qu'ils soient répartis sur des scaffolds différents. En outre, notons qu'ici l'intensité de la sélection est identique pour l'ensemble des incompatibilités. Chez les diploïdes la détection des barrières endogènes par un écart au ratio 1 :2 :1 est beaucoup plus facile à observer de part la présence des hétérozygotes que chez les haploïdes avec un ratio 1 :1. Inversement chez les haploïdes, seules les fortes différences d'intensité de sélection entre les paires d'allèles produisant les incompatibilités peuvent se traduire par une distorsion de ségrégation significative (p. ex. si interaction due à l'interaction entre A_1 et B_2 n'est pas autant délétère que celle entre A_2 et B_1 , voir Figure Supplémentaire 2 article 1).

La troisième information est que l'expression des incompatibilités tend à augmenter la taille de la carte génétique plutôt qu'à la réduire, contrairement à l'idée souvent admise [p. ex. (2)]. D'autres auteurs ont réalisé des simulations (3) et suggèrent que ces cartes ne sont que peu affectées par les marqueurs en distorsion de ségrégation.

Globalement, ces résultats suggèrent que l'histoire évolutive des populations utilisées pour les croisements *in vitro* peut fortement affecter les cartes génétiques. Ce résultat peut apparaître trivial pourtant il est souvent tentant d'utiliser des individus issus de populations divergentes pour maximiser le nombre de polymorphismes et recréer une carte avec une forte densité en marqueurs. Toutefois, si cette divergence a autorisé la mise en place de barrières d'isolement post-zygotiques, la carte génétique ne sera pas totalement représentative de la proximité génétique entre les locus. Le biais est d'ailleurs proportionnel au nombre d'incompatibilités et donc dépendant du temps de divergence. Par exemple, la carte génétique de référence au sein de notre laboratoire est un croisement entre une souche *virRvi6* et *avrRvi6*. De manière intéressante, notre carte génétique est la plus dense en marqueurs génétiques actuellement disponible chez *V. inaequalis*, toutefois, elle comprend aussi un nombre plus important de groupes de liaisons (4-6), ce qui peut résulter de l'expression de ces incompatibilités car les autres cartes ont été réalisées en utilisant des croisements de souches *avrRvi6* x *avrRvi6*. Ces résultats suggèrent donc que les cartes génétiques doivent être réalisées en utilisant des individus partageant la même histoire évolutive, même si ce choix peut réduire le nombre de polymorphismes.

2.2 Génotypage des outliers en populations naturelles

2.2.1 Déséquilibre de liaison entre les SNPs proches des outliers

Des matrices de déséquilibre de liaison (DL) ont été générées sur l'ensemble des comparaisons par paires de locus SNPs générés des régions présentant les 4 outliers. Le déséquilibre de liaison a été estimée par calcul du r^2 entre toutes les paires de locus i et j [(7-9) et Dillmann, communication personnelle]. Le DL a été étudié pour les populations naturelles dans la population hybride de P32 (A, figure 3) et la population *avrRvi6* de La Folletière (B, Figure 3). La population de P32 a été considérée comme une seule et même population pour obtenir des effectifs suffisants. De la même façon, la population *virRvi6* a été écartée du jeu de donnée car cette population présentait trop de données manquantes. La majorité des déséquilibres de liaison concerne des locus situés à l'intérieur d'une zone génomique. Ceci suggère que le déséquilibre de liaison se maintient sur une distance suffisamment grande pour que deux marqueurs proches soient en DL. A l'intérieur de chaque zone, le DL semble toutefois supérieur dans la population hybride à celui observé dans la population parentale. Cela semble montrer que la sélection contre les hybrides contribue à une augmentation du DL dans ces zones par rapport aux populations parentales.

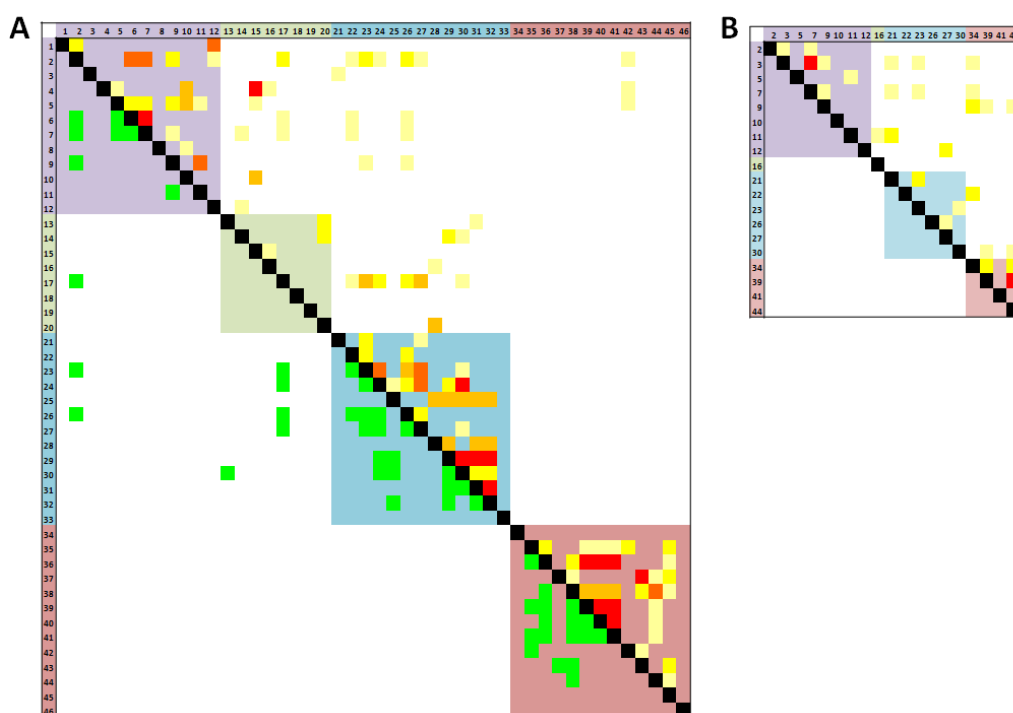


Figure 3 : Déséquilibre de liaison (DL) entre les SNPs génotypés sur les 4 zones génomiques différentes sur la population hybride de P32 (A) et sur la population *avrRvi6* de La Folletière (B). Les locus sont classés en fonction de leur position sur le scaffold. Les numéros des locus correspondent aux régions de PCTMCT6 (en violet, locus 1-12), PTGMGC27 (en vert, locus 13-20), PTGMT15 (en bleu, locus 21-33) et PAAMCT8 (en rose, locus 34-46). Les comparaisons entre les locus d'une même région sont signalées par un carré de couleur correspondant. La demi-matrice supérieure présente le déséquilibre de liaison entre couple de marqueurs (R^2) tandis que la demi-matrice inférieure reporte les comparaisons avec un R^2 significatif. * La significativité a été vérifiée par un test exact de Fisher au seuil 0.05 après 500 permutations du jeu de donnée puis correction de type FDR ('False Discovery Rate'). Remarquons l'absence de locus significativement en déséquilibre de liaison pour la population *avrRvi6* de La Folletière.

Un des résultats majeurs est que des SNPs présents dans différentes régions génomiques peuvent aussi être en DL entre eux. Ce résultat est observé à la fois dans la population hybride P32 et dans la population avirulente de la Folletière. Toutefois, seules les comparaisons entre des locus de la population hybride sont significatives après correction pour les tests multiples (en vert, figure 3). Ce résultat suggère que la contre-sélection des hybrides génère du DL entre les régions impliquées dans les barrières endogènes d'isolement reproductif. Ce résultat n'est toutefois vérifié que pour les SNPs dans les zones de PCTMCT6, PTGMGC27 et PTGMTC15. Aucun SNP de ces trois régions ne semble en fort déséquilibre de liaison avec les locus de la région de PAAMCT8. L'intensité de la contre-sélection dans cette zone est probablement trop faible pour observer un DL significatif avec les autres régions. En effet, le niveau de F_{ST} autour de PAAMCT8 est le plus faible des 4 zones étudiées (cf § Fig. 3 article 1). Enfin notons que si l'absence de DL dans la population parentale *avrRvi6* suit bien les attendus associés à l'accumulation de DMI, ce résultat devra être confirmée avec un échantillonnage plus exhaustif des allèles (ici $n \geq 16$) accompagné d'un génotypage de l'ensemble des locus pour valider la non-significativité du test.

2.2.2 *Génotypage dans de nouvelles populations*

Au cours de ce travail de thèse, une approche de génotypage des locus détectés comme outliers dans d'autres populations naturelles a été initiée. L'objectif de ce travail est de vérifier que les fréquences des allèles des barrières génétiques ne sont pas totalement corrélées aux habitats comme attendus dans le cadre de barrières exogènes (typiquement si cet allèle conférait l'adaptation à *Rvi6*).

2.2.2.1 Matériel et méthodes

La quantité de données ici présentée n'est absolument pas représentative du coût et du temps nécessaire pour les acquérir. En effet, pour cette approche le typage a été réalisé par la détection du polymorphisme causal par un jeu d'amorces spécifiques au locus outlier et cette approche a été réalisée par différents jeux de PCR classique¹².

2.2.2.1.1 Séquençage du produit AFLP

L'isolement des produits AFLP amplifiés a d'abord été initié par migration des produits d'amplifications spécifiques par migration sur gel d'acrylamide après révélation au nitrate d'argent (10). L'isolement de ces bandes et leur séquençage a parfois permis la localisation génomique de ces bandes par Blast en utilisant les 3 génomes disponibles de *V. inaequalis* comme référence. Toutefois, les produits de séquençage ont souvent été d'une qualité insuffisante pour localiser avec succès les bandes isolées à cause de la contamination des bandes d'intérêt par des bandes de taille voisine. Nous avons ensuite utilisé une autre approche basée sur la simulation d'une AFLP *in silico* à partir des données des 3 génomes de *V. inaequalis* avec le logiciel ISIF (11). Cette approche a permis l'identification de fragments

¹² Le choix de l'approche utilisée explique probablement la faible quantité des données acquises mais les limites associées à cette approche semble désormais pouvoir être contrée (voir résultats & discussions).

AFLP présents dans les génomes et de taille voisine du polymorphisme observé sur les outliers.

Après localisation de la bande AFLP, de nouvelles amorces spécifiques sur les régions flanquantes des outliers ont été définies afin de séquencer le locus pour des souches amplifiant la bande AFLP ou ne l'amplifiant pas. Les alignements locaux de ces séquences ont permis l'identification du polymorphisme causal. Deux grands types de polymorphismes ont été identifiés : soit la présence d'une mutation sur le site de restriction du fragment AFLP, soit l'existence d'indels dans les séquences.

2.2.2.1.2 Génotypage des fragments

Pour chacun des deux allèles détectés, une amorce spécifique de l'allèle a été utilisée dans un sens et une amorce consensus est utilisée dans l'autre. Les deux couples d'allèles spécifiques définis permettant une amplification de fragments de taille différente sur gel d'agarose ($\Delta > 100$ pb). La taille du fragment amplifié permet alors de connaître les allèles en présence au locus causal du polymorphisme AFLP. Cette approche a été réalisée sur 7 outliers différents. Plusieurs facteurs ont fortement limité la puissance de cette approche. Les deux facteurs qui se sont révélés les plus handicapants sont (i) le caractère multiallélique des locus amplifiés, (ii) le nombre de bandes non-spécifiques amplifiées. La quantité de données manquantes s'est traduit par des effectifs souvent trop faibles (< 10) pour calculer des fréquences alléliques et exclure l'existence de biais d'échantillonnage.

2.2.2.2 Résultats et discussion

Le déterminisme des deux locus ici présentés (Figure 4) étant bi-allélique, seul l'allèle type virulent dans les populations de P32 est représenté.

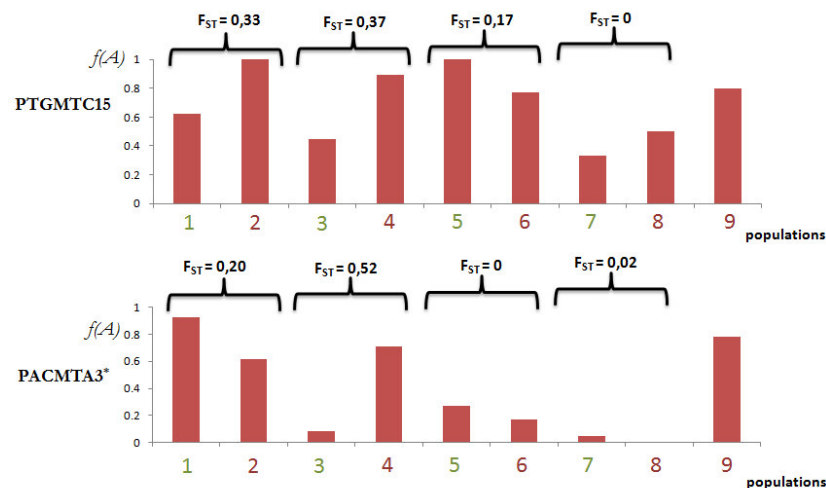


Figure 4 : Fréquence de l'allèle type virRvi6 en P32 au sein de différentes populations isolées virRvi6 et avrRvi6 et F_{ST} au locus entre les populations isolées d'un même site géographique. Les sites notés en vert et en rouge correspondent respectivement aux habitats Rvi6 et non-Rvi6. Les populations 1 et 2 sont isolées du verger normand de la Folletière (site 1 dans l'article précédent), les populations 3 et 4 correspondent au verger angevin P32 (site 2 dans l'article précédent), les populations 5 et 6 sont isolées du verger normand d'Yvrandes, les populations 7 et 8 correspondent à un verger échantillonné au Danemark et la « population » 9 correspond aux accessions de *M. x floribunda*. * Notons que PACMTA3 n'est pas détecté comme outlier dans l'article 3 mais que ce locus présente un fort niveau de différenciation très fort en P32 et est cartographié à proximité de l'outlier PTGMGC27.

Les fréquences alléliques de cet allèle sont très variables entre les 9 populations échantillonnées. Les fréquences alléliques ne sont pas corrélées aux habitats. *A contrario*, la variabilité observée est tout aussi grande en comparant les fréquences des habitats *Rvi6* (en rouge) et les habitats non-*Rvi6* (en vert). Par exemple, l'allèle « type virulent » au locus PTGMTC15 est fixé ($f(A)=1$) dans la population avr*Rvi6* d'Yvrandes (5) alors que ce locus est polymorphe chez les vir*Rvi6* ($f(A)=0.77$). Le même type de résultat est observé au locus PACMTA3 (locus voisin de l'outlier PTGMGC27 et présentant un fort F_{ST} en 2009), l'allèle associé à la virulence en P32 (site 2), est plutôt associé à l'avirulence à La Folletière (site 1). Dans les deux autres sites (Yvrandes et Danemark), les fréquences alléliques sont faibles dans les deux populations ($f(A)<0.3$). La conséquence majeure est que les F_{ST} associés à ces outliers est très variable entre les zones géographiques d'échantillonnage. En conséquence, La variabilité observée ne correspond aux pas attendus de barrières exogènes mais plus vraisemblablement à ceux attendus pour des barrières d'isolement post-zygotiques endogènes.

2.3 Barrières post-zygotiques de type DMI et gènes candidats

2.3.1 Hypothèses alternatives aux DMI

Les éléments transposables sont généralement considérées comme un moteur important de l'extension des génomes fongiques. Ces génomes sont, en effet, caractérisés par des régions riches en GC, porteuses de l'essentiel des gènes du génome (12); et des régions riches en AT appelées isochores. Ces isochores sont produites par un mécanisme de défense contre l'invasion de transposons et serait une zone particulièrement riche en petite protéine sécrétée de type effecteur¹³ (13).

Les barrières post-zygotiques pourraient être particulièrement accumulées dans les régions répétées et les isochores, compatibles avec les modèles de dépression hybrides dues à des remaniements chromosomiques et des duplications de gènes selon le modèle d'Oka (*cf* §1.2.2.1.2). Toutefois, les outliers localisés sur le génome (les 4 outliers caractérisés par un patron d'autostop dans l'article précédent) se situent dans des régions présentant un GC% élevé. De plus, la caractérisation des patrons d'autostop ne suggère pas que les pics de différenciation seraient centrés sur des isochores adjacents mais *a contrario* marquent aussi des régions riches en GC et des régions denses en gènes. L'existence même de patrons d'autostop suggère que cette hypothèse est peu vraisemblable avec des isochores non-détectées par la détection automatisée réalisée par l'URGI. En effet, ces régions se comportent théoriquement comme des blocs recombinants (13), ce qui maintiendrait la différenciation sur de grandes zones génomiques. Ces résultats confortent l'hypothèse d'une accumulation de barrières post-zygotiques endogènes de type Dobzhansky-Muller.

2.3.2 Caractérisations des zones génomiques et gènes candidats

La caractérisation fine des patrons d'autostop, réalisée dans l'article précédent,

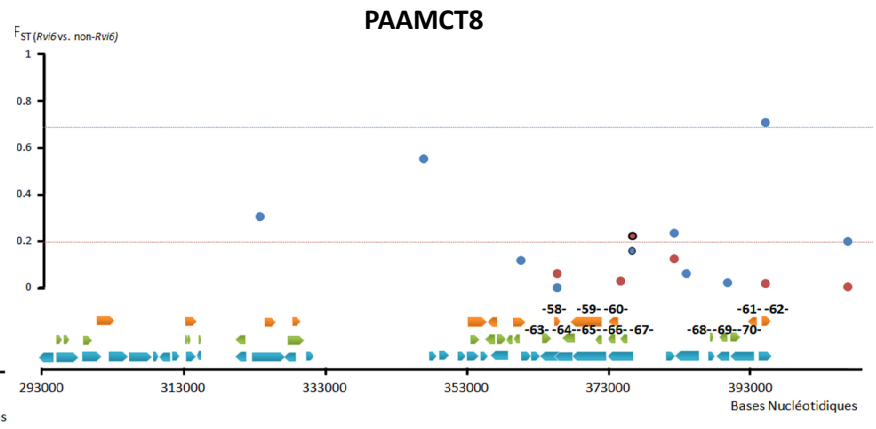
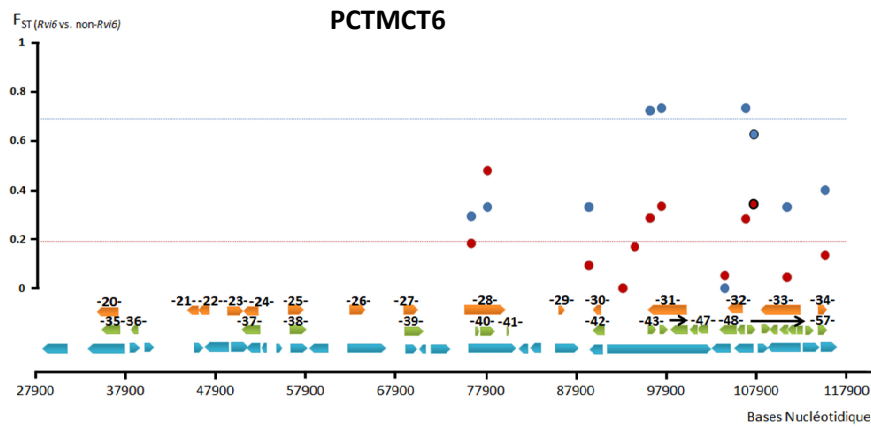
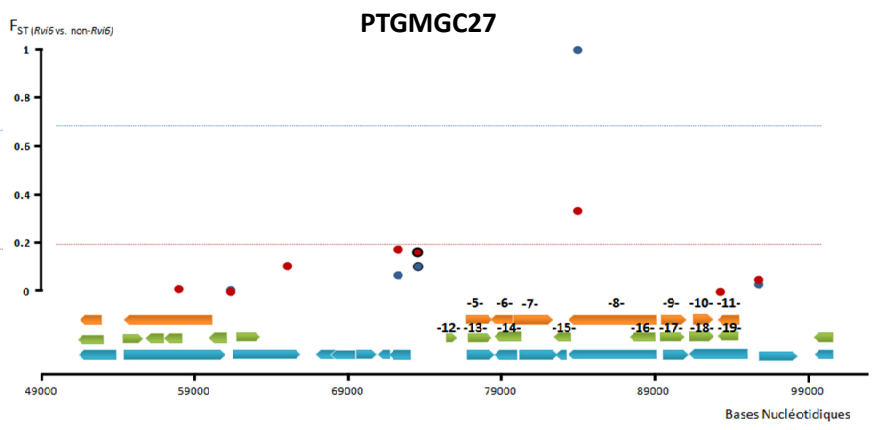
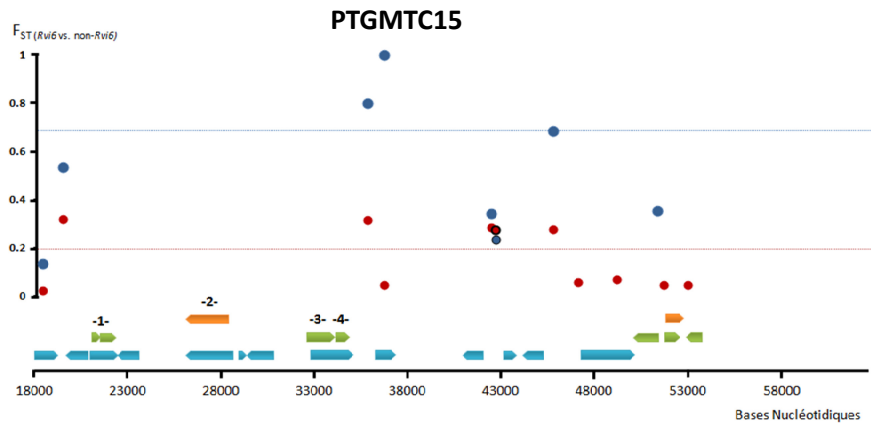
¹³ Chez *V. inaequalis*, les premières analyses réalisées par Jonathan Kreplak à l'URGI suggèrent toutefois que les structures en isochores AT de *V. inaequalis* présentent peu de gènes codant pour des petites protéines sécrétées.

suggère que la sélection n'influence que de petites zones génomiques. Cette hypothèse semble en effet vraisemblable car le déséquilibre de liaison aux locus « outliers » semble s'étendre sur des distances faibles sur le génome (cf Figure 3 ; Fig 3 article 1). La lecture des patrons d'autostop n'est en revanche pas facilitée par l'existence de polymorphismes récents. En effet, puisque les polymorphismes SNP utilisés ont été définis par alignements entre 3 génomes avirulents, les SNPs sont soit des polymorphismes ancestraux soit dérivés. Si les polymorphismes sont dérivés, ils sont apparus plus récemment et attendus en faible fréquence à l'intérieur de la population *avrRvi6*, ce qui ne permet pas d'observer un fort niveau de différenciation à ces locus. Des réseaux d'haplotypes ont été construits pour vérifier que les SNP présentant un fort niveau de F_{st} sont ancestraux et les SNP à faible F_{st} sont dérivés (données non-présentées). Cet attendu a été observé, les SNPs présentant un fort niveau de différenciation sont à la base du réseau alors que les polymorphismes se situent sur les branches périphériques. En absence de prédiction des patrons d'autostop en situation de contact secondaire, le meilleur moyen d'appréhender les patrons d'autostop est donc de s'intéresser aux locus présentant une plus forte différenciation.

En outre, la caractérisation fine des patrons d'autostop permet de discriminer la région de forte différenciation et de rechercher les gènes ou les promoteurs de ces derniers dans cette zone. Ces zones sont plutôt caractérisées par une forte densité en gènes (Figure 5), ce qui est assez représentatif des régions riches en GC chez les champignons (13). Pour chacun des quatre patrons décrits, les gènes de ces zones sont reportés (Figure 5 et Tableau 1), ce qui correspond soit à des gènes ayant une similitude de séquence avec des séquences présentes dans les bases de données Swissprot, soit à des séquences exprimées (EST) d'ARN messager produites par *V. inaequalis* en condition contrôlées. En effet, pour cette dernière catégorie, des bases de données d'ESTs ont été construites sous 4 conditions : une condition de croissance *in vitro* dans des conditions non-stressantes, des conditions de croissance en milieu carencé (stress azoté ou carboné) et en situation d'interaction avec son hôte *in planta* (Qin, communication personnelle). Les EST dans ces zones à fortes différenciation ne sont pas des EST spécifiques à l'interaction avec l'hôte mais plutôt des EST produits de manière constitutive ou exprimé dans des conditions de stress abiotiques relativement génériques. Les EST présents dans ces zones ont été blastés contre les bases de données bioinformatique du centre Américain de biotechnologie (NCBI). La plupart de ces EST non-soutenus par des similitudes de séquence avec des protéines présentes dans les bases de données Swissprot ne renvoient pas vers des gènes dont la fonction est connue, ce qui ne permet généralement pas de déterminer les fonctions et localisation spatiales des protéines. Cette première recherche permet un crible des gènes potentiellement responsables des barrières post-zygotiques. Notons que les gènes détectés se situent essentiellement dans des familles de gènes ubiquistes (kinases, oxydo-réductases, etc...) et non des protéines connues pour contribuer à la pathogénicité. Outre ces gènes connus, un grand nombre de gènes sont prédits dans ces zones par un algorithme de recherche automatisée de séquences codantes (ici Eugène). Conservons à l'esprit que (i) ces gènes prédits sont autant d'autres gènes putatifs pouvant être responsables de l'isolement reproductif observé; et que, (ii) plus largement encore, rien ne prouve que les DMI ne soient pas présentes dans les régions intergéniques de ces zones, aussi réduites soient-elles. Si ces patrons de différenciation sont simples pour les locus PTGMT15, PTGMGC27 et PAAMCT8 où un seul pic de F_{st} est présent, notons que le patron observé autour du locus

PCTMCT6 est lui plus complexe. Dans cette région, le Fst en P32 augmente graduellement entre les positions 110 000 et 70 000, ce qui suggère que le locus causal est situé en amont. Inversement, le pic de Fst basé sur les populations de La Folletière semble centré sur la zone 95000-110000). En conséquence, pour cette zone, l'ensemble des gènes présents ont été annotés (Tableau 1).

1. S. Neuenschwander, F. Hospital, F. Guillaume, J. Goudet, *Bioinformatics* 24, 1552 (Jul 1, 2008).
2. K. Semagn et al., *Genome* 49, 545 (May, 2006).
3. C. A. Hackett, L. B. Broadfoot, *Heredity (Edinb)* 90, 33 (Jan, 2003).
4. G. A. Brogini et al., *Fungal Genet Biol* 48, 166 (Feb, 2011).
5. J. M. Celton, A. Christoffels, D. J. Sargent, X. Xu, D. J. Rees, *BMC Biol* 8, 155 (2010).
6. X. Xu et al., *BMC Res Notes* 2, 163 (2009).
7. C. Edelist, Université Paris XI (2007).
8. P. W. Hedrick, *Genetics* 117, 331 (Oct, 1987).
9. T. Maruyama, in *Molecular Evolution, Protein Polymorphism and the Neutral Theory* M. Kimura, Ed. (Japan Societies Press, Tokyo, 1982) pp. 151-166.
10. S. Creste, A. Tulmann Neto, A. Figueira, *Plant Molecular Biology Reporter* 19, 299 (2001).
11. M. Paris, L. Despres, *Methods Mol Biol* 888, 55 (2010).
12. S. Raffaele, S. Kamoun, *Nat Rev Microbiol* 10, 417 (Jun, 2012).
13. T. Rouxel et al., *Nat Commun* 2, 202 (2011).



Région	locus	position	base de données	e-value*	Fonction putative du gène	Organisme*
PTGMTC15	-2-	26429-28022	Pubmed	1.00E-92	Perméase - acides aminés dicarboxyliques	<i>Saccharomyces</i>
PTGMTC15	-3-	32756-33589	NCBI	1.00E-35	oxydo-réductase FAD-dépendante	<i>Phaeosphaeria nodorum</i>
PTGMTC15	-4-	33775-34642	NCBI	2.00E-17	oxydo-réductase FAD-dépendante	<i>Mycosphaerella graminicola</i>
PTGMGC27	-5-	77180-78499	Pubmed	3.00E-22	Transporteur d'auxine (efflux)	<i>Schizosaccharomyces pombe</i>
PTGMGC27	-6- & -14-	79018-79635	Pubmed & NCBI	3.00E-50	Phosphatase - HK-MTPenyl-P phosphatase	<i>Saccharomyces cerevisiae</i>
PTGMGC27	-7-	80295-81946	Pubmed	1.00E-52	She9 - stabilité de la membrane interne de la mitochondrie	<i>Aspergillus niger</i>
PTGMGC27	-8- & -16-	83978-88818	Pubmed	3.00E-84 & 1.00E-75	Sérine-Thréonine Kinase	<i>Pichia pastoris</i>
PTGMGC27	-9- & -17-	89240-90079	Pubmed	1E-50 & 1.00E-125	Mannosyltransferase	<i>Candida albicans</i>
PTGMGC27	-10- & -18-	90473-90737	Pubmed	9.00E-26 & 1.00E-125	Mannosyltransferase	<i>Candida albicans</i>
PTGMGC27	-11- & -19-	91319-92410	Pubmed & NCBI	1.00E-58	Quinone oxydo-réductase	<i>Escherichia coli</i>
PTGMGC27	-12-	75796-76086	NCBI	1.00E-05	Ubiquinone oxydo-réductase NADH dépendante?	<i>Phaeosphaeria nodorum</i>
PCTMCT6	-20-	32627-33796	Pubmed	3.00E-61	Phospholipase	<i>Schizosaccharomyces pombe</i>
PCTMCT6	-21- & -22-	43943-44533	Pubmed	8.00E-26	Homologue de la protéine membranaire RIC1	<i>Homo sapiens</i>
PCTMCT6	-23-	47691-49104	Pubmed	2.00E-76	Guanosine-2'-O-methyltransferase trm11	<i>Schizosaccharomyces pombe</i>
PCTMCT6	-24- & -37-	49401-50276	Pubmed & NCBI	5.00E-21	putative 3-oxoacylréductase	<i>Bacillus subtilis</i>
PCTMCT6	-25- & -38-	55301-56407	Pubmed & NCBI	8.00E-78	Protéine mitochondriale LEU5 - transporteur Coenzyme A	<i>Saccharomyces cerevisiae</i>
PCTMCT6	-26-	62040-63275	Pubmed	2.00E-35	Metalloprotéase - homologue mep20	<i>Aspergillus flavus</i>
PCTMCT6	-27- & -39-	68266-69273	Pubmed & NCBI	2.00E-107 & 1.00E-168	Putative dioxygénase (tfdA)	<i>Schizosaccharomyces pombe</i>
PCTMCT6	-28- & -40-	76219-79833	Pubmed & NCBI	0	MAP kinase kinase kinase - wis4	<i>Schizosaccharomyces pombe</i>
PCTMCT6	-29-	85893-86489	Pubmed	4.00E-34	remodelage de la chromatine - sous-unité RSC7	<i>Saccharomyces cerevisiae</i>

Tableau 1 : Fonction putative des gènes candidats présentant une similitude de séquences avec des gènes codant des protéines référencées dans la base de donnée swissprot (blastx) ou similitude de séquence des ESTs sur les bases de données nucléotidiques du NCBI. * Les valeurs d'e-value et les organismes sont ceux des meilleurs « hits », la fonction peut par contre être déduite des résultats des blasts suivants. Seuls les résultats présentant une similitude sont reportés.

Région	locus	position	base de données	e-value*	Fonction putative du gène	Organisme*
PCTMCT6	-30- & -42-	88821-89528	Pubmed & NCBI	8.00E-79	Sous-unité 3A - facteur d'épissage	<i>Bos taurus</i>
PCTMCT6	-31-	96414-99836	Pubmed	2.00E-24	Adhésion	<i>Staphylococcus haemolyticus</i>
PCTMCT6	-32- & -48- & -49-	105007-105744	Pubmed	4.00E-61	Arginine N-méthyltransférase	<i>Mus Musculus</i>
PCTMCT6	-33- & -53-	109535-112030	Pubmed	1.00E-162	RMD11 - régulateur de la taille cellulaire & méiose	<i>Phaeosphaeria nodorum</i>
PCTMCT6	-34- & -57-	115377-116015	Pubmed	2.00E-46	Sérine-Thréonine kinase - Caséine Kinase II	<i>Bos taurus</i>
PCTMCT6	-35-	32399-33202	NCBI	2.00E-36	Protéine hypothétique - fonction non-décrite	<i>Phaeosphaeria nodorum</i>
PCTMCT6	-44- & -45-	98605-99246	NCBI	5.00E-20	Chaîne lourde de la myosine de classe II	<i>Pyrenophora tritici-repentis</i>
PCTMCT6	-46- & -47-	101531-103033	NCBI	9.00E-12	Protéine hypothétique - fonction non-décrite	<i>Pyrenophora teres</i>
PCTMCT6	-50-	107183-107920	NCBI	2.00E-09	Protéine hypothétique - fonction non-décrite	<i>Pyrenophora tritici-repentis</i>
PCTMCT6	-55-	112246-113113	NCBI	4.00E-32	Facteur de transcription spécifique des champignons	<i>Leptosphaeria maculans</i>
PAAMCT8	-58-	365774-366148	Pubmed	1.00E-22	Assemblage des protéines fer soufre - Matrice mitochondriale	<i>Schizosaccharomyces pombe</i>
PAAMCT8	-59-	368103-371002	Pubmed	0.00	Protéine inconnue - activité enzymatique	<i>Saccharomyces cerevisiae</i>
PAAMCT8	-60- & -66- & -67-	372835-373665	Pubmed	2.00E-69	MAP kinase kinase kinase A (Erk kinase A) - putative Ste11	<i>Dictyostelium discoideum</i>
PAAMCT8	-61-	391390-393158	Pubmed	6.00E-24	Butanediol Déshydrogénase	<i>Bacillus subtilis</i>
PAAMCT8	-62-	394892-396120	Pubmed	2.00E-84	Alcool déshydrogénase NADP dépendante	<i>Saccharomyces cerevisiae</i>
PAAMCT8	-63-	362514-363402	NCBI	1.00E-07	Protéine hypothétique - fonction non-décrite	<i>Phaeosphaeria nodorum</i>
PAAMCT8	-64-	367343-368211	NCBI	8.00E-35	Putative glycosidase	<i>Aspergillus nidulans</i>
PAAMCT8	-65-	370290-371174	NCBI	4.00E-94	Putative glycosidase	<i>Pyrenophora teres</i>
PAAMCT8	-69-	388925-390126	NCBI	2.00E-13	Protéine hypothétique - fonction non-décrite	<i>Phaeosphaeria nodorum</i>

Tableau 1: Fonction putative des gènes candidats (suite)

RESULTATS CHAPITRE 2

Evolution of pathogenic traits through secondary contact with gene flow between divergent pathogen populations: an unsuspected threat for agro-ecosystems.

**Evolution of pathogenic traits through secondary contact with gene flow
between divergent pathogen populations: an unsuspected threat
for agro-ecosystems.**

Thibault Leroy, Valérie Caffier, Bruno Le Cam & Christophe Lemaire

INTRODUCTION

Pathogenicity can be divided into virulence determined by the qualitative component of infection of a host (the Gene-for-Gene model [1,2], and aggressiveness, a quantitative component of pathogenicity [3]. Gene for gene relationship relies on a molecular interaction involving a major resistance gene (R) in plants which encode a protein that directly or indirectly triggers immunity against pathogens possessing the matching avirulence (Avr) gene [4]. Evidences for the Gene-for-Gene relationship led plant breeders to mainly focus on the way to overcome virulence by introgression of major resistances. Most plant breeding programs are based on R-genes identified in the wild and then introgressed into related agronomic species by breeders. However, resistant cultivars have frequently been rendered ineffective by rapid resistance breakdowns due to emergence of virulent strains. These emergences of virulence leading to ‘defeated’ R genes are common in agriculture (for an extensive review see [5]). Examples involve a large diversity of pathogens such as virus [6], bacteria [7], Oomycota [8] or fungi [9-11].

Rapid evolutionary changes from avirulence to virulence are attributed to the loss of function of the *Avr* gene within the agrosystems. Overcoming of an R gene is often described as a consequence of the spreading of a virulent mutant that invades resistant cultivars from infected susceptible ones (*i.e.* hosts lacking this R-gene). Those adaptations at the *Avr* locus might originate from non-synonymous point mutations, insertions of transposable elements or deletions of the *Avr* gene (*e.g.* [12-17]). Subsequently, we will refer to this scenario as the “mutation to virulence” model. Under this general model, Van der Plank [3] hypothesizes that a mutation to virulence is associated with a fitness cost on susceptible hosts. Indeed, fitness costs are required to explain how polymorphism at the avirulence locus can be maintained in

pathogens exposed to hosts carrying R-genes. Empirical examples of this evolutionary trade-off between virulence and fitness have been widely reported (*e.g.* [18-22] see also [8,19]).

Alternatively to this ‘mutation to virulence’ model, a recent empirical model has been proposed (Lemaire *et al.* in prep). Following these authors, loss of resistance would not be linked to new mutations within the agro-ecosystems but from migration of preexisting virulent populations from the wild that we will name “alien population”. After the emergence of the virulent population into the agrosystem, three different evolutionary scenarios are possible. First, the alien population is not divergent enough from the resident one. Then as no genetic barriers are expected, free gene flow occurs between the two populations. Recombination in the secondary contact might freely homogenize genomes. Secondly, if hybridization is totally prevented by strong reproductive barriers, *i.e.* when the two populations are too divergent, then gene flow is fully prevented leading to the maintenance of population genome integrities. Third, if populations are not completely reproductively isolated, viable hybrids between the two populations might be produced but less frequently than expected by chance because genetic barriers to gene flow partially contribute to the maintenance of the population differentiation in several genomic regions while regions devoid of barriers might freely homogenize. In their empirical evidence, Leroy *et al.* (in prep) validate this latter scenario. In this case, recombination can therefore be viewed as a diversifying force since it boosts pathogen variability and redistributed genotypes in the adaptive landscape, thus allowing discoveries of new optimal adaptive peaks. More broadly, recombination of pathogens populations might promote adaptive changes since pathogenicity can be largely promoted by such events of introgressions. Well-known examples of those deleterious consequences of recombination for animals or plants involve bacteria [23-24] or viruses [25].

In fungal or Oomycota pathogens, hybridization is generally related to interspecific hybridization events leading to an extension of host ranges [26-30]. This evolutionary change may be largely promoted by fusions of nuclei in dycaryotic cells (parasexuality), allowing changes in ploidy level and chromosomal rearrangements [31]. However recombination between divergent populations (*i.e.* at the intraspecific level) received much less support in fungi. Nevertheless, gene flow between divergent populations might generate complex negative or positive epistatic interactions in hybrids (*e.g.* [32-33]). Positive or negative epistasis depends on the relative fitness of the parents compared with the fitness of hybrids. Negative epistatic interaction involves Dobzhansky-Muller incompatibilities (DMI, [34-35]) between pairs of alleles from two divergent genomes which induce an outbreeding depression

in hybrids. Recent evidence for DMI in fungi has been demonstrated (Leroy *et al.* in prep) and suggests that recombinant genotypes are expected to be less fit than the parental populations (Leroy *et al.* in prep; [36-37]). However, evidence for DMI does not exclude that some hybrids could be fitter than their parents even in the parental environment (e.g. [38-39]). Consequently, a crucial question is the extent to which recombination between divergent fungal populations, such as resident pathogen populations and invasive exotic populations from the wild, may alter disease epidemiology and may constitute an emergence risk of fitter genetic combinations in parental environments. To address this question, we investigated the evolution of aggressiveness in a secondary contact zone with gene flow between two divergent populations of the fungal pathogen *Venturia inaequalis*.

V. inaequalis is a relevant model to investigate the epidemiological consequences of a secondary contact with gene flow between divergent fungal populations. First, the origin and evolutionary history of the *virRvi6* population - an emerging virulent population toward the *Rvi6* resistance genes- has been assessed recently (Lemaire *et al.* in prep). Interestingly, *virRvi6* population did not arise from a new mutation in agro-ecosystems but from migrations of preexisting virulent strain from the wild. Those *virRvi6* strains initially infected the wild progenitor of the *Rvi6* resistance gene (*M. x floribunda*) and subsequently migrated to the orchards after the deployment of the *Rvi6* cultivars (Lemaire *et al.* in prep). Second, a secondary contact zone with gene flow between *avrRvi6* and *virRvi6* of *V. inaequalis* had been detected in an apple orchard (Leroy *et al.* in prep). Given that *V. inaequalis* overwinters through pseudothecia (sexual fruiting bodies), sexual recombination between *avrRvi6* and *virRvi6* populations generated a hybrid swarm in this orchard by recurrent backcrossing with the two parental populations. Thereafter, intrinsic expressions of DMIs within the pathogen genome were identified thus explaining maintenance of the differentiation between the two populations (Leroy *et al.* in prep). This finding that is inconsistent with classical ecological divergence hypotheses in pathogens indicate that divergence time between the two populations was long enough to fix independent alleles in each population to produce such DMIs in hybrids. We thus hypothesize that divergence time might also favor independent adaptations within each population that could be exchanged in secondary contact zones and contribute to major evolutionary changes in pathogens.

The aim of the present study was to test if intraspecific hybridization between divergent *avrRvi6* and *virRvi6* populations promoted strong evolutionary changes for virulence or aggressiveness in *V. inaequalis*. Using pathogenicity tests, we first evaluated the risk of an accumulation of aggressiveness-related traits in hybrids between the resident

avirulent population and the emerging virulent one. Second, we investigated the temporal dynamic of invasion of the *virRvi6* allele in the secondary contact zone from 2008 to 2012.

RESULTS

Higher aggressiveness of mixed ancestry hybrids

Thirty *V. inaequalis* strains sampled in the secondary contact zone in Leroy *et al.* (in prep) were selected to maximize the range of hybrid indexes (ranged from 0.17 to 0.68). Strains were inoculated on two susceptible cultivars: *cv.* Gala and *cv.* Top Red. For all analyses, we used two summary measures of aggressiveness for each strain, the AUDPC, the area under the disease-progress curve based on the number of lesions, and severity, the percentage of the leaf infected by the pathogen. These thirty isolates were divided into two groups based on the phenotype at the *avrRvi6* locus of the pathogen: one avirulent toward *Rvi6* (*avrRvi6*) and the other virulent (*virRvi6*). Based on this two groups, no significant differences in normalized AUDPC were observed on Gala ($W=6612.5$, $p=0.5812$) but the normalized AUDPC on *cv.* Top Red were significantly different between the two groups ($W=7622$, $p=0.005$) which reveals a slightly higher aggressiveness for the *avrRvi6* samples. Note that a good correlation between normalized AUDPC and severity were observed on Gala ($R^2=0.45$, $b=0.50$, $p<1.10^{-15}$) and on Top Red ($R^2=0.69$, $b=0.44$, $p<1.10^{-15}$) indicating that both measures of aggressiveness might not be considered as independent. Indeed, severity is not exclusively dependent of the lesion growth but is also dependent of the number of lesions produced.

For all subsequent analysis, we used hybrid index values as a proxy of the proportion of the genome inherited from the *virRvi6* genome to address if recombination leads to evolutionary changes in hybrids.

First, strains with extreme values of hybrid index (0.174 and 0.682) express medium values of aggressiveness on our greenhouse inoculation trial. Alternatively, least aggressive strains were observed for a hybrid index of 0.417 and 0.44 for the normalized AUDPC and 0.417 and 0.583 for the normalized severity on *cv.* Gala and *cv.* TopRed respectively (A and B, Figure 1). At the opposite, the most aggressive strains had a hybrid index of 0.5 on *cv.* Gala and 0.48 on *cv.* Top Red for both measures of aggressiveness (A and B, Figure 1). Taken together, these results support that strains with a quasi-equal proportion of genome inherited from both populations expresses a wider range of aggressiveness values thus suggesting that

recombination contributes to an evolution of the overall aggressiveness in hybrids.

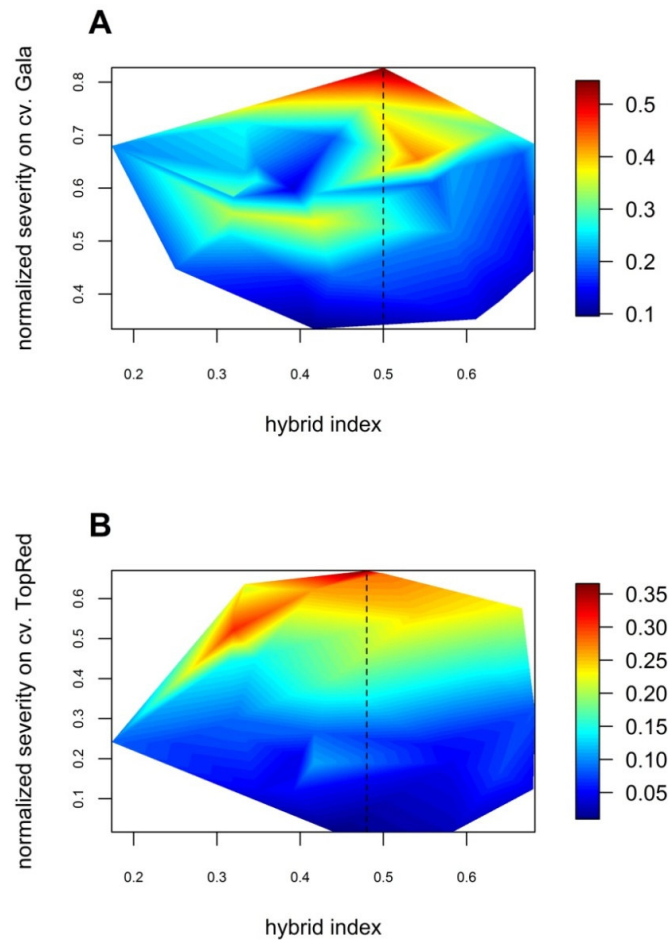


Fig 1: Linear interpolation of the aggressiveness of strains in function of their hybrid index on cv. Gala (A) and Top Red (B) respectively. Color bar indicates the range in normalized AUDPC. For each strain, value of the normalized AUDPC and severity is the median value over replicates. For each cultivar, dashed lines indicate the value of the hybrid index for the most aggressive strain. Note that variation in aggressiveness is maximal for hybrid index values between 0.4 and 0.6.

Second, the thirty isolates were divided into two other groups based on the main proportion of the ancestral inherited genome (mainly *avrRvi6* for hybrid index <0.5 or mainly *virRvi6* for hybrid index ≥ 0.5) (A, Fig 2). No significant regression between the normalized AUDPC and the hybrid index was observed for mainly *avrRvi6* strains neither on *cv. Gala* ($b=0.027$, $p=0.869$) nor on *cv. Top Red* ($b=0.250$, $p=0.154$). At the opposite, highly significant regressions were observed for mainly *virRvi6* strains whatever the cultivar ($b=-1.601$, $p<1.10^{-6}$ for *Gala*; $b=-0.651$, $p<0.002$ for *Top Red*). Negative values of the slope (b) suggest that recombination with the *avrRvi6* genome increases the aggressiveness. The same behavior was observed for regressions based on the normalized severity with a non significant

regression for mainly *avrRvi6* strains ($b=-0.128$, $p=0.583$ on *cv. Gala*; $b=0.135$, $p=0.658$ on *cv. Top Red*) but significant for mainly *virRvi6* strains ($b=-0.994$, $p=0.022$ on *cv. Gala*; $b=-1.194$, $p=0.015$ on *cv. Top Red*) (B, Fig 2).

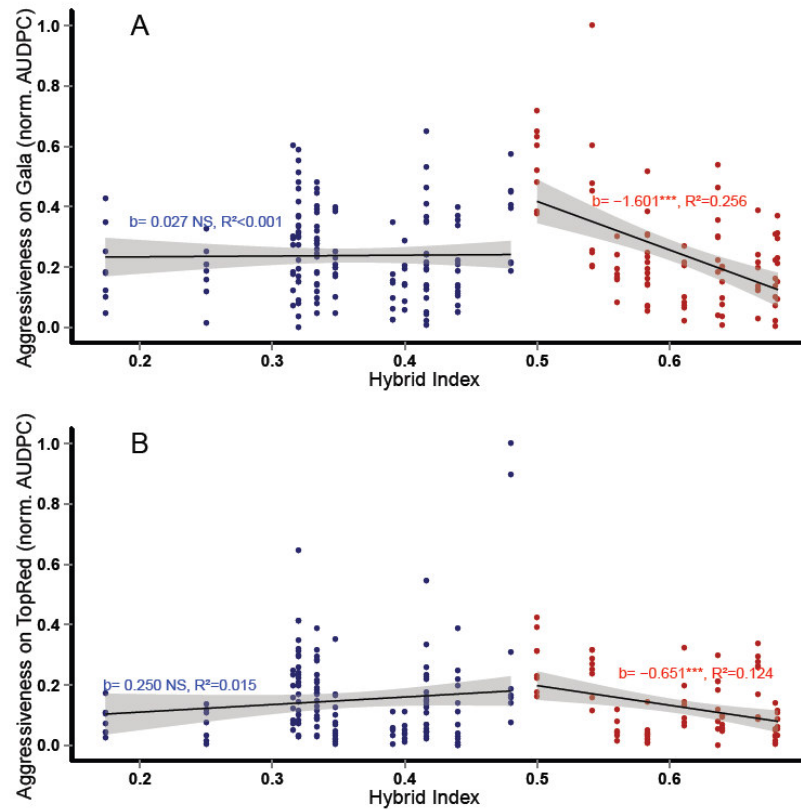


Fig 2: Linear regressions between the normalized AUDPC on Gala and the hybrid index on *cv. Gala* (A) and *Top Red* (B). The linear regression is shown in black and the 95% confidence interval as grey area. No significant correlation were observed for strain expressing a main *avrRvi6* genetic background (blue dots) while strains with a main *virRvi6* genetic background (red dots) is negatively correlated with the hybrid index values for both hosts.

Detection of aggressiveness-related loci through association mapping

A significant genetic structure between *avrRvi6* and *virRvi6* populations was detected based on polymorphisms at 969 loci (Weir & Cockerham's $F_{ST} = 0.046$, $p<0.001$). Subsequently, to avoid 'spurious' detection, first association tests were performed under STRAT to take into account the population structure between the two populations. Given that severity and AUDPC were not independent, we exclusively performed association analyses with the normalized AUDPC. Using STRAT, 104 and 31 loci were significantly associated with the normalized AUDPC on *cv. Gala* and *cv. Top Red*, respectively (Table 1). Furthermore, given that STRAT uses discrete classes of phenotype while normalized AUDPC

is continuously distributed (see material and methods for details); Mann-Whitney U-tests were performed independently for each locus to test if the two AFLP alleles had significant differences in aggressiveness. Compared to STRAT, U-tests detected a lower number of loci, 31 and 30 on cv. Gala and Top Red respectively (Table 1). Majority of loci significantly associated with the phenotype were exclusively detected in a single association mapping test thus suggesting variability between host cultivars. To avoid risks of false positives, only loci significantly detected by both methods were considered associated with phenotype. Ten loci were detected by both methods for aggressiveness on cv. Gala and only one locus was associated with aggressiveness on cv. Top Red (Table 1). Given the low number of loci associated with the normalized AUDPC on cv. Top Red, subsequent analyses were exclusively performed for loci associated with the normalized AUDPC on cv. Gala.

Table 1: Number of loci significantly associated with the normalized AUDPC on cv. Gala and cv. Top Red depending of STRAT or U-test methods. Note that in subsequent analyses a score is then calculated based on the proportion of allele contributing to the normalized AUDPC on cv. Gala and detected by both independent methods (10 loci).

Methods	Number of loci associated with		
	norm. AUDPC on cv. Gala	norm. AUDPC on cv. Top Red	<i>both norm. AUDPC</i>
STRAT	104 (0.107)	42 (0.043)	11 (0.011)
U-test	31 (0.032)	30 (0.031)	6 (0.006)
<i>both methods</i>	10 (0.010)	1 (0.001)	none

Score: a good indicator of the overall strain aggressiveness

For each strain, a score based on the number of alleles contributing to the aggressiveness over the ten loci were computed. First, in accordance with expectations, a significant linear regression between the score and the normalized AUDPC on cv. Gala was detected ($b=0.243$, $p<0.001$). Similar conclusions were also obtained with the normalized severity on cv. Gala ($b=0.185$, $p=0.032$) thus suggesting that these score based on loci associated to the normalized AUDPC might be a good indicator of the overall aggressiveness (Fig 3). More surprisingly, the score perform quite well for the variability observed on cv. Top Red with a significant slope between the normalized AUDPC and the score ($b=0.136$, $p=0.042$) and a close to significance regression for the normalized severity ($b=0.234$, $p=0.099$).

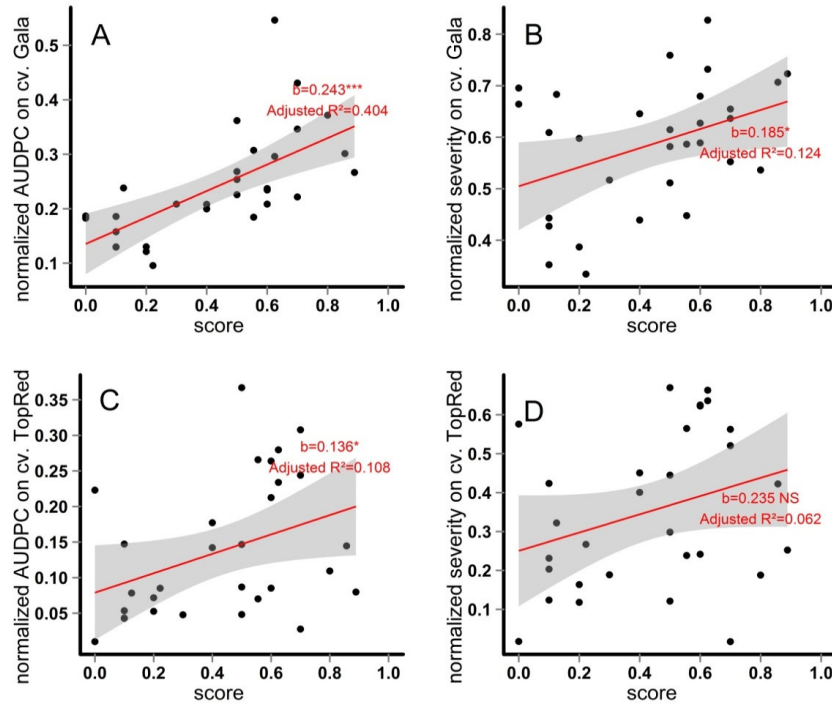


Fig 3: Significance of linear relationships between score and the overall increase of aggressiveness. Linear regressions were performed using normalized AUDPC (A, B) or normalized severity (C, D) on cv. Gala (A, C) and Top Red (B, D) respectively. The score correspond to the ratio of the number of alleles contributing to the normalized AUDPC on cv. Gala over the number of called allele at the ten loci detected by the association genetic approach (see main text). The linear regression is shown in black and the 95% confidence interval as grey area. Significances are reported with stars: * $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.**

After ranging the values of score into 4 distinct classes (from “---“ to “+++”) (Supplementary Data Fig 1), individuals belonging to the lowest class of score “---“ had a significant lower aggressiveness than individuals belonging to the two highest classes (“+-“ and “+++”) for cv. Gala. The same behavior was observed for the normalized AUDPC on cv. Top Red, with a significant lower aggressiveness of the “---“ class compared to the “+-“ class but not significantly compared to the upper class “+++” (B, Supplementary Data Fig 1).

Alleles contributing to the aggressiveness origin from both genomes

For the subset of the ten loci, a putative origin of alleles was investigated based on the contribution of alleles to the first axis of a CA in the secondary contact zone without gene flow (as mentioned in Leroy *et al.*, in prep). Therefore, a part of alleles is more likely inherited from *virRvi6* than the *avrRvi6*, while others are more likely inherited from *avrRvi6* than from *virRvi6*. Using the ten loci dataset, 6 more likely originated from the *avrRvi6* gene pool and 4 from the *virRvi6* one. This result needs to be cautiously interpreted because the

ancestry of each allele was indirectly evaluated through the contribution of the first axis of the CA assuming that the overall number of polymorphic alleles in each population is equal (see material and methods and Leroy *et al.* in prep). Nevertheless, this result suggests that genomic regions contributing to a higher aggressiveness probably originated from both genomes and therefore that new fitter combination arose by recombination.

Temporal dynamic of invasion of the Rvi6 virulence locus

After investigating the introgression of the aggressiveness-related traits, we studied the temporal dynamic of the virulence locus on *Rvi6* hosts and non-*Rvi6* hosts. Given that the *avirulent-Rvi6* allele that interacts with the *Rvi6* R-gene is unknown, we used the ‘virulent alleles’ at two outliers in strong linkage disequilibrium with the *Rvi6* virulence locus (Leroy *et al.*, in prep) as a proxy of the frequency of the *virRvi6* allele. Here, for each outlier, the virulent allele corresponded to the most frequent allele at each outlier loci in the *virRvi6* subpopulation in 2009 (for details see Leroy *et al.*, in prep). Using a three-year sampling effort over the last five years (*i.e.* 2008, 2009 and 2012) in the secondary contact zone, we investigated changes in allele frequencies at the two outliers. First, we observed an increase of virulent allele in the same proportion at the two loci suggesting that this behavior was likely to be representative of different genomic regions in strong linkage disequilibrium with the *virRvi6* locus (table 2). Interestingly, the virulent allele globally increased in the same way on both habitat (*Rvi6* host or non-*Rvi6* host). However, statistical power was not sufficient enough to fully reject that the frequency of the virulent allele remained stable over years in both habitats (table 2).

Table 2: Changes in allele frequencies at two outliers in strong linkage disequilibrium with the *virRvi6* locus during a four-year period. Virulent alleles correspond to most frequent alleles detected in *Rvi6* population in 2009. Significances of the increase of the virulence were assessed by Akaike Information Criterion (AIC), comparing a null model (no regression) and a linear model (linear regression). Best fit model were reported as the lower AIC value.

locus	habitat	frequency allele virulent			Akaike Information Criterion (AIC)		
		2008	2009	2012	no reg.	linear reg.	best fit model
PTGMGC27	<i>Rvi6</i>	0.3793	0.8214	0.9286	3.894	2.632	linear regression
	non- <i>Rvi6</i>	0.3333	0.7241	0.8519	3.445	1.705	linear regression
PTGMTC15	<i>Rvi6</i>	0.75	0.9167	0.8293	-3.609	-1.748	no regression
	non- <i>Rvi6</i>	0.4375	0.6364	0.6889	-0.825	-2.207	linear regression

DISCUSSION

Gene for gene relationships are one of the most striking examples of the strength of the host selective pressures exerted on the pathogen populations [40-43]. Breakdown of resistance gene is widely described as occurring through new mutations to virulence in pathogens infecting crops. Recently, we described a new scenario where virulence emerged through the accidental releasing of preexisting virulent population from the wild in the *V. inaequalis*/*Malus* pathosystem. In this scenario, orchards are secondary contact zones between the two divergent avirulent and virulent populations. As gene flow has been detected in one orchard, we explore here the consequences of this intraspecific hybridization for the evolution of both virulence and aggressiveness-related traits.

Evidence for outbreeding depression

In evolution, the main consequence of hybridization is founded on the expectation that recombinants between two divergent genotypes are on average less fit than the parental populations. Using inoculation trials, we assessed the fitness of *V. inaequalis* hybrids by inoculating them on two susceptible commercial apple cultivars. As expected, recombination generated several unfitted hybrids, with a lower aggressiveness than parental populations on the two different cultivars. Outbreeding depressions might origin from two different ways, ecological selection or selection against hybrids, which both contribute to the maintenance of species integrity. First, outbreeding depression might result from hybrids that are on average less ecologically adapted on both parental environments. F1 hybrids are disadvantageous in nature and introgression between parental populations is prevented. Second, outbreeding depression might also result from intrinsic DMIs which correspond to negative epistasis between pairs (or more) of interacting alleles originating from the two different genomes. This evidence for outbreeding depression corroborates a previous detection of DMI in the same secondary contact zone (Leroy *et al.* in prep). Here, it remains unclear if this outbreeding depression in hybrids might also partially be caused by the non-*Rvi6* plant maladaptation or exclusively from DMIs. Anyway, the lower fitness of several hybrids is in accordance with classic models of divergence where genome integrity is preserved by the lower fitness of hybrids [44].

Introgression of aggressiveness-related traits

Several fit hybrids were detected as individuals expressing medium hybrid index values suggesting that hybridization might lead to new genetic combinations that are even fitter in parental environments than parents. Indeed, based on an association genetic approach, we detected several loci contributing to a better fitness which probably originated from both parental genomes (4 from the *virRvi6*, 6 from the *avrRvi6* genetic background). This introgression suggests how hybridization between both genomes can boost genetic variance and favor new genetic combinations that can be then tested by natural selection. In theoretical models on the contribution of hybridization in evolution, fit hybrids are a general outcome (e.g. [38]). Indeed, if the first hybrid generation produces viable and fertile hybrids, the backcrossing with parental population generates a hybrid swarm with a wide variability of genetic and phenotypic traits from both parents. This burst can sporadically help hybrids to cross valleys and attain new higher adaptive peaks, *i.e.* generating rare genetic combinations that are fitter than their parents even in the parental environment (e.g. [39]). Keeping this expectation in mind and given the large fecundity generally assumed for pathogens, some fitter hybrids are expected to be produced quite abundantly in pathogens, and undoubtedly more easily than in many other organisms.

Invasion dynamic of *virRvi6* strains

Based on two different loci in strong linkage disequilibrium with the phenotype for avirulence/virulence, a rapid increase of the proportion of the ‘virulent alleles’ was observed on *Rvi6* hosts but also on non-*Rvi6* hosts. This observation is consistent with previous report of the unexpected high proportion of virulent strains on non-*Rvi6* hosts in this secondary contact zone in 2009 (47%, Leroy *et al.* in prep). Following theoretical models, in absence of a fitness cost, a rapid fixation of the virulence allele in both susceptible and resistant host population is expected because virulent strains are intrinsically advantaged by their ability to infect both hosts (e.g. [45]). Alternatively, if fitness cost exists, polymorphisms for the avirulence/virulence might be maintained over time. Here hybrids with a high hybrid index (close to parental *virRvi6* genome) seems to express a lower aggressiveness than hybrids with a low hybrid index suggesting that parental *virRvi6* population might be less adapted to non-*Rvi6* hosts or that virulent allele is costly. Lower aggressiveness in *virRvi6* than in *avrRvi6* population has also been reported for non-mixing population [46], we suggested that the lower aggressiveness might be easily compensated by the introgression of advantageous genetic loci from the *avrRvi6* genome. Given that frequency of virulent alleles seems to invade both populations, the cost of aggressiveness previously reported in *virRvi6* populations

was more likely due to a maladaptation to the Gala cultivar than to intrinsic cost of virulence. Consequently, fitter individuals are expected to be both virulent at the *avrRvi6* locus and harbouring an avirulent genetic background (Leroy *et al.* in prep) and thus expected to move away from this secondary contact zone to invade surroundings orchards.

Concluding remarks: is the cure worse than the disease?

Considerable effort are undertaken to introgress resistance, essentially monogenically inherited R-genes, from the wild relatives to agronomic plants. The contribution of adaptation to the R-genes breakdowns is widely admitted. Under this scenario, mutation to virulence promotes the emergence of new virulence from infected susceptible cultivars within the agro-ecosystems. Alternatively, R-genes might break down by a second way where virulence originates from the wild. Following this second model, virulence preexists on the wild (e.g. the progenitor of the resistance) and is accidentally released in agro-ecosystems after the deployment of resistant cultivars (Lemaire *et al.* in prep). Those two kinds of breakdown of R-genes similarly results in a complete efficiency loss of the plant resistance. However, we demonstrated that hybridization between divergent virulent and avirulent populations in secondary contact zones might also boost pathogen variability and lead to the introgression of advantageous loci from both genomes (virulence, aggressiveness-related traits). This second consequence induces much more deleterious effects since it not exclusively affects new resistant cultivars but the resident host population as a whole. Given that susceptible hosts generally represent much larger agricultural areas, this increase of aggressiveness through intraspecific hybridization is an unsuspected serious threat for crops. We strongly encourage plant pathologists to rigorously take into account the wild plant pathogen diversity and plant breeders to extensively investigate the occurrence of preexisting virulent populations on the wild progenitor before deploying new resistant cultivars in agricultural landscapes.

Materials and methods

Plant material

Gala and TopRed cultivars (respectively X4410 and X4003 from the INRA depository, France) were chosen as reference cultivars since their large presence in worldwide commercial apple orchards. In 2009, the cv. Gala was the second apple cultivar grown in the United States and in the European Union. Gala carries at least two QTL (quantitative trait

loci) of resistance to scab [47] but is often used as the susceptible reference standard in pathogenicity studies [48]. Top Red is a cultivar closely related to cv. Red Delicious, the most growing cultivar in the US in 2009. Gala and TopRed were grafted on rootstock MM106 (Pirard, France) and grown in the same conditions than previously reported [46]. The youngest fully expanded leaf was labelled F₀, the second youngest leaf below F₀ was labelled F₁, while the non-fully expanded leaf upper than F₀ was labeled as F₋₁. At the time of inoculation, each plant had 8–13 leaves.

Isolates of *Venturia inaequalis*

Isolates of *V. inaequalis* were collected in 2009 in the secondary contact zone between *avrRvi6* and *virRvi6* populations in Angers, France (Leroy *et al.* in prep). On the 58 previously published isolates, we chose 30 isolates (Supplementary Table 1). Choice of isolates was first based on their avirulence trait towards the major resistance gene *Rvi6*, 15 *avrRvi6* and 15 *virRvi6* strains were selected. Within each group, we maximized the available range in hybrid index values for each *avrRvi6* and *virRvi6* strains (Supplementary Table 1, see Leroy *et al.* for details). Hybrid index corresponds to the proportion of allele inherited from the *virRvi6* genome at 25 diagnostics loci (for details see Leroy *et al.* in prep).

Inoculations

Spore suspensions were prepared from the scabbed leaves and adjusted to a concentration of 80,000–95,000 spores mL⁻¹ (Coulter counter, Beckman, 6.2 μm < mean diameter < 12.8 μm). Three droplets of the suspension were deposited on malt agar to measure the germination rate, which varied from 59.5% to 90% depending on the isolate. The concentration of viable inoculum was calculated as the spore concentration of the inoculum suspension multiplied by its germination percentage. For each isolate, the suspension was applied with a sprayer (mainly on the upper side of the leaves) to sixteen plants (eight of cv. Gala and eight of cv. TopRed) enclosed in a plastic chamber to avoid cross-contamination of isolates during inoculation. Forty four hours after inoculation, the leaves were kept wet with an atomizer and the plants kept in darkness at 18°C. The plants were incubated under 16 h of light at 18°C and relative humidity fluctuating between 70% and 90%. During incubation the leaves remained dry, which prevented secondary infections.

Measurements of aggressiveness

Aggressiveness was evaluated on leaves F₋₁, F₀ and F₁, which were the most susceptible to scab, and developed full and well-defined lesions within 3 weeks after inoculation (for details see [46]).

Lesions density

Seven days after inoculation, each plant was inspected to confirm the absence of symptoms. From the 9th to the 16th day, the number of new lesions on each leaf was recorded every 2 or 3 days. Each lesion was marked with a permanent marker to avoid recounting. At the end of the experiment, each leaf was scanned and leaf area was obtained using the ImageJ software. Lesions density was subsequently computed as a number of lesions per squared centimeter of leaf. Lesion density was evaluated for leaves F₋₁, F₀ and F₁. Lesion density was expressed as the total number of lesions per square centimeter of leaf area. AUDPC is a disease-progress curve and was based on the kinetic of the lesion density over the experiment.

Severity

Sixteen days after inoculation, severity was evaluated on leaves F₋₁, F₀ and F₁ on a 0-7 scale, corresponding to the percentage of sporulating symptoms.

Data analysis

AUDPC and severity

For each leaf, the area under the disease progress curve (AUDPC) and the severity were used as a quantitative summary of the aggressiveness. Severity represents the percentage of the leaf area with sporulating symptoms. AUDPC and severity were performed independently for each leaf of each replicate; but subsequently, the values of AUDPC and severity were averaged over the three different leaves to obtain a single value for each replicate. Mean values of AUDPC and severity per replicate were then normalized between 0 and 1. When necessary, the mean value of the normalized value of AUDPC and severity over replicates were used.

All statistical analyses were performed using R version 2.13.0 [49]. R graphics were performed with the 'ggplot2' [50] and the 'akima' packages [51].

Association mapping

To identify locus associated with an increase of aggressiveness on apples, two different methods were used to perform an association mapping strategy. First an association mapping was performed under STRAT v. 1.1 [52]. STRAT enables association studies by taking into account the presence of population structure detected by STRUCTURE [53-56]. For STRUCTURE, the admixture mode was used and the Monte Carlo Markov Chain scheme was run for 1,500,000 iterations after an initial burn-in period of 500,000. We ran STRUCTURE for K=2 clusters assuming two groups related to *virRvi6* and *avrRvi6* (see Leroy et al. for details on the population structure of these strains). For each locus, STRAT tested the association between the locus and the phenotype. Here, the trait was binary coded assuming two phenotypes – zero for the 15 strain with the lowest value of AUDPC and one for the 15 strains with the highest values. Association was tested over 50,000 tests per locus for loci with at least 5 copies of each allele.

Second, given that association is not performed on continuous classes but over only two phenotypes using STRAT, we subsequently used a non-parametric Mann-Whitney U test to exclude loci that are likely to be artifacts where the difference in aggressiveness between the two alleles were not significant. In both STRAT and U tests, p values < 0.05 were considered statistically significant. Only loci both detected with STRAT and the Mann-Whitney U test were considered to be significantly associated with an increase of aggressiveness. Subsequently, a score was calculated for each sample as a ratio of the number of alleles contributing to an increase of aggressiveness over the total number of called alleles at these loci.

Putative origin of the alleles associated to the aggressiveness

As the native (*avrRvi6*) and the emerging (*virRvi6*) population did not exchange gene in some orchards, putative origin of each allele was investigated using AFLP genotyping assessed in these orchards (Lemaire *et al.* in prep; Leroy *et al.* in prep). We used the contribution of all AFLP alleles to the first axis of an AFC [57] as a proxy of the origin of the AFLP loci by analyzing the two populations from La Folletière (referred as site 1 in Leroy *et*

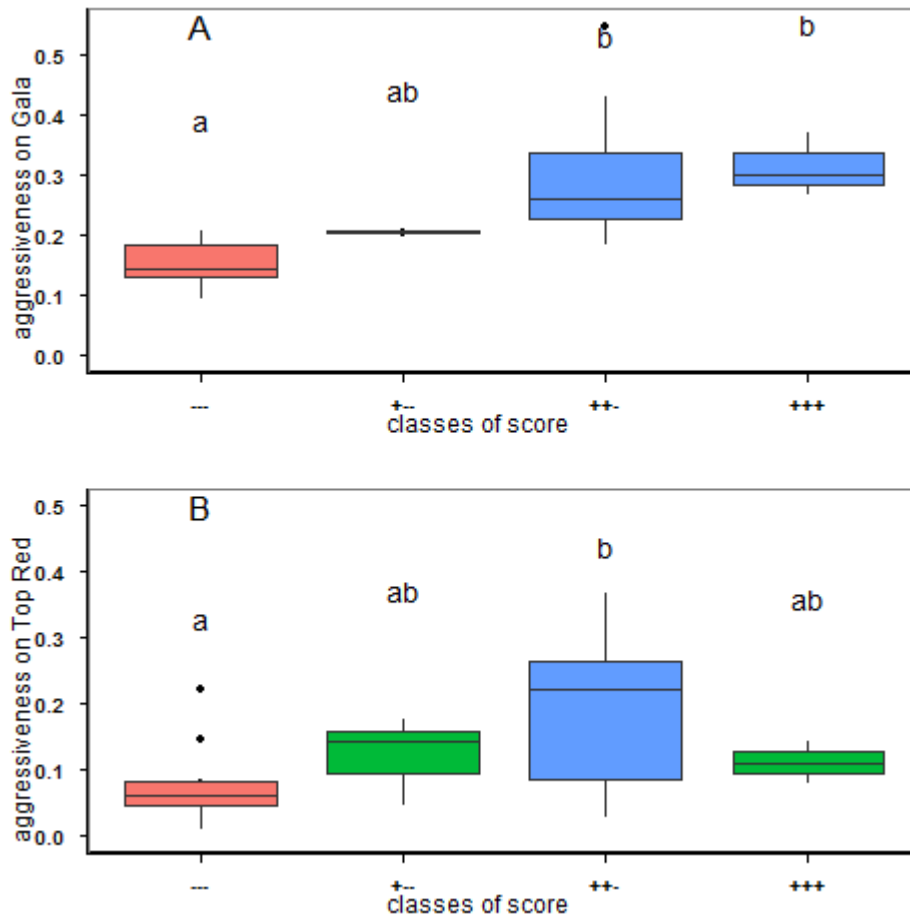
al. in prep). As the distribution of the contribution of the alleles to the first axis is not clearly bimodal, a median was used to cut off all alleles into two groups. Alleles that are related to a *virRvi6* origin were labeled as 1 while the alleles related to an *avrRvi6* were labeled as 0.

References

1. Flor HH (1955) Host-parasite interaction in flax rust - its genetics and other implications. *Phytopathology* 45: 680–685.
2. Flor HH (1971) Current status of the gene-for-gene concept. *Annual Review of Phytopathology* 9: 275-278.
3. Van der Plank JE (1963) *Plant diseases: epidemics and control*. New York: Academic Press. 349 pp.
4. Jones JD, Dangl JL (2006) The plant immune system. *Nature* 444: 323-329.
5. Parlevliet JE (2002) Durability of resistance against fungal, bacterial and viral pathogens; present situation. *124*: 147-156.
6. Qiu W, Moyer JW (1999) Tomato Spotted Wilt Tospovirus Adapts to the TSWV N Gene-Derived Resistance by Genome Reassortment. *Phytopathology* 89: 575-582.
7. Li ZK, Luo LJ, Mei HW, Paterson AH, Zhao XH, et al. (1999) A "defeated" rice resistance gene acts as a QTL against a virulent strain of *Xanthomonas oryzae* pv. *oryzae*. *Mol Gen Genet* 261: 58-63.
8. Peressotti E, Wiedemann-Merdinoglu S, Delmotte F, Bellin D, Di Gaspero G, et al. (2010) Breakdown of resistance to grapevine downy mildew upon limited deployment of a resistant variety. *BMC Plant Biol* 10: 147.
9. Bayles RA, Flath K, Hovmoller MS, de Vallavieille-Pope C (2000) Breakdown of the *Yr 17* resistance to yellow rust of wheat in northern Europe. *Agronomie* 20: 805-811.
10. de Vallavieille-Pope C, Ali S, Leconte M, Enjalbert J, Delos M, et al. (2011) Virulence dynamics and regional structuring of *Puccinia striiformis* f. sp. *tritici* in France between 1984 and 2009. *Plant Disease* 96: 131–140.
11. McIntosh RA, Brown GN (1997) Anticipatory breeding for resistance to rust diseases in wheat. *Annu Rev Phytopathol* 35: 311-326.
12. Catanzariti AM, Dodds PN, Lawrence GJ, Ayliffe MA, Ellis JG (2006) Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell* 18: 243-256.
13. Fudal I, Ross S, Brun H, Besnard AL, Ermel M, et al. (2009) Repeat-induced point mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans*. *Mol Plant Microbe Interact* 22: 932-941.
14. Gout L, Kuhn ML, Vincenot L, Bernard-Samain S, Cattolico L, et al. (2007) Genome structure impacts molecular evolution at the *AvrLm1* avirulence locus of the plant pathogen *Leptosphaeria maculans*. *Environ Microbiol* 9: 2978-2992.
15. Joosten MH, Vogelsang R, Cozijnsen TJ, Verberne MC, De Wit PJ (1997) The biotrophic fungus *Cladosporium fulvum* circumvents *Cf-4*-mediated resistance by producing unstable AVR4 elicitors. *Plant Cell* 9: 367-379.
16. Orbach MJ, Farrall L, Sweigard JA, Chumley FG, Valent B (2000) A telomeric avirulence gene determines efficacy for the rice blast resistance gene *Pi-ta*. *Plant Cell* 12: 2019-2032.
17. Stergiopoulos I, De Kock MJ, Lindhout P, De Wit PJ (2007) Allelic variation in the effector genes of the tomato pathogen *Cladosporium fulvum* reveals different modes of adaptive evolution. *Mol Plant Microbe Interact* 20: 1271-1283.
18. Huang Y-J, Evans N, Fitt BDL, Li Z-Q, Rouxel T, et al. (2006) Fitness cost associated with loss of the *AvrLm4* avirulence function in *Leptosphaeria maculans* (phoma stem canker of oilseed rape). In: Fitt BDL, Evans N, Howlett BJ, Cooke BM, editors.

- Sustainable strategies for managing *Brassica napus* (oilseed rape) resistance to *Leptosphaeria maculans* (phoma stem canker): Springer.
19. Leach JE, Vera Cruz CM, Bai J, Leung H (2001) Pathogen fitness penalty as a predictor of durability of disease resistance genes. *Annu Rev Phytopathol* 39: 187-224.
 20. Montarry J, Hamelin FM, Glais I, Corbi R, Andrivon D (2010) Fitness costs associated with unnecessary virulence factors and life history traits: evolutionary insights from the potato late blight pathogen *Phytophthora infestans*. *BMC Evol Biol* 10: 283.
 21. Thrall PH, Burdon JJ (2003) Evolution of virulence in a plant host-pathogen metapopulation. *Science* 299: 1735-1737.
 22. Vera Cruz CM, Bai J, Ona I, Leung H, Nelson RJ, et al. (2000) Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proc Natl Acad Sci U S A* 97: 13500-13505.
 23. Feil EJ, Spratt BG (2001) Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol* 55: 561-590.
 24. Goss EM, Kreitman M, Bergelson J (2005) Genetic diversity, recombination and cryptic clades in *Pseudomonas viridiflava* infecting natural populations of *Arabidopsis thaliana*. *Genetics* 169: 21-35.
 25. Lai MM (1992) Genetic recombination in RNA viruses. *Curr Top Microbiol Immunol* 176: 21-32.
 26. Brasier CM, Cooke DE, Duncan JM (1999) Origin of a new *Phytophthora* pathogen through interspecific hybridization. *Proc Natl Acad Sci U S A* 96: 5878-5883.
 27. Goss EM, Cardenas ME, Myers K, Forbes GA, Fry WE, et al. (2011) The plant pathogen *Phytophthora andina* emerged via hybridization of an unknown *Phytophthora* species and the Irish potato famine pathogen, *P. infestans*. *PLoS One* 6: e24543.
 28. Newcombe G, Stirling B, McDonald SK, and Bradshaw HD (2000) *Melampsora x columbiana*, a natural hybrid of *M. medusae* and *M. occidentalis*. *Mycol Res* 104.
 29. Stukenbrock EH, Christiansen FB, Hansen TT, Dutheil JY, Schierup MH (2012) Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proc Natl Acad Sci U S A* 109: 10954-10959.
 30. Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agroecosystems. *Annual Review of Phytopathology* 46: 75-100.
 31. Schardl CL, Craven KD (2003) Interspecific hybridization in plant-associated fungi and oomycetes: a review. *Mol Ecol* 12: 2861-2873.
 32. Burke JM, Arnold ML (2001) Genetics and the fitness of hybrids. *Annu Rev Genet* 35: 31-52.
 33. Burke JM, Carney SE, Arnold ML (1998) Hybrid fitness in the Louisiana irises: analysis of parental and F1 performance. *Evolution* 52: 37-43.
 34. Dobzhansky T (1937) *Genetics and the origin of species*. New York: Columbia University Press.
 35. Muller HJ (1942) Isolating mechanisms, evolution and temperature. *Biol Symp* 6: 71-125.
 36. Le Gac M, Hood ME, Giraud T (2007) Evolution of reproductive isolation within a parasitic fungal species complex. *Evolution* 61: 1781-1787.
 37. Sloan DB, Giraud T, Hood ME (2008) Maximized virulence in a sterilizing pathogen: the anther-smut fungus and its co-evolved hosts. *J Evol Biol* 21: 1544-1554.
 38. Barton NH (2001) Speciation. *Trends Ecol Evol* 16: 325.
 39. Rieseberg LH, Whitton J, Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* 152: 713-727.
 40. Giraud T (2006) Speciation in parasites: host switching does not automatically lead to allopatry. *Trends Parasitol* 22: 151-152; author reply 152-153.

41. Giraud T, Gladieux P, Gavrillets S (2010) Linking the emergence of fungal plant diseases with ecological speciation. *Trends Ecol Evol* 25: 387-395.
42. Kolmer JA (1997) Virulence dynamics and genetics of cereal rust populations in North America. In: Crute IR, Holub EB, Burdon JJ, editors. *The gene-for-gene relationship in plant-parasite interactions*. New York.: CAB International. pp. 139–156.
43. Slusarenko A (2001) *Annual Plant Reviews Vol. 4*. Sheffield: Sheffield Academic Press. 286 pp.
44. Coyne JA, Orr HA (2004) *Speciation*. Sunderland, Massachusetts: Sinauer Associates. 545 pp.
45. Sapoukhina N, Durel CE, Le Cam B (2009) Spatial deployment of gene-for-gene resistance governs evolution and spread of pathogen populations. *Theoretical Ecology* 2: 229-238.
46. Caffier V, Didelot F, Pumo B, Causeur D, Durel CE, et al. (2010) Aggressiveness of eight *Venturia inaequalis* isolates virulent or avirulent to the major resistance gene *Rvi6* on a non-*Rvi6* apple cultivar. *Plant Pathology* 59: 1072-1080.
47. Soufflet-Freslon V, Gianfranceschi L, Patocchi A, Durel CE (2008) Inheritance studies of apple scab resistance and identification of *Rvi14*, a new major gene that acts together with other broad-spectrum QTL. *Genome* 51: 657-667.
48. Bus V, Rikkerink E, Aldwinckle HS, Caffier V, Durel CE, et al. (2009) A proposal for the nomenclature of *Venturia inaequalis* races. *Acta Horticulturae* 814: 739-746.
49. R Development Core Team. R (2011) *A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria.
50. Wickham H (2009) *ggplot2: elegant graphics for data analysis*. New York: Springer.
51. Gebhardt A (2009) *akima: Interpolation of irregularly spaced data*. 0.5-7 ed.
52. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. *American Journal of Human Genetics* 67: 170-181.
53. 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.
54. Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* 7: 574-578.
55. Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 9: 1322-1332.
56. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
57. Daguin C, Bonhomme F, Borsa P (2001) The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism at locus mac-1. *Heredity* 86: 342-354.



Supplementary Data Fig 1: Aggressiveness of hybrid strains on cv. Gala (A) and cv. Top Red (B) depending of classes of score. The score correspond to the ratio of the number of alleles contributing to the normalized AUDPC on cv. Gala over the number of called allele at the ten loci detected by the association genetic approach (see text). Classes of score are labeled "--", "+-", "++-", and "+++" which correspond to scores from 0 to 0.25, 0.25 to 0.5, 0.5 to 0.75 and 0.75 to 1 respectively. Different letters indicate significant differences between classes (Tukey HSD with $\alpha=0.05$ performed after an ANOVA).

Supplementary Figure 1: strains used in this study. For each strain, hybrid index, score at the 10 loci associated with a higher aggressiveness on cv. Gala and four different measures of aggressiveness were reported (mean were based on up to eight inoculated plants).

Strain	Phenotype at the <i>Rvi6</i> locus	Hybrid Index	Score	Classes of score	Mean normalized AUDPC on cv. Gala	Mean normalized AUDPC on cv. Top Red	Mean normalized severity on cv. Gala	Mean normalized severity on cv. Top Red
09BCZ001	<i>avrRvi6</i>	0.440	0.400	--+	0.200	0.177	0.439	0.400
09BCZ008	<i>avrRvi6</i>	0.417	0.800	+++	0.372	0.109	0.536	0.188
09BCZ009	<i>avrRvi6</i>	0.417	0.222	---	0.096	0.085	0.334	0.267
09BCZ014	<i>avrRvi6</i>	0.333	0.889	+++	0.267	0.080	0.723	0.252
09BCZ015	<i>avrRvi6</i>	0.500	0.625	++-	0.546	0.280	0.827	0.663
09BCZ016	<i>avrRvi6</i>	0.480	0.500	++-	0.362	0.367	0.759	0.670
09BCZ017	<i>avrRvi6</i>	0.333	0.600	++-	0.234	0.212	0.627	0.625
09BCZ021	<i>avrRvi6</i>	0.440	0.000	---	0.187	0.010	0.695	0.017
09BCZ026	<i>avrRvi6</i>	0.174	0.600	++-	0.208	0.085	0.680	0.242
09BCZ029	<i>avrRvi6</i>	0.316	0.500	++-	0.254	0.147	0.511	0.445
09BCZ031	<i>avrRvi6</i>	0.333	0.625	++-	0.296	0.234	0.732	0.636
09BCZ032	<i>avrRvi6</i>	0.417	0.600	++-	0.237	0.264	0.589	0.622
09BCZ043	<i>avrRvi6</i>	0.320	0.700	++-	0.346	0.308	0.552	0.521
09BCZ048	<i>avrRvi6</i>	0.250	0.556	++-	0.184	0.070	0.448	0.238
09BCZ049	<i>avrRvi6</i>	0.400	0.100	---	0.158	0.043	0.427	0.203
09BCZ020	<i>virRvi6</i>	0.640	0.200	---	0.121	0.072	0.387	0.163
09BCZ022	<i>virRvi6</i>	0.348	0.400	--+	0.208	0.142	0.645	0.451
09BCZ039	<i>virRvi6</i>	0.320	0.500	++-	0.226	0.087	0.582	0.298
09BCZ046	<i>virRvi6</i>	0.636	0.857	+++	0.301	0.145	0.707	0.422
09BCZ064	<i>virRvi6</i>	0.680	0.100	---	0.129	0.043	0.443	0.124
09BCZ066	<i>virRvi6</i>	0.320	0.556	++-	0.307	0.266	0.586	0.565
09BCZ068	<i>virRvi6</i>	0.583	0.700	++-	0.221	0.028	0.636	0.017
09BCZ075	<i>virRvi6</i>	0.667	0.000	---	0.183	0.223	0.664	0.576
09BCZ086	<i>virRvi6</i>	0.542	0.700	++-	0.431	0.244	0.655	0.563
09BCZ089	<i>virRvi6</i>	0.682	0.125	---	0.238	0.078	0.683	0.322
09BCZ090	<i>virRvi6</i>	0.391	0.200	---	0.130	0.053	0.598	0.118
09BCZ092	<i>virRvi6</i>	0.560	0.100	---	0.186	0.054	0.609	0.231
09BCZ094	<i>virRvi6</i>	0.583	0.300	--+	0.208	0.048	0.517	0.189
09BCZ113	<i>virRvi6</i>	0.611	0.100	---	0.130	0.147	0.352	0.423
09BCZ117	<i>virRvi6</i>	0.348	0.500	++-	0.268	0.048	0.615	0.121

DISCUSSION

Le but de mes travaux de thèse était de mieux appréhender l'intensité de l'isolement reproductif entre les populations divergentes *avrRvi6* résidentes sur le pommier cultivé et les populations envahissantes *virRvi6* originaires de *M. x floribunda*. La mise en évidence d'un contact secondaire avec flux de gènes entre les deux populations est, en effet, une situation favorable à la compréhension de la nature des barrières génétiques impliquées dans le maintien de l'intégrité des deux génomes. Appliquée aux pathogènes, cette quantification de l'intensité du flux génique entre ces deux populations est nécessaire pour prédire les risques de changements évolutifs associé à l'introgession adaptative de gènes responsables de la pathogénie. Nous disposons maintenant de quelques éléments de réponse qui seront ici discutés.

1. Emergences à partir du compartiment non-cultivé

Le modèle d'émergence associé à une migration de souches virulentes préexistantes dans le compartiment non-cultivé tel que proposé par Lemaire *et al.* (en préparation) apporte de nouveaux attendus par rapport au modèle classique de l'émergence par mutation vers la virulence à l'intérieur des agrosystèmes. La différence majeure entre ces deux modèles d'émergence réside dans la divergence possible entre la population avirulente et la population virulente pendant la période d'allopatrie. L'existence d'une telle divergence favoriserait donc l'apparition de barrières génétiques contribuant au maintien de l'intégrité des génomes des populations divergentes. On peut alors s'attendre à ce que les conséquences épidémiologiques associées à une remise en contact secondaire des populations, comme la possibilité de recombinaison et/ou de transfert de gènes associés à la virulence ou l'agressivité, découlent de l'intensité et du nombre de ces barrières génétiques.

Avant de discuter en détail de l'importance et de la nature des barrières génétiques, notons que l'importance de ces travaux est très dépendante du caractère générique de ce type d'émergence chez les phytopathogènes. Pour l'heure et à ma connaissance, *V. inaequalis* est le seul pathogène pour lequel on a démontré que l'introduction d'une résistance à partir d'un hôte sauvage favorise l'émergence d'une population virulente infectant initialement le compartiment non-cultivé. L'absence d'exemple avéré similaire d'un tel processus chez d'autres champignons pathogènes peut provenir du fait qu'il est difficile à distinguer du modèle considéré comme étant *a priori* le plus probable où la virulence provient du pool d'individus avirulents. Les programmes d'amélioration végétale pour la résistance aux pathogènes utilisent pourtant régulièrement des gènes de résistance provenant du compartiment sauvage pour l'amélioration de la résistance des espèces cultivées. L'introgession de gènes d'hôtes sauvages peut alors conférer un caractère de résistance chez certains génotypes d'hôtes domestiques. Il est cependant difficile de généraliser ce caractère à l'ensemble de l'aire de répartition de l'hôte. En effet la résistance observée localement peut simplement résulter de l'absence de la population virulente correspondante du pathogène. Cette population serait alors uniquement présente dans la population de l'hôte sauvage d'où provient la résistance conformément aux hypothèses de coévolution généralement admises (p. ex. Hamilton, 1980 ; Bell, 1982 ; Bell & Maynard-Smith, 1987 ; Hamilton, 1990 ; Clay & Kover 1996). Par conséquent, nous proposons que ce modèle d'émergence par amélioration pour la résistance soit bien plus fréquent que généralement admis dans le processus de contournement des résistances dans les agrosystèmes. Des études plus approfondies permettront soit d'affirmer soit d'infirmer le caractère généraliste de ce type d'émergence.

La connaissance du modèle d'émergence des virulences est un préalable fondamental avant de formaliser les attendus sur les mécanismes possibles – écologiques ou non-écologiques - responsables de la diversification des populations de pathogènes. Les méthodes

ABC (Tavaré *et al.* 1997 ; Beaumont, 2002 ; Roux, 2010) offrent une possibilité simple et efficace de tester ces deux types de scénarios d'émergence. La difficulté réside dans le fait de connaître les hôtes potentiellement sources des populations virulentes et de pouvoir échantillonner sur ces hôtes dans le compartiment non-cultivé. L'interprétation des scans génomiques étant très dépendante du cadre conceptuel envisagé (Bierne *et al.* 2011), les résultats présentés au cours de cette thèse auraient été probablement interprétés d'une manière différente sous le scénario alternatif d'une mutation vers la virulence.

2. Contacts secondaires et importance des barrières post-zygotiques

2.1 Contact secondaire avec flux génique entre des populations *avrRvi6* et *virRvi6*

Les zones de contact secondaire sont de précieux laboratoires pour détecter les barrières génétiques d'isolement reproductif contribuant au maintien de l'intégrité des génomes (Hewitt 1988). Nos résultats confortent l'hypothèse que ces zones existent aussi chez les pathogènes et que celles-ci sont pertinentes pour étudier les processus de divergence des pathogènes. En effet, notre étude met en évidence un contact secondaire suivi de flux de gènes entre les populations *avrRvi6* et *virRvi6* de *V. inaequalis*. Pourtant les attendus théoriques sur les zones hybrides, notamment *via* la contribution de N. Barton et ses collaborateurs, ont été peu intégrés à la littérature des phytopathologistes. La cause majeure est une prédominance des hypothèses écologiques dans la diversification des pathogènes.

Une des questions intrigantes et pour l'heure non-élucidée de notre étude est la différence observée entre les contacts secondaires de populations infectant les pommiers à couteau (chapitre 1) et les pommiers à cidre (Guérin *et al.* 2007 ; Gladieux *et al.* 2011). En effet, dans ces derniers vergers, le contact secondaire ne semble pas aboutir à un flux génique entre les deux populations (au moins entre 2000 et 2005, Gladieux *et al.* 2011). Plusieurs hypothèses ont été proposées pour expliquer cette spécificité. D'abord, une asynchronie dans le développement de la maladie pourrait expliquer ce maintien de la différenciation. En effet, les pommiers à cidre sont plus tardifs que les pommiers à couteau. Or, les variétés *Rvi6* utilisées dans les vergers cidricoles présentent majoritairement un fond génétique de pommiers à couteau et non de pommier à cidre, elles présentent effectivement un retard d'une dizaine de jours dans la floraison. Toutefois, les libérations d'ascospores se prolongeant sur plusieurs mois au printemps, cette hypothèse apparaît difficile à soutenir. De plus, les variétés non-*Rvi6* de pomme à cidre sont plutôt constituées de cultivars anciens généralement considérées comme porteurs de nombreux gènes de résistance, ce qui pourrait contribuer à une plus forte spécialisation écologique des populations dans ces vergers voire à une exclusion mutuelle des deux populations. En effet, des tests d'inoculations croisés ont été réalisés au cours de la thèse de Fabien Guérin (2004). L'exclusion des habitats a été vérifiée en inoculant sept souches *virRvi6* isolées des vergers normands de Pont Audemer et Yvrandes sur 7 cultivars non-*Rvi6* de pommier à cidre, suggérant une exclusion mutuelle des populations *virRvi6* sur les hôtes non-*Rvi6*. En revanche, le génotypage réalisé sur les populations isolées en 2004 du site de La Folletière (site 1 dans le chapitre 1) indique un certain degré d'admixture entre *avrRvi6* et *virRvi6* sur plusieurs souches suggérant des possibilités de flux de gènes entre les deux populations. Au cours de l'été 2012, un nouvel échantillonnage a été réalisé dans l'un de ces vergers (La Folletière, « site 1 » dans le chapitre 1) et le génotypage de ces populations permettra de vérifier si la différenciation génétique entre *virRvi6* et *avrRvi6* est toujours maintenue ou si celle-ci a été érodée par la

recombinaison.

2.2 Barrières génétiques post-zygotiques endogènes chez les pathogènes

Les barrières détectées par notre approche de scan génomique ne confortent pas les hypothèses écologiques généralement admises telles que la prédominance du locus d'avirulence dans la rupture du flux génique. Certains auteurs suggèrent que le gène d'avirulence pourrait agir considéré comme un « caractère magique » capable d'isoler les populations avirulentes et virulentes en sympatrie (Giraud *et al.* 2010; Gladieux *et al.* 2011; Servedio *et al.* 2011). *A contrario*, les barrières génétiques révélées dans cette zone de contact semblent toutes de type post-zygotiques endogènes. En effet, aucun outlier cartographié ne semble être au voisinage du locus d'avirulence, ni véritablement lié au caractère de virulence. Ces barrières post-zygotiques endogènes semblent donc jouer un rôle dans le maintien de l'intégrité des génomes parentaux entre des populations divergentes de pathogènes, mais ce rôle est partiel puisque l'isolement reproductif est insuffisant pour observer une absence de flux génique sur la majorité du génome. Bien évidemment, notre scan génomique est basé sur un échantillonnage d'un nombre restreint de locus, ce qui ne permet pas une détection exhaustive de l'ensemble des barrières génétiques. De plus, l'utilisation de marqueurs anonymes sur le génome tels que les locus AFLP rend la localisation génomique de ces outliers particulièrement ardue, ce qui ne permet pas une étude complète de la localisation de l'ensemble des barrières génétiques et l'étude approfondie de leur environnement génomique. Il serait donc abusif de considérer que le locus d'avirulence n'est pas une barrière importante au flux génique entre les populations *avrRvi6* et *virRvi6*. L'isolement reproductif observé peut, en effet, être la résultante de l'accumulation d'un grand nombre d'incompatibilités génétiques et de un ou quelques locus de barrières exogènes. Toutefois, le gène d'avirulence n'agit pas comme un caractère magique car son effet est insuffisant pour empêcher la recombinaison sur la plus grande partie du génome. Les outliers correctement localisés sont présents sur des zones différentes du génome, ce qui, compte tenu du nombre d'outliers détecté suggère un nombre relativement important de barrières post-zygotiques. La non-prise en compte des barrières post-zygotiques endogènes contribuent alors à une surévaluation de la contribution des barrières pré-zygotiques exogènes. En conséquence, l'intensité de la barrière exogène n'est pas compatible avec des modèles de divergence sympatrique. Le maintien de la différenciation entre les populations avirulentes et virulentes en condition de faible flux génique n'est donc pas exclusivement due à la non-viabilité des migrants (habitats exclusifs) mais semblerait plutôt favorisée par l'existence de barrières post-zygotiques endogènes de type non-viabilité ou stérilité des hybrides. Enfin, sous l'hypothèse d'un modèle de mutation vers la virulence, l'exclusion des populations sur leurs hôtes nécessite une quasi non-viabilité écologique des individus virulents sur l'hôte sensible (un très fort coût de fitness associé à la virulence) afin que les populations virulentes se retrouvent incapables de se reproduire sur l'hôte sensible. Cette dernière condition semble peu compatible avec les coûts de fitness généralement observés en conditions contrôlées (Leach *et al.* 2001; Montarry *et al.* 2010; Peressotti *et al.* 2010). Les hypothèses de spécialisation écologique doivent être avancées avec prudence lorsque la diversification ne corrèle pas avec des attendus forts de divergence allopatrique, d'où l'importance de tester avec rigueur l'hypothèse d'une émergence à partir du compartiment non-cultivé.

Les barrières post-zygotiques endogènes détectées semblent corrélées avec les modèles proposés par Dobzhansky (1937) et Muller (1942). En effet, nos résultats démontrent que les outliers (et particulièrement les SNPs les plus proches du locus) sont en fort déséquilibre de liaison entre eux chez les hybrides (*in vitro* ou populations naturelles) alors que ces derniers

sont répartis sur des scaffolds distincts du génome. L'utilisation d'une approche de cartographie génétique *via* l'obtention de croisements *in vitro*, permettant de s'affranchir de toute sélection liée à l'hôte, a mis en évidence des distorsions de ségrégation sur certains locus outliers suggérant que le déséquilibre de liaison observé entre ces outliers résulte d'une forte contre-sélection des hybrides de première génération. En effet, cette contre-sélection est suffisamment forte pour générer une distorsion de ségrégation significative sur la F1 du croisement. La distorsion observée sur les outliers devra toutefois être conforté par des marqueurs autres des marqueurs AFLP. En effet on ne peut exclure que cette distorsion puisse résulter d'une ségrégation de deux locus polymorphes amplifiant une même taille de fragment et générant une ségrégation de type 3 :1 (p. ex. Cloutier *et al.* 1997 ; Ky *et al.* 2000). Néanmoins, étant donné le faible nombre de polymorphismes observés pour chacune des combinaisons AFLP amplifié sur la descendance ($n=7^{\pm 3,33}$), la probabilité d'une co-ségrégation de deux fragments polymorphismes sur deux des quatre outliers reste faible. Si cette contre-sélection est confirmée, ce résultat confirmerait aussi que les outliers détectés ne sont pas liés à des processus démographiques comme par un exemple un effet de surf. Cette hypothèse semble toutefois pouvoir être exclue par l'évidence de patrons d'autostop et d'autres locus à fort F_{st} pour 4 des 6 zones génomiques testées. Même en absence de distorsion de sélection, ce résultat ne contredit pas pour autant l'hypothèse des DMI car les conditions d'observation de ce type de distorsions sont plus stringentes chez les haploïdes que chez les diploïdes. En effet, une telle distorsion nécessite que la contre-sélection des hybrides soit fortement différente entre les deux combinaisons recombinantes possibles générant ces incompatibilités. En effet, chez un organisme haploïde, la première génération permet de générer 4 classes de génotypes possibles, deux parentaux (A_1B_1 et A_2B_2) et deux recombinés (A_1B_2 et A_2B_1). Cette distorsion n'est alors possible que si la dépression hybride infecte différemment les deux combinaisons recombinées ($S_{A_1B_2} \neq S_{A_2B_1}$, voir figure supplémentaire 2, chapitre 1).

Les barrières post-zygotiques peuvent aussi résulter de deux autres processus différents des DMI, tels que l'existence de forts remaniements chromosomiques ou la perte de fonction des gènes dupliqués (Modèle d'Oka). La localisation des outliers hors des « régions répétées inversées » ou structurées en « isochores riches en AT » semble plus en accord avec une hypothèse d'un isolement post-zygotiques conférés par des DMIs car ces régions riches en GC sont plutôt représentatives d'un « core génome » et non de régions riches en gènes dupliqués (« îlots de transposons », Raffaele & Kamoun 2012). De plus, l'analyse des gènes présents dans ces 4 zones ne met pas en évidence des gènes communs qui auraient été dupliqués entre les régions comme attendus sous le modèle d'Oka. Notre résultat suggère plutôt que le temps de divergence entre les populations *avrRvi6* et *virRvi6* fut suffisamment long pour que les incompatibilités puissent s'être accumulées. Toutefois, la fixation des allèles d'incompatibilité dans les populations divergentes ne résulte pas forcément du seul effet de la dérive génétique mais pourrait avoir été accélérée par un avantage adaptatif conféré par ces mutations dans les fonds génétiques parentaux.

L'utilisation d'une approche de caractérisation fine des patrons d'autostop est une première étape visant à réduire sensiblement le nombre de gènes candidats responsables de ces barrières post-zygotiques. Cette approche a été limitée par l'absence de modèles prédictifs associés au modèle d'autostop en situation de contact secondaire, ce qui n'a pas permis pas de préciser les localisations des locus responsables de ces incompatibilités. Toutefois, une approche fonctionnelle devrait permettre de préciser cette localisation et, plus largement encore, de valider ou d'infirmer l'hypothèse selon laquelle ces barrières post-zygotiques sont de type DMIs. En outre, la description moléculaire de ces barrières post-zygotiques offre une perspective particulièrement intéressante à ce travail. En effet, bien que l'accumulation d'incompatibilités de type Dobzhansky-Muller soit invoquée comme le modèle général

expliquant la divergence de la plupart des organismes actuels (Coyne & Orr 2004), les locus (gènes, régions non-codantes) responsables de ces incompatibilités n'ont été identifiées que chez très peu d'organismes [*Arabidopsis* (Bomblies *et al.* 2007 ; Bikard *et al.* 2009), *Drosophila* (Sawamura & Yamamoto 1997; Barbash *et al.* 2003; Masly *et al.* 2006), *Mus musculus* (Mihora *et al.* 2008) et *Saccharomyces* (Lee *et al.* 2008)].

L'interprétation des scans génomiques étant particulièrement dépendante du cadre conceptuel adopté, l'exclusion des hypothèses d'isolement post-zygotiques peut avoir des répercussions majeures sur la localisation du locus d'avirulence. Le temps nécessaire à l'étude fine de la localisation des outliers étant bien plus importante que leur simple détection, l'interprétation des barrières endogènes comme exogènes peut alors conduire à augmenter considérablement le temps nécessaire à la découverte des seuls gènes d'avirulence. Cet effet est d'ailleurs proportionnel au temps de divergence des populations en allopatrie car on s'attend à ce que les barrières post-zygotiques endogènes s'accumulent à un rythme important sous l'hypothèse d'un effet boule de neige (au carré du temps de divergence pour les DMIs entre paires de locus, Coyne & Orr 2004 ; Matute *et al.* 2010). Les données actuellement disponibles pour les ascomycètes tels que *V. inaequalis* suggèrent toutefois que l'isolement reproductif suivrait plutôt un modèle linéaire que quadratique (Giraud & Gourbière 2012). Chez des populations très divergentes, la recherche des barrières exogènes peut donc s'avérer particulièrement délicate. Dans ce cas, l'abondance en barrières endogènes peut même réduire l'intérêt de la recherche du locus d'avirulence. En effet, une telle recherche n'a de sens que si les barrières exogènes sont considérés par défaut comme ayant initié le processus de divergence. Or des travaux théoriques récents prédisent que des barrières endogènes seules peuvent initier les premières étapes de divergence et que leur couplage avec des barrières exogènes suffise ensuite à renforcer l'isolement reproductif (Barton & De Cara 2009 ; Bierne *et al.* 2011).

En résumé, l'évolution d'un organisme phytopathogène ne peut être réduite à la seule interaction avec son hôte. La sélection naturelle n'est que l'une des quatre forces évolutives contribuant à l'évolution des populations de pathogènes. En outre seule une proportion des effets sélectifs peut être attribuée aux pressions de sélection exercées par l'hôte et ainsi contribuer à la mise en place de la spécialisation. L'étude de la diversification des pathogènes hors du seul prisme de l'adaptation à son hôte peut apparaître pertinente. Chez les organismes modèles où des analyses de scans génomiques ont été réalisées sur des centaines de milliers de polymorphismes (p. ex. *Homo*, *Drosophila*, *Arabidopsis*), le nombre de locus détectés comme étant sous sélection s'est révélé faible, ce qui est conforme à une évolution majoritairement gouvernée par la dérive génétique (Kimura 1983; Lynch 2007b, a). Chez les pathogènes, les pressions de sélection exercées par l'hôte sont fortes, particulièrement dans le cas des interactions de type Gène-pour-Gène ici étudiées. Pour autant, même sous ces hypothèses, l'évolution des populations avirulentes et virulentes des pathogènes peut aussi être gouvernée par une grande part de hasard. En conséquence, l'importance des forces stochastiques (p. ex. fixation aléatoire par dérive lors de grandes fluctuations démographiques) dans la diversification des pathogènes reste à déterminer.

Chez les champignons phytopathogènes, les caractéristiques permettant de formuler des attendus de spécialisation écologique ont été bien formulées (Giraud 2006a, b; Giraud *et al.* 2010). Cependant, outre ces caractéristiques favorables à la mise en place de barrières écologiques, les pathogènes et particulièrement les champignons présentent aussi des caractéristiques favorables à la mise en place de barrières post-zygotiques endogènes.

2.3 Particularité des champignons pathogènes

Les champignons phytopathogènes présentent des génomes généralement considérés comme particulièrement flexibles notamment par la forte mobilité des éléments transposables dans les génomes. Si la duplication de gènes est généralement considérée comme une source de nouvelles adaptations (Marques *et al.* 2008; Powell *et al.* 2008; Spanu *et al.* 2010; Kondrashov 2012), la duplication de gènes conduit, dans la majorité des cas, à une extinction rapide de l'expression d'une des deux copies de ces gènes dupliqués (Watterson 1983; Lynch & Conery 2000; Lynch & Force 2000). L'arrêt de l'expression d'une des copies résulterait d'une accumulation de mutations (Force *et al.* 1999; Lynch & Force 2000; Lynch *et al.* 2001). Or comme la majorité des mutations affectant la valeur sélective sont délétères (Lynch & Walsh 1998), ces mutations induisent alors la perte de fonction du gène. D'autres auteurs suggèrent que cet arrêt d'expression pourrait être d'origine épigénétique (Rodin & Riggs 2003; Rodin *et al.* 2005). Considérons que suite à un événement de duplication ancestral, la mutation vers des allèles non-fonctionnels affecte des copies différentes dans les deux populations en allopatrie, la remise en contact peut conduire à la non-viabilité des hybrides qui ne disposent d'aucune copie fonctionnelle (voir modèle d'Oka, introduction § 1.2.2.1.2). La mutation des gènes dupliqués pourrait être accélérée par les mécanismes d'extinction de l'activité des rétrotransposons (RIP) observés chez certains champignons comme *L. maculans* ou *V. inaequalis* (Rouxel *et al.* 2011). La mise en place de barrières endogènes par duplications de gènes pourrait alors être particulièrement importante dans l'isolement reproductif des populations de pathogènes. De la même façon, cette forte mobilité des génomes pourraient entraîner des remaniements chromosomiques importants et jouer un rôle important dans la mise en place de ce type de barrières post-zygotiques endogènes (§ 1.2.2.1.1).

Le couplage des barrières exogènes et endogènes est donc raisonnablement attendu chez les organismes pathogènes et particulièrement dans le cas des populations avirulentes et virulentes. En effet, le couplage des barrières endogènes et exogènes est d'autant plus rapide que le cline de fréquence allélique au locus de la barrière exogène est marqué (Bierne *et al.* 2011). Par conséquent, même sous un modèle où le gène d'avirulence a un rôle prépondérant dans la rupture du flux génique, les attendus associés au contact secondaire de deux populations ayant divergé en allopatrie sont un couplage rapide des barrières post-zygotiques avec la barrière exogène, c'est-à-dire ici le gène d'avirulence (Bierne *et al.* 2011). Par conséquent, plus la sélection écologique est importante, *via* par exemple la non-viabilité écologique des immigrants, plus la probabilité que les barrières endogènes soient spatialement couplées avec les barrières exogènes est forte. Enfin si de forts taux de sélection favorisent intrinsèquement ce processus, Bierne et ses collaborateurs notent que ce couplage est toutefois plus attendu pour des taux de sélection dits « modérés » ($s=0.5$) car un taux encore plus fort tend ensuite à réduire la largeur du cline exogène et donc la probabilité que les deux types de barrières se rencontrent dans l'espace et se couplent. Il est donc fondamental de considérer que les deux types de barrières comme complémentaires par leur couplage et leur contribution à l'isolement reproductif total.

Nos résultats suggèrent qu'un isolement reproductif fort est nécessaire pour maintenir stable les zones d'hybridations, c'est-à-dire la mise en place de zones hybrides *sensus stricto*. En effet, la forte fécondité associée aux populations de pathogènes (p. ex. Andrivon *et al.* 2007) rend la recombinaison particulièrement efficace pour générer des combinaisons sans incompatibilité chez les organismes haploïdes. De plus, dans notre cas, on observe une rapide augmentation des allèles type virulent aux outliers au cours d'une période de 4 ans, ce qui suggère que l'allèle de virulence est favorable dans l'ensemble de l'habitat (hôtes *Rvi6* et non-*Rvi6*) et accompagnerait le cortège d'allèles type virulent aux barrières endogènes, ce qui

conduirait à la fixation des allèles type virulent. En absence d'un grand nombre de barrières post-zygotiques et d'une barrière exogène forte, il apparaît alors peu vraisemblable que les zones d'hybridation puisse rester durablement stables au cours du temps. Des modèles prédictifs sont nécessaires pour formaliser ces attendus, mais si cette hypothèse est vérifiée, elle pourrait expliquer pourquoi aucune zone hybride n'est actuellement décrite chez les pathogènes (Barton, communication personnelle). Toutefois, aussi ponctuelles soient-elles, ces zones présentent des risques majeurs pour l'évolution de la pathogénie des populations de pathogènes. En effet, outre l'absence d'incompatibilités, la recombinaison chez des populations de pathogènes pourrait rapidement générer un large panel de combinaisons génétiques dont certaines pourraient contribuer à une augmentation de la valeur sélective des populations. Sous cette hypothèse, ces zones de contact avec flux de gènes ne favoriseraient pas l'établissement de zones hybrides mais plutôt contribuerait à un remplacement des populations parentales.

3. Asymétrie de l'introgession et conséquences épidémiologiques

3.1 Sens de l'introgession du fond génétique

L'introgession détectée semble particulièrement asymétrique dans la zone d'hybridation avec une prépondérance du flux génique dans le sens de la population native *avrRvi6* vers la population invasive *virRvi6*. Ce résultat correspond bien aux attendus théoriques d'un cas de contact secondaire entre une population native et une population invasive (Currat & Excoffier 2004; Currat *et al.* 2008; Excoffier *et al.* 2009a). En effet, ces auteurs ont réalisé des simulations démo-génétiques et suggèrent que la dynamique d'invasion peut expliquer ce patron d'introgession à elle seule et ne nécessite donc pas d'invoquer l'intervention de la sélection naturelle¹⁴. Les raisons de cette asymétrie sont multiples (Excoffier *et al.* 2009). Premièrement, l'invasion progressive par des souches virulentes va contribuer à une dilution accrue du fond génétique de la population invasive, ce qui va conduire à une admixtion plus importante de ces individus le long du front d'expansion. Deuxièmement, les deux populations ne sont pas au même stade dans l'infection de leurs habitats car la population invasive augmente fortement en fréquence dans ces nouvelles zones, tandis que la taille de la population locale n'augmente pas (reste limité par la capacité du milieu) voire diminue à cause de la compétition avec les populations envahissantes. En conséquence, le gène introgressé dans l'espèce invasive a peu de chance d'être éliminé par la dérive génétique si la population augmente rapidement en fréquence dans les nouveaux habitats colonisés. Troisièmement, un allèle introgressé dans la population invasive peut « surfer » et atteindre des fortes fréquences dans les zones nouvellement colonisées.

Toutefois, les attendus de ce modèle suggèrent que le fond génétique avirulent ne devrait recevoir que peu d'allèles du compartiment virulent. En effet, la théorie prévoit que les gènes du fond virulent n'introgessent le fond avirulent que s'ils sont sous forte sélection positive (Moran 1981; Buggs 2007; Currat *et al.* 2008). Or si l'introgession semble bien asymétrique, le fond génétique avirulent semble toutefois avoir introgressé une part importante d'allèles originaires du fond génétique virulent. Nous avons alors suggéré que

¹⁴ Ce patron est observé sous de nombreuses conditions démographiques incluant une plus forte densité de la population invasive que native ou une compétition des deux espèces conduisant à l'extinction de l'espèce native. Néanmoins, si la fitness des hybrides est très faible (inférieure à 10%), les patrons sont toutefois plus complexes, via des gradients d'introgession ou des introgession bidirectionnelles dans la zone de contact (pour plus de détails voir Currat *et al.* 2008).

certains caractères présents dans le fond génétique virulent pouvaient être avantageux pour les populations *avrRvi6*.

3.2 Introgression des facteurs d'agressivité et conséquences épidémiologiques

Les tests d'agressivités réalisés en conditions contrôlées révèlent que les souches les plus agressives sur les cultivars Gala et Top Red sont les souches hybrides présentant une part relativement équivalente des deux génomes. De manière intéressante, les locus liés à l'agressivité semblent provenir des deux fonds génétiques (six semblent provenir du fond avirulent, quatre du fond virulent) suggérant que les facteurs d'agressivité présents dans chacun des deux fonds génétiques peuvent être introgressés dans un même fond génétique. Les hybrides possédant la majorité de ces facteurs d'agressivité semblent alors plus agressifs sur un cultivar communément déployé dans les vergers de pommiers tel que Gala. De plus, une telle introgression peut permettre de contrebalancer les coûts de fitness associés à l'allèle de virulence, si ceux-ci existent. Dans cette situation, les conséquences évolutives associées à une remise en contact de populations virulentes et avirulentes en vergers peuvent donc être majeures : (i) la résistance n'a plus aucun effet sur la population car la virulence est marginalement favorable dans l'ensemble de l'habitat et le polymorphisme à ce locus n'est plus maintenu par le coût de fitness associé à la virulence. La virulence envahit donc l'ensemble de la population et la résistance est alors totalement perdue. (ii) L'introgression de l'ensemble des facteurs d'agressivité conduit à une augmentation globale de l'agressivité du pathogène sur l'ensemble des hôtes (résistants mais aussi sensibles). Ce dernier effet est le plus néfaste car il n'impacte pas uniquement les seuls arbres porteurs du gène de résistance mais la population d'hôte dans son ensemble. Cette conséquence, spécifique au modèle d'émergence par migration de souches virulentes préexistantes, conduirait alors à un paradoxe ; les programmes de sélection pour la résistance visant à réduire les populations de pathogène peuvent indirectement contribuer à une augmentation de l'agressivité des populations de pathogènes.

4. Réflexions sur la durabilité des résistances

Le déploiement des gènes de résistance dans les agrosystèmes n'a été véritablement durable que dans bien peu de pathosystèmes (mais voir Kolmer 2003). Généralement, le contournement de ces résistances résulte de nouvelles mutations au locus d'avirulence au sein de populations infectant les cultivars sensibles permettant ainsi l'infection des hôtes porteurs du gène de résistance. Ce contournement s'accompagne d'un coût de fitness du pathogène virulent sur l'hôte sensible. Un des enseignements majeurs de notre travail est que l'émergence de la virulence à partir du compartiment non-cultivé génère des attendus différents.

Sous ce modèle, la durabilité de la résistance n'est pas seulement dépendante de l'éventuelle mise en place de nouvelles mutations mais aussi des capacités de migration des souches virulentes autorisant la sortie du compartiment non-cultivé et l'infection des agrosystèmes. Cette capacité de migration nécessite une bonne connaissance de la gamme d'hôte des pathogènes, ce qui implique une connaissance exhaustive des hôtes alternatifs de ces populations. Les travaux de Lê Van *et al.* (2011) montrent que la recherche de virulence dans le compartiment sauvage reste laborieuse et très aléatoire. De plus, les hôtes pouvant être totalement dépourvus d'intérêt agronomique. Or, la gamme d'hôtes et, plus largement encore,

la gamme d'habitats possibles des pathogènes est probablement sous-estimée par notre focalisation sur les seuls agrosystèmes (Morris *et al.* 2007; Morris *et al.* 2009). Ces habitats alternatifs « non-agronomiques » pourraient pourtant être des zones particulièrement importantes dans la diversification des populations de pathogènes *via* par exemple l'introgession de facteurs d'agressivité entre différents pathotypes. L'utilisation des approches de métagénomique basées sur des données de séquençage à haut débit offre la possibilité de mieux appréhender les habitats alternatifs des pathogènes et de renseigner la connectivité éventuelles entre les populations divergentes afin de mieux prévenir les risques liés à l'introgession de locus liés à la pathogénie entre ces populations.

Un autre enseignement important est que les plantations composées de génotypes issus des programmes de sélection (comme le verger de sélection, site 2) sont des zones particulièrement propices pour observer des zones d'hybridation entre populations divergentes. Dans ces zones, la présence de souches virulentes du compartiment non-cultivé est particulièrement attendue favorisant d'emblée une remise en contact secondaire avec des populations avirulentes. Les risques sont renforcés par plusieurs facteurs inhérents au criblage des descendants résistants au champ : (i) les parents (dont la source de la résistance) sont généralement plantés au voisinage des descendants, (ii) le gène de résistance n'est pas protégé par des fongicides ce qui favorise la présence de souches virulentes et (iii) les descendants du croisement d'hôtes présentent un large panel de phénotypes pour la résistance, ce qui contribue à la création d'habitats intermédiaires. Ces pratiques sont évidemment difficiles à modifier, surtout pour des programmes de sélection impliquant des plantes ligneuses¹⁵. Toutefois, l'isolement géographique des parents et des descendants semble être une mesure préventive simple permettant de réduire les risques d'émergence et de recombinaison entre les pools génétiques dans ces habitats particuliers.

Toutefois, dans le cas de la résistance *Rvi6*, ce type de contact s'est probablement aussi produit dans d'autres contextes. En effet, ce contact secondaire avec flux de gènes a aussi été observé dans d'autres vergers composés des variétés commerciales Ariane (cv. *Rvi6*), Gala et Reine des Reinettes (cv. non-*Rvi6*) (La Rétuzière, Maine-et-Loire ; Caffier, communication personnelle). Le nombre de cas de contacts secondaires entre les populations virulentes et avirulentes est élevé (au moins 8 pays concernés en Europe), qu'il soit suivi de flux de gènes reste plus difficile à évaluer. Plusieurs arguments permettent pourtant de suggérer que leur nombre peut être élevé. En effet, *M. x floribunda*, le progéniteur de la résistance *Rvi6*, est une espèce utilisée en ornement pour fleurir les jardins et les parcs publics. Hors des vergers de production, les contacts secondaires pourraient donc se produire dans des jardins publics ou privés. De plus, outre la création des variétés *Rvi6*, l'utilisation de *M. x floribunda* dans les programmes de sélection contre la tavelure a permis la création de cultivars *Rvi6* utilisés uniquement comme pollinisateurs dans les vergers de pommiers sensibles (le cv. *Rvi6* Everest), décuplant encore le nombre de zones de contacts secondaires possibles.

Plus largement, une réflexion sur le rôle de la recombinaison des souches différenciellement adaptées doit être engagée à la suite de ce travail. En effet, le contournement rapide des résistances génétiques monogéniques conduit actuellement à une diminution de l'usage de ces ressources génétiques dans les programmes de sélection. L'émergence de nouveaux programmes centrés sur l'utilisation de résistances quantitatives est motivé par l'hypothèse que ces dernières présenteraient des résistances plus durables car générant des pressions de sélection plus faibles sur les populations de pathogènes (p. ex. Lindhout, 2002 ; Parlevliet, 2002). Toutefois, ces résistances semblent aussi pouvoir être rapidement érodées dans l'environnement (Pariaud *et al.* 2009 ; Montarry *et al.* 2012 ; Caffier, en préparation). Si des adaptations locales partielles de ces résistances sont préexistantes dans

¹⁵ Pour des arbres, il est en effet inimaginable de réaliser ces tests en conditions contrôlées.

différentes populations d'un pathogène, la recombinaison pourrait générer de nouvelles combinaisons expliquant en partie ces cas d'érosions rapides de résistances. Une réflexion globale sur la gestion durable des résistances (majeure et partielle) doit intégrer une redéfinition des agrosystèmes eux-mêmes afin d'introduire plus d'hétérogénéité dans les habitats, *via* des mélanges variétaux et plus efficacement encore des associations de différentes espèces végétales, et à plus de rotations de cultures pour les plantes annuelles afin de réduire la taille des populations et de limiter les émergences. Nous revenons ainsi aux fondements même de l'agriculture (Stukenbrock & McDonald 2008; McDonald 2010).

CONCLUSION ET PERSPECTIVES

1. Conclusion

Depuis la découverte des relations Gène-pour-Gène par Flor (1955, 1971), les résistances monogéniques ont été intégrées dans un grand nombre de programmes d'amélioration variétale pour la résistance aux bioagresseurs. Dans la majorité des cas, le déploiement des gènes de résistance dans l'environnement s'est accompagné d'une perte rapide de l'efficacité de la résistance dans les agrosystèmes (Parleviet 2002). Pour les résistances monogéniques les moins éphémères, la résistance est restée efficace pendant quelques dizaines d'années, ce qui reste insuffisant pour satisfaire pleinement les définitions de durabilité, c'est-à-dire au minimum une efficacité de la résistance pendant toute la durée d'utilisation commerciale du cultivar (Leach *et al.* 2001). Les modèles actuels de diversification des pathogènes suggèrent que la mutation de l'avirulence vers la virulence est la raison majeure de ces contournements rapides. Dans ce travail, nous proposons un autre modèle où la virulence serait préexistante dans le compartiment non-cultivé et migrerait accidentellement dans les vergers. Ce dernier modèle a été récemment démontré comme étant le plus probable pour l'émergence des populations *virRvi6* de *V. inaequalis* (Lemaire *et al.* en préparation). Au cours de cette thèse, nous avons utilisé le pathosystème pommier/*V. inaequalis* pour prédire les conséquences associées à ce nouveau modèle d'émergence. Si les deux modèles d'émergence conduisent tous deux de façon similaire à la perte de la résistance, notre modèle repose sur un changement récent de régime de migration des populations, passant de l'alopatrie vers la parapatrie et ainsi autorisant l'établissement de contacts secondaires avec flux de gènes.

La première conséquence de ce modèle d'émergence est que l'isolement reproductif ne résulte plus uniquement du seul gène d'avirulence, mais aussi de l'existence de barrières post-zygotiques accumulées pendant la divergence en alopatrie. On peut alors s'attendre à ce que le temps de divergence entre les deux populations soit un bon estimateur du nombre de barrières génétiques post-zygotiques accumulées. Dans le couple *V. inaequalis*/*Malus*, la divergence entre la population *avrRvi6* et *virRvi6* semble être récente (< à 10 000 ans, Lemaire *et al.* en préparation), ce qui suggère que le nombre de barrières post-zygotiques attendu est faible. Pourtant, bien que notre approche de scan génomique ne permette pas un échantillonnage exhaustif des barrières, plusieurs outliers marqueurs de barrières différentes ont été détectés, ce qui est incompatible avec la seule détection du locus d'avirulence. De plus, notre approche de cartographie génétique a permis de révéler que les marqueurs SNP autour des outliers sont cartographiés sur un même groupe de liaison, ce qui suggère que ces derniers sont en fort déséquilibre de liaison entre eux chez les hybrides même en absence de pression de sélection par l'hôte. Nous avons démontré qu'une part importante de l'isolement reproductif observé provient des barrières post-zygotiques et proposé que ces barrières résultent d'incompatibilités génétiques de type Dobzhansky-Muller (DMI).

La seconde conséquence est que la recombinaison de ces populations divergentes peut avoir des retombées épidémiologiques majeures. La première est que le coût de fitness associée à la perte de l'avirulence qui peut être très ancien, semble moindre que sous le modèle de mutation vers la virulence, l'attendu est donc une fixation rapide de l'allèle de virulence dans l'ensemble de la population. La conservation d'un effet résiduel de la résistance (voir p ex. Li *et al.* 1999) est donc peu vraisemblable dans ce type d'émergence. De plus, si les populations possèdent des adaptations locales différentes, le contact secondaire peut conduire à l'établissement de nouvelles combinaisons porteuses de l'ensemble des facteurs d'agressivité sur l'hôte résistant mais aussi sur l'hôte sensible. Or, si les programmes de sélection développent des nouvelles variétés résistantes, c'est justement parce que le niveau de résistance des variétés présentes cultivées n'est pas satisfaisant. Ce modèle

d'émergence conduit alors à un paradoxe : les programmes de sélection peuvent contribuer à rendre ces variétés encore plus sensibles aux pathogènes. Chez *V. inaequalis*, cette conséquence est particulièrement néfaste car les variétés résistantes ne représentent que des surfaces marginales des cultures (les variétés *Rvi6* ne représente que 0.5% de la surface de production des vergers Européens), cette augmentation de l'agressivité donc avoir un impact beaucoup plus important sur les vergers de pommiers sensibles. Le risque associé à cette augmentation de l'agressivité pourrait toutefois prendre la forme d'un recours plus important aux fongicides pour contrôler le développement des épidémies. Pour finir par une touche enfin optimiste, on peut noter que la situation pourrait s'avérer encore plus dramatique si les populations *virRvi6* présentaient une résistance accrue aux fongicides, or aucune étude dans ce sens n'a été menée jusqu'à présent. Pour conclure, nous proposons un modèle prédictif de risque destiné aux sélectionneurs afin de contribuer à une meilleure durabilité des résistances génétiques monogéniques et à réduire le risque associé à l'introgression de facteurs d'agressivité en situation de contacts secondaires avec flux de gènes entre les populations divergentes (Figure 1).

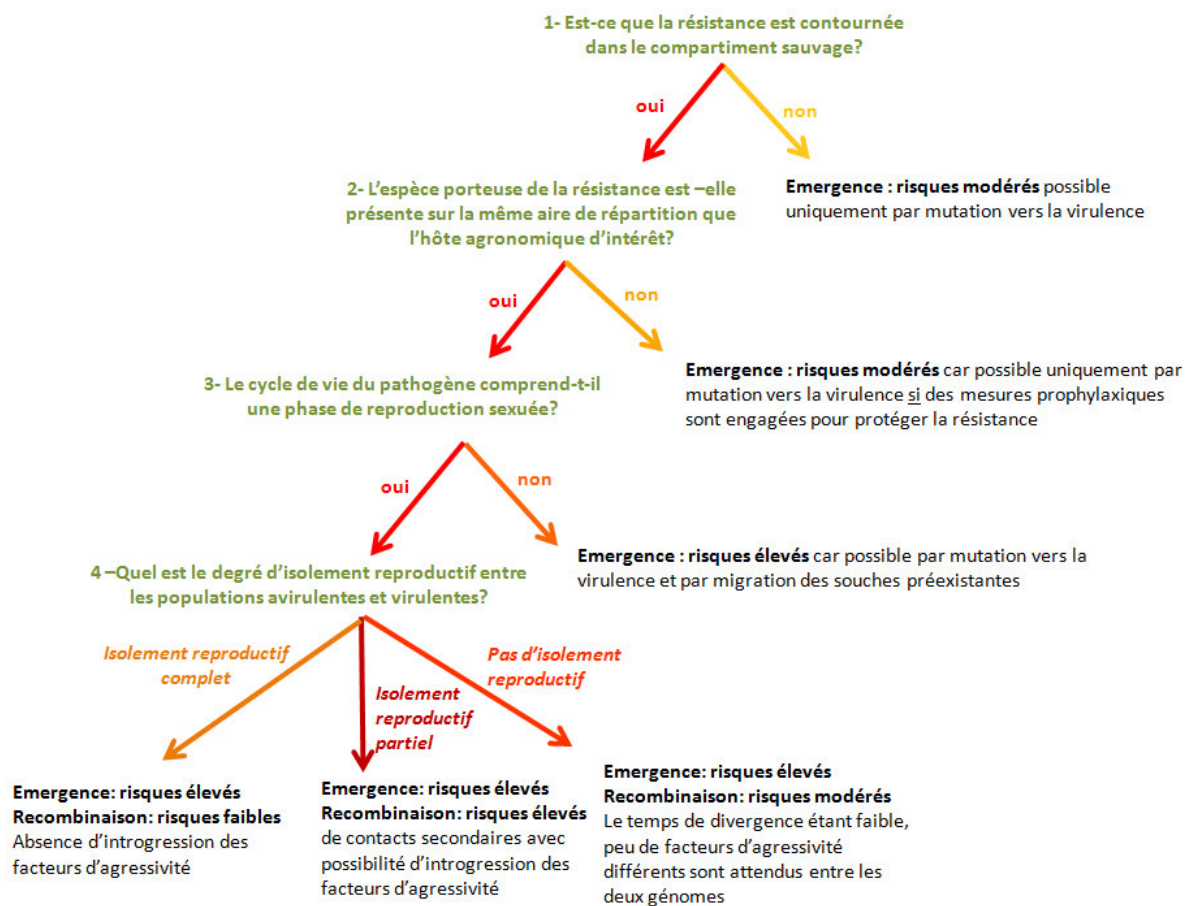


Figure 1 : Modèle prédictif associé aux risques d'émergences par mutation, par migration de souches virulentes préexistantes dans le compartiment sauvage et aux risques de contact secondaire avec flux de gènes. Considérant que la sortie générale des résistances monogéniques est le contournement, un risque modéré ne sous-entend pas ici que l'émergence n'est pas vraisemblable mais plutôt le temps nécessaire au contournement est sous la dépendance du taux de mutation au locus d'avirulence.

2. Perspectives

2.1 A court terme

A l'heure actuelle, seule une petite partie des outliers ont été correctement localisés sur le génome. Au cours de cette thèse, la localisation la plus efficace des bandes AFLP a été obtenue par simulation d'une AFLP *in silico* (Paris & Despres 2010) en utilisant les séquences des 3 génomes *avrRvi6* actuellement disponibles. L'acquisition de la séquence d'une souche *virRvi6* devrait permettre de localiser plus d'outliers sur le génome. La validation des polymorphismes AFLP pourra être effectuée directement par le génotypage de marqueurs SNPs dans les régions probables de ces nouveaux fragments, afin de valider si ces marqueurs présentent des patrons d'autostop comme observé sur les autres outliers et de les localiser sur la carte génétique.

Deux situations bien contrastées ont été observées entre les populations *avrRvi6* et *virRvi6* infectant les vergers à couteau et les vergers à cidre. La situation sur le site P32 était particulièrement favorable à l'étude car plusieurs générations de reproduction sexuée entre *avrRvi6* et *virRvi6* ont eu lieu, permettant d'observer des patrons d'autostop sur des zones génomiques assez restreintes (< 100kb). *A contrario*, dans les vergers de pommier à cidre, la différenciation semblait se maintenir au cours d'un suivi pluriannuel de 5 ans (Gladieux *et al.* 2011). L'échantillonnage des populations du site de La Folletière réalisé en 2012 et le génotypage des populations sur les mêmes marqueurs microsatellites que dans l'étude de Gladieux *et al.* (2011) permettra de répondre à cette question.

2.2 A plus long terme

Notre approche de génomique des populations et la mise en évidence des patrons d'autostop est une première étape permettant la recherche des gènes candidats de barrières génétiques. Cette approche nécessitera d'abord de préciser la localisation des gènes candidats en génotypant plusieurs autres SNPs dans ces zones. L'acquisition de la séquence du premier génome *virRvi6* permettra de cibler préférentiellement des polymorphismes dérivés (la recherche de SNPs « spécifiques » entre les 3 génomes *avrRvi6* et le génome *virRvi6*) afin de mieux caractériser la zone présentant le maximum de différenciation et ainsi réduire le nombre de gènes candidats. Ce criblage des gènes candidats ouvre ensuite la voie vers une caractérisation fonctionnelle de ces incompatibilités. La mise en évidence du fonctionnement des barrières post-zygotiques est rare et les évidences ne sont résumées qu'à un faible nombre d'organismes modèles. A terme, *V. inaequalis* pourrait donc intégrer cette liste. Chez cet organisme, la transformation homologue du champignon est maîtrisée dans mon équipe d'accueil (Qin, communication personnelle). Un changement d'allèle au niveau des barrières est donc possible et permettrait de démontrer le lien entre la dépression hybride et la présence de DMIs. Ces effets seraient d'ailleurs quantifiables *in vitro* par différentes mesures de croissance fongique et de capacité de sporulation.

Lorsque j'ai débuté ce travail de thèse, un des premiers objectifs était de localiser le locus *avrRvi6* sur le génome. Nous avons démontré que la détection de ce seul locus entre les génomes *avrRvi6* et *virRvi6* par scan génomique est difficile, malgré le nombre de locus génotypés (969). En effet, les barrières post-zygotiques endogènes génèrent les mêmes types de patrons de différenciation que les barrières exogènes. Or comme les barrières post-zygotiques endogènes peuvent rapidement dépasser le nombre de barrières exogènes, la

découverte du locus d'avirulence est de plus en plus complexe à mesure que le temps de divergence est long. Concernant la localisation du gène d'avirulence, plusieurs sorties sont alors envisageables, (i) dans la situation la plus favorable, la cartographie des autres outiers permet une détection de la barrière exogène *via* une colocalisation de l'outier avec le locus *virRvi6* localisé sur la carte génétique grâce au phénotype ; (ii) si cette première approche reste infructueuse, une approche de génomique des populations ciblée sur le reséquençage de génomes est une piste possible pouvant permettre la mise en évidence de toutes les barrières génétiques dont le gène d'avirulence *avrRvi6*.

Le modèle de prédiction des risques proposé dans le cadre d'une gestion durable des résistances (Figure 1 ci-contre) nécessite d'être simplifié pour une utilisation efficace dans les programmes de sélection. En effet, sous ce modèle, il faut d'abord renseigner le risque d'une préexistence de la virulence dans le compartiment non-cultivé, ce qui nécessite - au minimum¹⁶ - une bonne connaissance de l'aire de répartition de l'hôte progéniteur de la résistance. Si cette hypothèse est vérifiée, il est ensuite nécessaire de caractériser l'intensité de l'isolement reproductif entre les populations avirulentes des agrosystèmes et virulentes du compartiment non-cultivé. Bien que cette approche soit possible *in vitro* chez les organismes cultivables sur l'ensemble du cycle de vie, ces travaux sont assez lourds et ne permettent pas de déterminer avec exactitude si les populations seront totalement imperméables au flux génique dans l'environnement¹⁷. Alternativement, je propose de préciser ce schéma par une meilleure connaissance des cas potentiels d'émergence à partir du compartiment non-cultivé. Une méta-analyse des émergences des virulences (mutation/migration) sera ensuite nécessaire pour déterminer si des caractères d'histoire de vie des hôtes et des pathogènes sont particulièrement sujets à ce type d'émergence. Par exemple, est-ce que des taux de mutations différents existent aux locus d'avirulence entre les différentes espèces de pathogènes ? Et si tel est le cas, est-ce que cette différence s'explique par un compromis évolutif entre le taux de mutation et le caractère pérenne de son hôte dans l'environnement ?

Pour conclure, je pense que la durabilité des résistances aux pathogènes n'est possible que par une plus grande focalisation sur le pathogène dès les premières étapes des programmes de sélection. Dans ce contexte, les nouvelles approches ciblées sur la détection d'effecteurs conservés chez les pathogènes et la recherche ensuite de gènes de résistance chez les plantes (Bart *et al.* 2012) est un bon espoir d'observer un jour des résistances plus durables dans les agrosystèmes. Cette stratégie sera ensuite d'autant plus durable que ces résistances génétiques seront protégées par une lutte chimique raisonnée. La prophylaxie *via* un déploiement spatial plus raisonné des résistances (pas de variétés résistantes dans la zone d'origine de la résistance), un usage restreint à un secteur professionnel et accompagné d'une redéfinition de nos agrosystèmes (McDonald 2010) sont autant de facteurs supplémentaires qui contribueront à conférer à cette nouvelle génération de résistances un caractère plus durable.

¹⁶ Cette information est importante mais n'est pas pour autant suffisante. Le gène de résistance peut en effet être présent dans d'autres espèces végétales que l'espèce utilisée pour l'introgession de la résistance. Chez les *Malus*, la résistance *Rvi6* n'est pas observée exclusivement chez *M. x floribunda*, mais aussi chez *M. baccata* et *M. ioensis*, voir annexe)

¹⁷ Le nombre de croisements dans l'environnement sera bien supérieur à ce qui est testable *in vitro*. La recombinaison peut avoir lieu à une fréquence faible dans l'environnement, même en intégrant en plus l'isolement reproductif conféré par la barrière exogène. Certaines nouvelles combinaisons génétiques présentant des meilleures fitness que les populations parentales pourraient donc être produites à une fréquence faible.

BIBLIOGRAPHIE

- Aderhold R. (1899) Auf welche Weise können wir dem immer weiteren Umsichgreifen des Fusicladiums in unseren Apfelkulturen gegengen und welche Sorten haben sich bisher dem Pilz gegenüber am widerstrandsfähigsten gezeigt. *Pomol Monatsh*, XLV, 266-272
- Aguilar A. (2006) Analysis of FST outliers at allozyme loci in Pacific salmon: implications for natural selection. *Environmental Biology of Fishes*, 76, 329-339
- Alexopoulos C.J., Mims C.W. & Blackwell M. (1996) *Introductory mycology* 4th Edition edn. Wiley, New York, USA.
- Anagnostakis S.L. (1987) Chestnut blight: the classical problem of an introduced pathogen. *Mycologia*, 79, 23-37
- Andrion D., Pilet F., Montarry J., Hafidi M., Corbiere R., Achbani el H., Pelle R. & Ellisseche D. (2007) Adaptation of *Phytophthora infestans* to Partial Resistance in Potato: Evidence from French and Moroccan Populations. *Phytopathology*, 97, 338-43
- Aylor D.E. (1998) The aerobiology of apple scab. *Plant Disease* 82, 838-849.
- Bagga H.S. & Boone D.M. (1968) Genes in *Venturia inaequalis* controlling pathogenicity to crabapples. *Phytopathology*, 58, 1176-1182
- Barbash D.A., Siino D.F., Tarone A.M. & Roote J. (2003) A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proc Natl Acad Sci U S A*, 100, 5302-7
- Bart R., Cohn M., Kassen A., McCallum E.J., Shybut M., Petriello A., Krasileva K., Dahlbeck D., Medina C., Alicai T., Kumar L., Moreira L.M., Rodrigues Neto J., Verdier V., Santana M.A., Kositcharoenkul N., Vanderschuren H., Gruissem W., Bernal A. & Staskawicz B.J. (2012) High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc Natl Acad Sci U S A*, 109, E1972-9
- Barton N. & Rouhani S. (1991) The probability of fixation of a new karyotype in a continuous population. *Evolution*, 45, 499-517
- Barton N.H. (1983) Multilocus clines. *Evolution*, 37, 454-471
- Barton N.H. (2001) Speciation. *Trends Ecol Evol*, 16, 325
- Barton N.H. (2010) What role does natural selection play in speciation? *Philos Trans R Soc Lond B Biol Sci*, 365, 1825-40
- Barton N.H. & de Cara M.A. (2009) The evolution of strong reproductive isolation. *Evolution*, 63, 1171-90
- Barton N.H. & Gale K.S. (1993) Genetic analysis of hybrid zones. In: *Hybrid zones and the evolutionary process* (ed. Harrison RG), pp. 13-45. Oxford university press, Oxford
- Barton N.H. & Hewitt G.M. (1985) Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics*, 16, 113-148
- Barton N.H. & Shpak M. (2000) The effect of epistasis on the structure of hybrid zones. *Genet Res*, 75, 179-98
- Bateson W. (1909) Heredity and variation in modern lights. In: *Darwin and modern science* (ed. Seward A), pp. 85-101. Cambridge University Press, Cambridge
- Bayles R.A., Flath K., Hovmoller M.S. & de Vallavieille-Pope C. (2000) Breakdown of the *Yr 17* resistance to yellow rust of wheat in northern Europe. *Agronomie*, 20, 805-811
- Bazin E., Dawson K.J. & Beaumont M.A. (2010) Likelihood-free inference of population structure and local adaptation in a Bayesian hierarchical model. *Genetics*, 185, 587-602
- Bazykin A.D. (1969) Hypothetical mechanism of speciation. *Evolution*, 23, 685-687
- Beaumont M.A. & Balding D.J. (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, 13, 969-980
- Beaumont M.A. & Nichols R.A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 263, 1619-1626

- Beaumont M.A., Zhang W. & Balding D.J. (2002) Approximate Bayesian computation in population genetics. *Genetics*, 162, 2025-35
- Belfanti E., Silfverberg-Dilworth E., Tartarini S., Patocchi A., Barbieri M., Zhu J., Vinatzer B.A., Gianfranceschi L., Gessler C. & Sansavini S. (2004) The *HcrVf2* gene from a wild apple confers scab resistance to a transgenic cultivated variety. *Proc Natl Acad Sci U S A*, 101, 886-90
- Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. (1996-2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. In. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier
- Bell G. (1982) *The Masterpiece of Nature: The Evolution and Genetics of Sexuality*, Berkeley, California, USA.
- Bell G. & Maynard Smith J. (1987) Short-term selection for recombination among mutually antagonistic species. *Nature*, 328, 66-68
- Benaouf G. & Parisi L. (2000) Genetics of Host-Pathogen Relationships Between *Venturia inaequalis* Races 6 and 7 and *Malus* Species. *Phytopathology*, 90, 236-42
- Berbee M.L., Carmean D.A. & Winka K. (2000) Ribosomal DNA and resolution of branching order among the ascomycota: how many nucleotides are enough? *Mol Phylogenet Evol*, 17, 337-44
- Beresford R.M. & Manktelow D.W.L. (1994) Economics of reducing fungicide use by weather-based disease forecasts for control of *Venturia inaequalis* in apples. *New Zealand Journal of Crop and Horticultural Science*, 22, 113-120
- Bert T.M. & Arnold W.S. (1995) An Empirical-Test of Predictions of 2 Competing Models for the Maintenance and Fate of Hybrid Zones - Both Models Are Supported in a Hard-Clam Hybrid Zone. *Evolution*, 49, 276-289
- Bierne N. (2010) The distinctive footprints of local hitchhiking in a varied environment and global hitchhiking in a subdivided population. *Evolution*, 64, 3254-72
- Bierne N., Welch J., Loire E., Bonhomme F. & David P. (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, 20, 2044-72
- Bomblies K., Lempe J., Epple P., Warthmann N., Lanz C., Dangl J.L. & Weigel D. (2007) Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol*, 5, e236
- Bonhomme M., Chevalet C., Servin B., Boitard S., Abdallah J.M., Blott S. & San Cristobal M. (2010) Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. *Genetics*, 186, 241-262
- Boone D.M. & Keitt G.W. (1957) *Venturia inaequalis* (Cke.) Wint. XII. Genes controlling pathogenicity of wild-type lines. *Phytopathology*, 47, 403-409
- Boughman J.W., Rundle H.D. & Schluter D. (2005) Parallel evolution of sexual isolation in sticklebacks. *Evolution*, 59, 361-373
- Brasier C.M. (1995) Episodic selection as a force in fungal microevolution with special reference to clonal speciation and hybrid introgression. *Canadian Journal of Botany*, 73, 1213-1221
- Brasier C.M., Cooke D.E. & Duncan J.M. (1999) Origin of a new *Phytophthora* pathogen through interspecific hybridization. *Proc Natl Acad Sci U S A*, 96, 5878-83
- Brasier C.M., Kirk S.A., Pipe N.D. & Buck K.W. (1998) Rare interspecific hybrids in natural populations of the Dutch elm disease pathogens *Ophiostoma ulmi* and *O. novo-ulmi*. *Mycological Research*, 102, 45-57
- Brideau N.J., Flores H.A., Wang J., Maheshwari S., Wang X. & Barbash D.A. (2006) Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science*, 314, 1292-5

- Broggini G.A., Bus V.G., Parravicini G., Kumar S., Groenwold R. & Gessler C. (2011) Genetic mapping of 14 avirulence genes in an EU-B04 x 1639 progeny of *Venturia inaequalis*. *Fungal Genet Biol*, 48, 166-76
- Broggini G.A., Le Cam B., Parisi L., Wu C., Zhang H.B., Gessler C. & Patocchi A. (2007) Construction of a contig of BAC clones spanning the region of the apple scab avirulence gene *AvrVg*. *Fungal Genet Biol*, 44, 44-51
- Brown J., Hovmøller M., Wyand R. & Yu D. (2002) Oases in the desert: dispersal and host specialization of biotrophic fungal pathogens of plants. In: (eds. Bullock J, Kenward R & Hails R), pp. 395-409. Blackwell Sciences, Oxford
- Brown J.K. & Hovmøller M.S. (2002) Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science*, 297, 537-41
- Buckley J., Butlin R.K. & Bridle J.R. (2012) Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Mol Ecol*, 21, 267-80
- Buerkle C.A. (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology Notes*, 5, 684-687
- Buggs R.J. (2007) Empirical study of hybrid zone movement. *Heredity (Edinb)*, 99, 301-12
- Burdon J.J. & Thrall P.H. (1999) Spatial and temporal patterns in coevolving plant and pathogen associations. *Am Nat*, 153(Suppl), S15-S33
- Burke J.M. & Arnold M.L. (2001) Genetics and the fitness of hybrids. *Annu Rev Genet*, 35, 31-52
- Burke J.M., Carney S.E. & Arnold M.L. (1998) Hybrid fitness in the Louisiana irises: analysis of parental and F1 performance. *Evolution*, 52, 37-43
- Bus V., Rikkerink E., Aldwinckle H.S., Caffier V., Durel C.E., Gardiner S., Gessler C., Groenwold R., Laurens F., Le Cam B., Luby J., Meulenbroek B., Kellerhals M., Parisi L., Patocchi A., Plummer K., Schouten H.J., Tartarini S. & van de Weg W.E. (2009) A proposal for the nomenclature of *Venturia inaequalis* races. *Acta Horticulturae*, 814, 739-746
- Bus V.G., Rikkerink E.H., Caffier V., Durel C.E. & Plummer K.M. (2011) Revision of the nomenclature of the differential host-pathogen interactions of *Venturia inaequalis* and *Malus*. *Annu Rev Phytopathol*, 49, 391-413
- Bus V.G.M., Plummer K.M., Rikkerink E.H.A. & Luby J. (2000) Evaluation of the pathogenicity of two scab isolates from the *Vf*-resistant apple cultivar "Baujade". In: *Integrated control of pome fruit diseases*, pp. 231-237. IOBC WPRS Bulletin
- Butlin R.K., Galindo J. & Grahame J.W. (2008) Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philos Trans R Soc Lond B Biol Sci*, 363, 2997-3007
- Caffier V., Didelot F., Pumo B., Causeur D., Durel C.E. & Parisi L. (2010) Aggressiveness of eight *Venturia inaequalis* isolates virulent or avirulent to the major resistance gene *Rvi6* on a non-*Rvi6* apple cultivar. *Plant Pathology*, 59, 1072-1080
- Caffier V., Le Cam B., Expert P., Tellier M., Devaux M., Giraud M. & Chevalier M. (2012) A new scab-like disease on apple caused by the formerly saprotrophic fungus *Venturia asperata*. *Plant Pathology*, 61, 915-924
- Calenge F., Faure A., Goerre M., Gebhardt C., Van de Weg W.E., Parisi L. & Durel C.E. (2004) Quantitative Trait Loci (QTL) Analysis Reveals Both Broad-Spectrum and Isolate-Specific QTL for Scab Resistance in an Apple Progeny Challenged with Eight Isolates of *Venturia inaequalis*. *Phytopathology*, 94, 370-9
- Cambareri E.B., Jensen B.C., Schabtach E. & Selker E.U. (1989) Repeat-induced G-C to A-T mutations in *Neurospora*. *Science*, 244, 1571-5
- Cambareri E.B., Singer M.J. & Selker E.U. (1991) Recurrence of repeat-induced point

- mutation (RIP) in *Neurospora crassa*. *Genetics*, 127, 699-710
- Carter K. (2007) Low-Chill Apples. URL (<http://ucanr.org/sites/UrbanHort/files/80158.pdf>)
- Catanzariti A.M., Dodds P.N., Lawrence G.J., Ayliffe M.A. & Ellis J.G. (2006) Haustorially expressed secreted proteins from flax rust are highly enriched for avirulence elicitors. *Plant Cell*, 18, 243-56
- Causse M.A., Fulton T.M., Cho Y.G., Ahn S.N., Chunwongse J., Wu K., Xiao J., Yu Z., Ronald P.C., Harrington S.E. & et al. (1994) Saturated molecular map of the rice genome based on an interspecific backcross population. *Genetics*, 138, 1251-74
- Cavalli-Sforza L.L. (1966) Population structure and human evolution. *Proc R Soc Lond B Biol Sci*, 164, 362-379
- Charlesworth B., Nordborg M. & Charlesworth D. (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research*, 70, 155-74
- Chevreau E. & Laurens F. (1987) The pattern of inheritance in apple: further results from leaf isozyme analysis. *Theoretical and Applied Genetics*, 75
- Chevreau E., Lespinasse Y. & Gallet M. (1985) Inherence of pollen enzymes and polyploid origin of apple. *Theoretical and Applied Genetics*, 71, 268-277
- Chou J.Y., Hung Y.S., Lin K.H., Lee H.Y. & Leu J.Y. (2010) Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biol*, 8, e1000432
- Clay K. & Kover P.X. (1996) The Red Queen Hypothesis and plant/pathogen interactions. *Annu Rev Phytopathol*, 34, 29-50
- Cloutier S., Cappadocia M. & Landry B.S. (1997) Analysis of RFLP mapping inaccuracy in *Brassica napus* L. *Theor Appl Genet* 95, 83-91
- Coart E., Van Glabeke S., De Loose M., Larsen A.S. & Roldan-Ruiz I. (2006) Chloroplast diversity in the genus *Malus*: new insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). *Molecular Ecology*, 15, 2171-2182
- Combes C. (1995) *Interactions durables: écologie et évolution du parasitisme*. Masson, Paris.
- Cornille A., Gladieux P., Smulders M.J., Roldan-Ruiz I., Laurens F., Le Cam B., Nersesyan A., Clavel J., Olonova M., Feugey L., Gabrielyan I., Zhang X.G., Tenailon M.I. & Giraud T. (2012) New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet*, 8, e1002703
- Couch B.C., Fudal I., Lebrun M.H., Tharreau D., Valent B., van Kim P., Notteghem J.L. & Kohn L.M. (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, 613-30
- Coyne J.A. & Orr H.A. (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.
- Creste S., Tulmann Neto A. & Figueira A. (2001) Detection of Single Sequence Repeat Polymorphisms in Denaturing Polyacrylamide Sequencing Gels by Silver Staining. *Plant Molecular Biology Reporter*, 19, 299-306
- Crosby J.A., Janick J., Pecknold P.C., Korban S.S., O'Connon P.A., Ries S.M., Goffreda J. & Voordeckers A. (1992) Breeding apples for scab resistance: 1945–1990. *Fruit Varieties Journal*, 46, 145–166
- Currat M. & Excoffier L. (2004) Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLoS Biol*, 2, e421
- Currat M., Ruedi M., Petit R.J. & Excoffier L. (2008) The hidden side of invasions: massive introgression by local genes. *Evolution*, 62, 1908-20
- Daguin C., Bonhomme F. & Borsa P. (2001) The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism

- at locus *mac-1*. *Heredity (Edinb)*, 86, 342-54
- Dangl J.L. & Jones J.D. (2001) Plant pathogens and integrated defence responses to infection. *Nature*, 411, 826-33
- de la Cruz F. & Davies J. (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol*, 8, 128-33
- de Meeus T., Michalakakis Y. & Renaud F. (1998) Santa rosalia revisited: or why are there so many kinds of parasites in 'the garden of earthly delights'? *Parasitol Today*, 14, 10-3
- de Vallavieille-Pope C., Ali S., Leconte M., Enjalbert J., Delos M. & Rouzet J. (2011) Virulence dynamics and regional structuring of *Puccinia striiformis* f. sp. *tritici* in France between 1984 and 2009. *Plant Disease*, 96, 131-140
- Dettman J.R. & Taylor J.W. (2004) Mutation and evolution of microsatellite loci in *Neurospora*. *Genetics*, 168, 1231-48
- Diamond J. (1997) *Guns, Germs, and Steel: The Fates of Human Societies*. Norton, W.W., New York.
- Diamond J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature*, 418, 700-707
- Didelot F., Brun L. & Parisi L. (2007) Effects of cultivar mixtures on scab control in apple orchards. *Plant Pathology*, 56, 1014-1022
- Dobrindt U. & Hacker J. (2001) Whole genome plasticity in pathogenic bacteria. *Curr Opin Microbiol*, 4, 550-7
- Dobzhansky T. (1937) *Genetics and the origin of species*. 364 pp. edn. Columbia University Press, New York.
- Doebli M. (1996) A quantitative genetic competition model for sympatric speciation. *J. Evol. Biol.*, 9, 893-909
- Dunemann F., Gläss R., Bartsch S., Saad Eldin M., Peil A. & Bus V. (2012) Molecular cloning and analysis of apple *HcrVf* resistance gene paralogs in a collection of related species. *Tree Genetics & Genomes*, 1-15
- Durand E., Jay F., Gaggiotti O.E. & Francois O. (2009) Spatial inference of admixture proportions and secondary contact zones. *Mol Biol Evol*, 26, 1963-73
- Edmands S. (1999) Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. *Evolution*, 53, 1757-1768
- Elias M., Faria R., Gompert Z. & Hendry A. (2012) Factors Influencing Progress toward Ecological Speciation. *International Journal of Ecology*
- Ellis M.A., Ferree D.C. & Madden L.V. (1998) Effects of an apple scab-resistant cultivar on use patterns of inorganic and organic fungicides and economics of disease control. *Plant Disease*, 82, 428-433
- Endler J.A. (1977) *Geographic variation, speciation, and clines*. Princeton University Press, Princeton, New Jersey.
- EUROSTAT (2007). URL (<http://epp.eurostat.ec.europa.eu/portal/page/portal/agriculture/data/database>)
- Excoffier L., Foll M. & Petit R.J. (2009a) Genetic Consequences of Range Expansions. *Ecology, Evolution, and Systematics*, 40, 481-501
- Excoffier L., Hofer T. & Foll M. (2009b) Detecting loci under selection in a hierarchically structured population. *Heredity (Edinb)*, 103, 285-98
- Falush D., Stephens M. & Pritchard J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164, 1567-87
- Falush D., Stephens M. & Pritchard J.K. (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, 7, 574-578

- FAOSTAT (2012). URL <http://faostat.fao.org/>
- Faure M.F., David P., Bonhomme F. & Bierne N. (2008) Genetic hitchhiking in a subdivided population of *Mytilus edulis*. *BMC Evol Biol*, 8, 164
- Feil E.J. & Spratt B.G. (2001) Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol*, 55, 561-90
- Felsenstein J. (1981) Skepticism Towards Santa Rosalia, or Why Are There So Few Kinds of Animals. *Evolution*, 35, 124-138
- Ferree P.M. & Barbash D.A. (2009) Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol*, 7, e1000234
- Fishman L., Kelly A.J., Morgan E. & Willis J.H. (2001) A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics*, 159, 1701-16
- Flor H.H. (1955) Host-parasite interaction in flax rust - its genetics and other implications. *Phytopathology* 45, 680-685
- Flor H.H. (1971) Current status of the gene-for-gene concept. *Annual Review of Phytopathology*, 9, 275-278
- Foll M. & Gaggiotti O. (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. *Genetics*, 180, 977-993
- Force A., Lynch M., Pickett F.B., Amores A., Yan Y.L. & Postlethwait J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151, 1531-45
- Forsline P.L. (1995) Adding diversity to the national apple germplasm collection: Collecting wild apples in Kazakhstan. *New York Fruit Quarterly*, 3, 3-6
- Forsline P.L., Aldwinckle H.S., Dickson E.E., Luby J.L. & Hokanson S.C. (2003) Collection, maintenance, characterization and utilization of wild apples of central Asia. *Horticultural Reviews*, 29
- Forsline P.L., Dickson E.E. & Luby J.L. (1994) Collection of wild *Malus*, *Vitis* and other fruit species genetic resources in Kazakhstan and neighbouring republics. *Horticultural Reviews*, 29, 2-61
- Fourcade Y., Chapat-Bardy A., Secondi J., Fleurant C. and Lemaire C. Is local selection so widespread in river organisms? Fractal geometry of river networks leads to high bias in outlier detection. *Molecular Ecology* (in press)
- Friesen T.L., Stukenbrock E.H., Liu Z., Meinhardt S., Ling H., Faris J.D., Rasmussen J.B., Solomon P.S., McDonald B.A. & Oliver R.P. (2006) Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat Genet*, 38, 953-6
- Fudal I., Ross S., Brun H., Besnard A.L., Ermel M., Kuhn M.L., Balesdent M.H. & Rouxel T. (2009) Repeat-induced point mutation (RIP) as an alternative mechanism of evolution toward virulence in *Leptosphaeria maculans*. *Mol Plant Microbe Interact*, 22, 932-41
- Funk D.J. (1998) Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. *Evolution*, 52, 1744-1759
- Gaggiotti O.E., Lee C.E. & Wardle G.E. (1997) The effect of overlapping generations and population structure on gene-frequency clines. In: *Structured-population models in marine, terrestrial and freshwater systems* (eds. Tuljapulkar S & Caswell H), pp. 355-369. Chapman & Hall, New York
- Gautier M., Flori L., Riebler A., Jaffrezic F., Laloe D., Gut I., Moazami-Goudarzi K. & Foulley J.L. (2009) A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics*, 10, 550
- Gavrilets S. (1997a) Evolution and speciation on holey adaptive landscapes. *Trends Ecol Evol*, 12, 307-12

- Gavrilets S. (1997b) Hybrid zones with Dobzhansky-type epistatic selection. *Evolution*, 51, 1027-1035
- Gavrilets S. (1997c) Single-locus clines. *Evolution*, 51, 979-983
- Gavrilets S. (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, 57, 2197-215
- Gebhardt A. (2009) akima: Interpolation of irregularly spaced data. In:
- Gessler C., Patocchi A., Sansavini S., Tartarini S. & Gianfranceschi L. (2006) *Venturia inaequalis* resistance in apple *Critical Reviews in Plant Sciences*, 25, 473-503
- Gessler C. & Pertot I. (2012) Vf scab resistance of *Malus*. *Trees-structure and function*, 26, 95-108
- Giraud T. (2006a) Speciation in parasites: host switching does not automatically lead to allopatry. *Trends Parasitol*, 22, 151-2; author reply 152-3
- Giraud T. (2006b) Speciation: selection against migrant pathogens: the immigrant inviability barrier in pathogens. *Heredity (Edinb)*, 97, 316-8
- Giraud T., Gladieux P. & Gavrilets S. (2010) Linking the emergence of fungal plant diseases with ecological speciation. *Trends Ecol Evol*, 25, 387-95
- Giraud T. & Gourbiere S. (2012) The tempo and modes of evolution of reproductive isolation in fungi. *Heredity (Edinb)*, 109, 204-14
- Giraud T., Refregier G., Le Gac M., de Vienne D.M. & Hood M.E. (2008) Speciation in fungi. *Fungal Genet Biol*, 45, 791-802
- 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
- Gislason D., Ferguson M.M., Skulason S. & Snorrason S.S. (1999) Rapid and coupled phenotypic and genetic divergence in Icelandic arctic char (*Salvelinus alpinus*). *Can. J. Fish. Aquat. Sci.*, 56, 2229-34
- Gladieux P., Guerin F., Giraud T., Caffier V., Lemaire C., Parisi L., Didelot F. & B L.E.C. (2011) Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Mol Ecol*, 20, 4521-32
- Gladieux P., Zhang X.G., Afoufa-Bastien D., Sanhueza R.M.V., Sbaghi M. & Le Cam B. (2008) On the Origin and Spread of the Scab Disease of Apple: Out of Central Asia. *Plos One*, 3
- Gladieux P., Zhang X.G., Roldan-Ruiz I., Caffier V., Leroy T., Devaux M., Van Glabeke S., Coart E. & Le Cam B. (2010) Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Molecular Ecology*, 19, 658-674
- Gomez-Alpizar L., Carbone I. & Ristaino J.B. (2007) An Andean origin of *Phytophthora infestans* inferred from mitochondrial and nuclear gene genealogies. *Proc Natl Acad Sci U S A*, 104, 3306-11
- Gompert Z. & Buerkle C.A. (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Mol Ecol*, 18, 1207-24
- Gompert Z. & Buerkle C.A. (2010) introgress: a software package for mapping components of isolation in hybrids. *Mol Ecol Resour*, 10, 378-84
- Gompert Z. & Buerkle C.A. (2011) Bayesian estimation of genomic clines. *Mol Ecol*, 20, 2111-27
- Gompert Z. & Buerkle C.A. (2012) bgc: Software for Bayesian estimation of genomic clines. *Mol Ecol Resour*
- Gompert Z., Parchman T.L. & Buerkle C.A. (2012) Genomics of isolation in hybrids. *Philos Trans R Soc Lond B Biol Sci*, 367, 439-50
- Goss E.M., Cardenas M.E., Myers K., Forbes G.A., Fry W.E., Restrepo S. & Grunwald N.J.

- (2011) The plant pathogen *Phytophthora andina* emerged via hybridization of an unknown *Phytophthora* species and the Irish potato famine pathogen, *P. infestans*. *PLoS One*, 6, e24543
- Goss E.M., Kreitman M. & Bergelson J. (2005) Genetic diversity, recombination and cryptic clades in *Pseudomonas viridiflava* infecting natural populations of *Arabidopsis thaliana*. *Genetics*, 169, 21-35
- Gourbiere S. & Mallet J. (2010) Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball". *Evolution*, 64, 1-24
- Gout L., Kuhn M.L., Vincenot L., Bernard-Samain S., Cattolico L., Barbetti M., Moreno-Rico O., Balesdent M.H. & Rouxel T. (2007) Genome structure impacts molecular evolution at the *AvrLm1* avirulence locus of the plant pathogen *Leptosphaeria maculans*. *Environ Microbiol*, 9, 2978-92
- Gracia D. (2011) Histoire évolutive de la virulence virVf en Europe chez *Venturia inaequalis*, agent de la tavelure du pommier. 28 pp. Université d'Angers, Angers
- Grayburn W.S. & Selker E.U. (1989) A natural case of RIP: degeneration of the DNA sequence in an ancestral tandem duplication. *Mol Cell Biol*, 9, 4416-21
- 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
- Guérin F. & Le Cam B. (2004) Breakdown of the scab resistance gene *Vf* in apple leads to a founder effect in populations of the fungal pathogen *Venturia inaequalis*. *Phytopathology*, 94, 364-369
- Haldane J.B.S. (1922) Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* , 12, 101–109
- Hall M.C. & Willis J.H. (2005) Transmission ratio distortion in intraspecific hybrids of *Mimulus guttatus*: implications for genomic divergence. *Genetics*, 170, 375-86
- Hamilton W.D. (1980) Sex versus non-sex parasite. *Oikos*, 35, 282–90
- Hamilton W.D., Axelrod R. & Tanese R. (1990) Sexual reproduction as an adaptation to resist parasites. *Proceedings of the National Academy of Sciences*, 87, 3566-3573
- Han Y., Liu X., Benny U., Kistler H.C. & VanEtten H.D. (2001) Genes determining pathogenicity to pea are clustered on a supernumerary chromosome in the fungal plant pathogen *Nectria haematococca*. *Plant J*, 25, 305-14
- Harris S.A., Robinson J.P. & Juniper B.E. (2002) Genetic clues to the origin of the apple. *Trends in Genetics*, 18, 426-430
- Harrison R.G. (1986) Pattern and process in a narrow hybrid zone. *Heredity*, 56, 337–349
- Harrison R.G. (1990) Hybrid zones: windows on evolutionary processes. *Oxf. Surv. Evol. Biol.*, 7, 69-128
- Harushima Y., Nakagahra M., Yano M., Sasaki T. & Kurata N. (2001) A genome-wide survey of reproductive barriers in an intraspecific hybrid. *Genetics*, 159, 883-92
- Hatfield T. & Schluter D. (1999) Ecological speciation in sticklebacks: Environment-dependent hybrid fitness. *Evolution*, 53, 866-873
- Hatta R., Ito K., Hosaki Y., Tanaka T., Tanaka A., Yamamoto M., Akimitsu K. & Tsuge T. (2002) A conditionally dispensable chromosome controls host-specific pathogenicity in the fungal plant pathogen *Alternaria alternata*. *Genetics*, 161, 59-70
- Hatta R., Shinjo A., Ruswandi S., Kitani K., Yamamoto M., Akimitsu K. & Tsuge T. (2006) DNA transposon fossils present on the conditionally dispensable chromosome controlling AF-toxin biosynthesis and pathogenicity of *Alternaria alternata*. *J. Gen. Plant Pathol*, 72, 210–19
- Heath D.D., Shrimpton J.M., Hepburn R.I., Jamieson S.K., Brode S.K. & Docker M.F. (2006) Population structure and divergence using microsatellite and gene locus markers in

- Chinook salmon (*Oncorhynchus tshawytscha*) populations. *Canadian Journal of Fisheries and Aquatic Sciences*, 63, 1370–1383
- Hendry A.P. (2009) Ecological speciation! Or the lack thereof? *Can. J. Fish. Aquat. Sci.*, 66, 1383–1398
- Hewitt G.M. (1988) Hybrid zones-natural laboratories for evolutionary studies. *Trends Ecol Evol*, 3, 158-67
- Hirsch N., Zimmerman L.B. & Grainger R.M. (2002) Xenopus, the next generation: *X. tropicalis* genetics and genomics. *Dev Dyn*, 225, 422-33
- Hoekstra H.E., Drumm K.E. & Nachman M.W. (2004) Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*, 58, 1329-41
- Hofer T., Ray N., Wegmann D. & Excoffier L. (2009) Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Ann Hum Genet*, 73, 95-108
- Hofman S., Spolsky C., Uzzell T., Cogalniceanu D., Babik W. & Szymura J.M. (2007) Phylogeography of the fire-bellied toads *Bombina*: independent Pleistocene histories inferred from mitochondrial genomes. *Mol Ecol*, 16, 2301-16
- Holb I.J. (2006) Effect of six sanitation treatments on leaf litter density, ascospore production of *Venturia inaequalis* and scab incidence in integrated and organic apple orchards. *European Journal of Plant Pathology*, 115, 293-307
- Holb I.J., DeJong P.F. & Heijne B. (2003a) Efficacy and phytotoxicity of lime sulfur in organic apple production. *Annals of Applied Biology*, 142, 225–233
- Holb I.J., Heijne B. & Jeger M.J. (2003b) Summer epidemics of apple scab: Their relationship between measurements and their implications for the development of predictive models and threshold levels under different disease control regimes. *Journal of Phytopathology* 151, 335–343
- Hough L.F. (1944) A survey of the scab resistance of the foliage on seedlings in selected apple progenies. *Proc. Amer. Soc. Hort. Sci.*, 4, 260-272
- Hough L.F., Williams E.B., Goonewardene H.F., Janick J. & Korban S.S. (1953) Apple scab resistance from *Malus floribunda* Sieb. *Proceeding of the American Society for Horticultural Sciences*, 62, 341-347
- Hough L.F., Williams E.B., Goonewardene H.F., Janick J. & Korban S.S. (1988) Progress in developing environmentally and commercially acceptable apple cultivars. In: *International Symposium on Horticultural Germplasm and Wild*, pp. 205-208, Beijing, China
- Huang Y.-J., Evans N., Fitt B.D.L., Li Z.-Q., Rouxel T. & Balesdent M.H. (2006) Fitness cost associated with loss of the *AvrLm4* avirulence function in *Leptosphaeria maculans* (phoma stem canker of oilseed rape). In: *Sustainable strategies for managing Brassica napus (oilseed rape) resistance to Leptosphaeria maculans (phoma stem canker)* (eds. Fitt BDL, Evans N, Howlett BJ & Cooke BM). Springer
- Hubisz M.J., Falush D., Stephens M. & Pritchard J.K. (2009) Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour*, 9, 1322-32
- Inderbitzin P., Davis R.M., Bostock R.M. & Subbarao K.V. (2011) The ascomycete *Verticillium longisporum* is a hybrid and a plant pathogen with an expanded host range. *PLoS One*, 6, e18260
- Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N., Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A., Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B., Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S., Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B., Biemont C., Skalli Z.,

- Cattolico L., Poulain J., De Berardinis V., Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J., Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S., Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J., Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M., Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C., Wincker P., Lander E.S., Weissenbach J. & Roest Crolius H. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, 431, 946-57
- James T.Y., Kauff F., Schoch C.L., Matheny P.B., Hofstetter V., Cox C.J., Celio G., Gueidan C., Fraker E., Miadlikowska J., Lumbsch H.T., Rauhut A., Reeb V., Arnold A.E., Amtoft A., Stajich J.E., Hosaka K., Sung G.H., Johnson D., O'Rourke B., Crockett M., Binder M., Curtis J.M., Slot J.C., Wang Z., Wilson A.W., Schussler A., Longcore J.E., O'Donnell K., Mozley-Standridge S., Porter D., Letcher P.M., Powell M.J., Taylor J.W., White M.M., Griffith G.W., Davies D.R., Humber R.A., Morton J.B., Sugiyama J., Rossman A.Y., Rogers J.D., Pfister D.H., Hewitt D., Hansen K., Hambleton S., Shoemaker R.A., Kohlmeyer J., Volkmann-Kohlmeyer B., Spotts R.A., Serdani M., Crous P.W., Hughes K.W., Matsuura K., Langer E., Langer G., Untereiner W.A., Lucking R., Budel B., Geiser D.M., Aptroot A., Diederich P., Schmitt I., Schultz M., Yahr R., Hibbett D.S., Lutzoni F., McLaughlin D.J., Spatafora J.W. & Vilgalys R. (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, 443, 818-22
- Janick J. (2006) The PRI apple breeding program. *Hort-Science*, 41, 8-10
- Janick J., Cummins J.N., Brown S.K. & Hemmat M. (1996) Apples. In: *Fruit Breeding* (eds. Janick J & Moore JN). Wiley, New York
- Jenczewski E., Gherardi M., Bonnin I., Prospero J.-M., Olivieri I. & Huguet T. (1997) Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). *Theor. Appl. Genet*, 94, 682-691.
- Jha G., Thakur K. & Thakur P. (2009) The *Venturia* apple pathosystem: pathogenicity mechanisms and plant defense responses. *J Biomed Biotechnol*, 2009, 680160
- Jiang C.X., Chee P.W., Draye X., Morrell P.L., Smith C.W. & Paterson A.H. (2000) Multilocus interactions restrict gene introgression in interspecific populations of polyploid *Gossypium* (cotton). *Evolution*, 54, 798-814
- Johnson P.A. & Gullberg U. (1998) Theory and models of sympatric speciation. In: *Endless Forms: Species and Speciation* (eds. Howard DJ & Berlocher SH), pp. 79-89. Oxford University press, Oxford
- Jones D.A., Thomas C.M., Hammond-Kosack K.E., Balintkurti P.J. & Jones J.D.G. (1994) Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. *Science*, 266, 789-793
- Jones J.D. & Dangl J.L. (2006) The plant immune system. *Nature*, 444, 323-9
- Joosten M.H., Vogelsang R., Cozijnsen T.J., Verberne M.C. & De Wit P.J. (1997) The biotrophic fungus *Cladosporium fulvum* circumvents *Cf-4*-mediated resistance by producing unstable AVR4 elicitors. *Plant Cell*, 9, 367-79
- Juniper B.E. & Maberley D.J. (2006) *The story of the apple*. Timber Press, Grantham, United Kingdom.
- Key K.H. (1968) The concept of stasipatric speciation. *Syst. zool.*, 17, 14-22
- Kim Y. & Stephan W. (2002) Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, 160, 765-77
- Kimura M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.
- Kirkpatrick M. & Ravigné V. (2002) Speciation by natural and sexual selection: models and experiments. *Am Nat*, 159 Suppl 3, S22-35

- Klopfstein S., Currat M. & Excoffier L. (2006) The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol*, 23, 482-90
- Kolaczkowski B., Kern A.D., Holloway A.K. & Begun D.J. (2010) Genomic differentiation between temperate and tropical Australian populations of *Drosophila melanogaster*. *Genetics*, 187, 245-60
- Köller W. & Parker D.M. (1989) Purification and characterization of cutinase from *Venturia inaequalis*. *Phytopathology*, 79, 278–283
- Köller W., Parker D.M. & Becker C.M. (1991) Role of cutinase in the penetration of apple leaves by *Venturia inaequalis*. *Phytopathology*, 81, 1375–1379
- Kolmer A., Long D.L. & Hughes M.E. (2005) Physiologic Specialization of *Puccinia triticina* on Wheat in the United States in 2003. *Plant Disease*, 89, 1201-1206
- Kolmer J.A. (1997) Virulence dynamics and genetics of cereal rust populations in North America. In: *The gene-for-gene relationship in plant-parasite interactions* (eds. Crute IR, Holub EB & Burdon JJ), pp. 139–156. CAB International, New York.
- Kolmer J.A. (2003) Postulation of Leaf Rust Resistance Genes in Selected Soft Red Winter Wheats. *Crop Science*, 43, 1266-1274
- Kondrashov A.S. (1986) Multilocus model of sympatric speciation. III. Computer simulations. *Theor Popul Biol*, 29, 1-15
- Kondrashov F.A. (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc Biol Sci*
- Korban S.S. & Skirvin R.M. (1984) Nomenclature of the cultivated apple. *Hort-Science*, 19, 177-180
- Koutinas N., Pepelyankov G. & Lichev V. (2010) Flower induction and flower bud development in apple and sweet cherry. *Biotechnology & Biotechnological Equipment*, 24, 1549-1558
- Kozak G.M., Rudolph A.B., Colon B.L. & Fuller R.C. (2012) Postzygotic Isolation Evolves before Prezygotic Isolation between Fresh and Saltwater Populations of the Rainwater Killifish, *Lucania parva*. *Int J Evol Biol*, 2012, 523967
- Kruuk L.E., Baird S.J., Gale K.S. & Barton N.H. (1999a) A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics*, 153, 1959-71
- Kruuk L.E.B., Gilchrist J.S. & Barton N.H. (1999b) Hybrid dysfunction in fire-bellied toads (*Bombina*). *Evolution*, 53, 1611-1616
- Ky C.L., Barre P., Lorieux M., Trouslot P., Akaffou S., Louarn J., Charrier A., Hamon S. & Noirot M. (2000) Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea* sp.). *Theoretical and Applied Genetics*, 101, 669-676
- Lai M.M. (1992) Genetic recombination in RNA viruses. *Curr Top Microbiol Immunol*, 176, 21-32
- Langerhans R.B., Gifford M.E. & Joseph E.O. (2007) Ecological speciation in *Gambusia* fishes. *Evolution*, 61, 2056-74
- Le Cam B., Devaux M. & Parisi L. (2001) Specific polymerase chain reaction identification of *Venturia nashicola* using internally transcribed spacer region in the ribosomal DNA. *Phytopathology*, 91, 900-904
- Le Cam B., Parisi L. & Arene L. (2002) Evidence of Two Formae Speciales in *Venturia inaequalis*, Responsible for Apple and Pyracantha Scab. *Phytopathology*, 92, 314-20
- Le Gac M. & Giraud T. (2008) Existence of a pattern of reproductive character displacement in Homobasidiomycota but not in Ascomycota. *J Evol Biol*, 21, 761-72
- Le Gac M., Hood M.E. & Giraud T. (2007) Evolution of reproductive isolation within a parasitic fungal species complex. *Evolution*, 61, 1781-7
- Lê Van A., Durel C.E., Le Cam B. & Caffier V. (2011) The threat of wild habitat to scab

- resistant apple cultivars. *Plant Pathology*, 60, 621-630
- Lê Van A., Gladieux P., Lemaire C., Cornille A., Giraud T., Durel C.-E., Caffier V. & Le Cam B. (2012) Evolution of pathogenicity traits in the apple scab fungal pathogen in response to the domestication of its host. *Evolutionary Applications* (in press)
- Leach J.E., Vera Cruz C.M., Bai J. & Leung H. (2001) Pathogen fitness penalty as a predictor of durability of disease resistance genes. *Annu Rev Phytopathol*, 39, 187-224
- Lee H.Y., Chou J.Y., Cheong L., Chang N.H., Yang S.Y. & Leu J.Y. (2008) Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell*, 135, 1065-73
- Lenormand T., Bourguet D., Guillemaud T. & Raymond M. (1999) Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. *Nature*, 400, 861-4
- Lepinasse Y. (1990) Le pommier. In: *Amélioration des espèces végétales cultivées*, pp. 580-592. INRA éditions, Paris
- Lewontin R.C. & Krakauer J. (1973) Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. *Genetics*, 74, 175-195
- Li W.H. & Nei M. (1977) Persistence of common alleles in two related populations or species. *Genetics*, 86, 901-14
- Li Z.K., Luo L.J., Mei H.W., Paterson A.H., Zhao X.H., Zhong D.B., Wang Y.P., Yu X.Q., Zhu L., Tabien R., Stansel J.W. & Ying C.S. (1999) A "defeated" rice resistance gene acts as a QTL against a virulent strain of *Xanthomonas oryzae* pv. *oryzae*. *Mol Gen Genet*, 261, 58-63
- Lindhout P. (2002) The perspectives of polygenic resistance in breeding for durable disease resistance. *Euphytica*, 124, 217-226
- Lorieux M. (2012) MapDisto: fast and efficient computation of genetic linkage maps. *Molecular Breeding*, 30, 1231-1235
- Lynch M. (2007a) The evolution of genetic networks by non-adaptive processes. *Nat Rev Genet*, 8, 803-13
- Lynch M. (2007b) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc Natl Acad Sci U S A*, 104 Suppl 1, 8597-604
- Lynch M. & Conery J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science*, 290, 1151-5
- Lynch M. & Force A. (2000) Gene duplication and the origin of interspecific genomic incompatibility. *Am Nat*, 156, 590-605
- Lynch M., O'Hely M., Walsh B. & Force A. (2001) The probability of preservation of a newly arisen gene duplicate. *Genetics*, 159, 1789-804
- Lynch M., Sung W., Morris K., Coffey N., Landry C.R., Dopman E.B., Dickinson W.J., Okamoto K., Kulkarni S., Hartl D.L. & Thomas W.K. (2008) A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A*, 105, 9272-7
- Lynch M. & Walsh J.B. (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland.
- Ma L.J., van der Does H.C., Borkovich K.A., Coleman J.J., Daboussi M.J., Di Pietro A., Dufresne M., Freitag M., Grabherr M., Henrissat B., Houterman P.M., Kang S., Shim W.B., Woloshuk C., Xie X., Xu J.R., Antoniw J., Baker S.E., Bluhm B.H., Breakspear A., Brown D.W., Butchko R.A., Chapman S., Coulson R., Coutinho P.M., Danchin E.G., Diener A., Gale L.R., Gardiner D.M., Goff S., Hammond-Kosack K.E., Hilburn K., Hua-Van A., Jonkers W., Kazan K., Kodira C.D., Koehrsen M., Kumar L., Lee Y.H., Li L., Manners J.M., Miranda-Saavedra D., Mukherjee M., Park G., Park J., Park S.Y., Proctor R.H., Regev A., Ruiz-Roldan M.C., Sain D., Sakthikumar S., Sykes S., Schwartz D.C., Turgeon B.G., Wapinski I., Yoder O., Young S., Zeng Q., Zhou S.,

- Galagan J., Cuomo C.A., Kistler H.C. & Rep M. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*, 464, 367-73
- MacCallum C.J., Nürnberg B., Barton N.H. & Szymura J.M. (1998) Habitat preference in the *Bombina* hybrid zone in Croatia. *Evolution*, 52, 227-239
- MacHardy W.E. (1996) *Apple scab: Biology, Epidemiology and Management*. APS Press, St. Paul, MN.
- MacHardy W.E., Gadoury D.M. & Gessler C. (2001) Parasitic and biological fitness of *Venturia inaequalis*: relationship to disease management strategies. *Plant Disease*, 85, 1036-1051
- Maheshwari S. & Barbash D.A. (2011) The genetics of hybrid incompatibilities. *Annu Rev Genet*, 45, 331-55
- Maliepaard C., Alston F.H., van Arkel G., Brown L.M., Chevreau E., Dunemann F., Evans K.M., Gardiner S., Guilford P., van Heusden A.W., Janse A.W., Laurens F., Lynn J.R., Manganaris A.G., den Nijs A.P.M., Periam N., Rikkerink E., Roche P., Ryder C., Sansavini S., Schmidt H., Tartarini S., Verhaegh J.J., Vrieling-van Ginkel M. & King G.J. (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theoretical and Applied Genetics*, 97, 60-73
- Mallet J., Barton N., Lamas G., Santisteban J., Muedas M. & Eeley H. (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. *Genetics*, 124, 921-36
- Marques A.C., Vinckenbosch N., Brawand D. & Kaessmann H. (2008) Functional diversification of duplicate genes through subcellular adaptation of encoded proteins. *Genome Biol*, 9, R54
- Masly J.P., Jones C.D., Noor M.A., Locke J. & Orr H.A. (2006) Gene transposition as a cause of hybrid sterility in *Drosophila*. *Science*, 313, 1448-50
- Matute D.R., Butler I.A., Turissini D.A. & Coyne J.A. (2011) A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science*, 329, 1518-21
- Mayden R.L. (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. In: *Species: the Units of Diversity* (eds. Claridge MA, Dawah HA & Wilson MR), pp. 381-424. Chapman & Hall, London
- Maynard Smith J. (1966) Sympatric speciation. *American Naturalist*, 100, 637-650
- Maynard Smith J. & Haigh J. (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, 23, 23-35
- Mayr E. (1942) *Systematics and the Origin of Species*. Columbia University Press, New York.
- Mayr E. (1963) *Animal Species and Evolution*. Harvard University Press, Cambridge.
- Mayr E. (1970) *Populations, Species and Evolution*. Harvard University Press, Cambridge.
- McCoy K.D. (2003) Sympatric speciation in parasites--what is sympatry? *Trends Parasitol*, 19, 400-4
- McDonald B.A. (2010) How can we achieve durable disease resistance in agricultural ecosystems. *New Phytologist*, 185, 3-5
- McDonald B.A. & Linde C. (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, 40, 349-79
- McIntosh R.A. & Brown G.N. (1997) Anticipatory breeding for resistance to rust diseases in wheat. *Annu Rev Phytopathol*, 35, 311-26
- Meirmans P.G. (2012) The trouble with isolation by distance. *Mol Ecol*, 21, 2839-46
- Menon R. (1956) Studies on Venturiaceae on Rosaceous Plants. *Phytopath. Z.*, 27, 117-146
- Mihola O., Trachtulec Z., Vlcek C., Schimenti J.C. & Forejt J. (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science*, 323, 373-5
- Mizuta Y., Harushima Y. & Kurata N. (2010) Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. *Proc Natl Acad Sci U S A*, 107, 20417-22

- Montarry J., Cartier E., Jacquemond M., Palloix A. & Moury B. (2012) Virus adaptation to quantitative plant resistance: erosion or breakdown? *J Evol Biol* (in press)
- Montarry J., Hamelin F.M., Glais I., Corbi R. & Andrivon D. (2010) Fitness costs associated with unnecessary virulence factors and life history traits: evolutionary insights from the potato late blight pathogen *Phytophthora infestans*. *BMC Evol Biol*, 10, 283
- Moore W.S. (1977) An evaluation of narrow hybrid zones in vertebrates. *The Quarterly Review of Biology*, 52, 263-277
- Moran C. (1981) Genetic demarcation of geographical distribution by hybrid zones. *Proc. Ecol. Soc. Aust.*, 11, 67-73
- Moran N.A. & Jarvik T. (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*, 328, 624-7
- Morgan J. & Richards A. (1993) *The book of Apples*. Brogdale Horticultural Trust in association with Ebury Press, London.
- Morris C.E., Bardin M., Kinkel L.L., Moury B., Nicot P.C. & Sands D.C. (2009) Expanding the Paradigms of Plant Pathogen Life History and Evolution of Parasitic Fitness beyond Agricultural Boundaries. *Plos Pathogens*, 5
- Morris C.E., Kinkel L.L., Kun X., Prior P. & Sands D.C. (2007) A surprising niche for the plant pathogen *Pseudomonas syringae*. *Infection, Genetics and Evolution*, 7, 84-92
- Moullia C., Le Brun N., Dallas J., Orth A. & Renaud F. (1993) Experimental evidence of genetic determinism in high susceptibility to intestinal pinworm infection in mice: a hybrid zone model. *Parasitology*, 106, 387-393
- Moyle L.C. & Nakazato T. (2010) Hybrid incompatibility "snowballs" between *Solanum* species. *Science*, 329, 1521-3
- Muller H.J. (1942) Isolating mechanisms, evolution and temperature. *Biol Symp*, 6, 71-125
- Munkacsi A.B., Stoxen S. & May G. (2007) Domestication of maize, sorghum, and sugarcane did not drive the divergence of their smut pathogens. *Evolution*, 61, 388-403
- Munkacsi A.B., Stoxen S. & May G. (2008) *Ustilago maydis* populations tracked maize through domestication and cultivation in the Americas. *Proc Biol Sci*, 275, 1037-46
- Myburg A.A., Vogl C., Griffin A.R., Sederoff R.R. & Whetten R.W. (2004) Genetics of postzygotic isolation in *Eucalyptus*: whole-genome analysis of barriers to introgression in a wide interspecific cross of *Eucalyptus grandis* and *E. globulus*. *Genetics*, 166, 1405-18
- Nagel L. & Schluter D. (1998) Body size, natural selection, and speciation in sticklebacks. *Evolution*, 52, 209-218
- Narum S.R., Banks M., Beacham T.D., Bellinger M.R., Campbell M.R., Dekoning J., Elz A., Guthrie C.M., 3rd, Kozfkay C., Miller K.M., Moran P., Phillips R., Seeb L.W., Smith C.T., Warheit K., Young S.F. & Garza J.C. (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Mol Ecol*, 17, 3464-77
- Nei M. (1976) Mathematical models of speciation and genetic distance. In: *Population genetics and ecology* (eds. Karlin S & Nevo E), pp. 723-765. Academic Press, New York
- Nei M. & Maruyama T. (1975) Letters to the editors: Lewontin-Krakauer test for neutral genes. *Genetics*, 80, 395
- Nei M., Maruyama T. & Wu C.I. (1983) Models of evolution of reproductive isolation. *Genetics*, 103, 557-79
- Nei M. & Nozawa M. (2011) Roles of mutation and selection in speciation: from Hugo de Vries to the modern genomic era. *Genome Biol Evol*, 3, 812-29
- Newcombe G., Stirling B., McDonald S.K. & Bradshaw H.D. (2000) *Melampsora x columbiana*, a natural hybrid of *M. medusae* and *M. occidentalis*. *Mycol. Res.*, 104

- Nielsen R. (2005) Molecular signatures of natural selection. *Annu Rev Genet*, 39, 197-218
- Nosil P., Funk D.J. & Ortiz-Barrientos D. (2009) Divergent selection and heterogeneous genomic divergence. *Mol Ecol*, 18, 375-402
- Nosil P., Vines T.H. & Funk D.J. (2005) Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, 59, 705-19
- Ohno S. (1970) *Evolution by gene duplication*. Springer, Berlin.
- Ohno S. (1998) The notion of the Cambrian pananimalia genome and a genomic difference that separated vertebrates from invertebrates. *Prog Mol Subcell Biol*, 21, 97-117
- Oka H.I. (1953) The mechanisms of sterility in the intervarietal hybrids. Phylogenetic differentiation of cultivated rice. *Japan J Breed*, 2, 217-224
- Oka H.I. (1957) Genic analysis for the sterility of hybrids between distantly related varieties of cultivated rice. *J Genet*, 55, 397-409
- Oka H.I. (1974) Analysis of genes controlling F1 sterility in rice by the use of isogenic lines. *Genetics*, 77, 521-534
- Orbach M.J., Farrall L., Sweigard J.A., Chumley F.G. & Valent B. (2000) A telomeric avirulence gene determines efficacy for the rice blast resistance gene Pi-ta. *Plant Cell*, 12, 2019-32
- Orr H.A. (1995) The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*, 139, 1805-13
- Orr H.A. & Turelli M. (2001) The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution*, 55, 1085-94
- Ortiz-Barrientos D., Reiland J., Hey J. & Noor M.A. (2002) Recombination and the divergence of hybridizing species. *Genetica*, 116, 167-78
- Otto S.P. & Whitton J. (2000) Polyploid incidence and evolution. *Annu Rev Genet*, 34, 401-437
- Palopoli M.F. & Wu C.I. (1996) Rapid evolution of a coadapted gene complex: evidence from the Segregation Distorter (SD) system of meiotic drive in *Drosophila melanogaster*. *Genetics*, 143, 1675-88
- Pardo-Manuel de Villena F., de la Casa-Esperon E., Briscoe T.L. & Sapienza C. (2000) A genetic test to determine the origin of maternal transmission ratio distortion. Meiotic drive at the mouse *Om* locus. *Genetics*, 154, 333-342
- Pariaud B., Ravigne V., Halkett F., Goyeau H., Carlier J. & Lannou C. (2009) Aggressiveness and its role in the adaptation of plant pathogens. *Plant Pathology*, 53, 409-424
- Paris M. & Despres L. (2010) In silico fingerprinting (ISIF): a user-friendly in silico AFLP program. *Methods Mol Biol*, 888, 55-64
- Parisi L., Durel C.E. & Laurens F. (2000) First report on the presence of *Venturia inaequalis* race 7 in french apple orchards. In: *Integrated control of pome fruit disease* (ed. Parisi L), pp. 99-104, Gent
- Parisi L., Laurens F., Guillaumes J. & Kruger J. (1993) A new race of *Venturia inaequalis* virulent to apples with resistance due to the *Vf* gene. *Phytopathology*, 83, 533-537
- Parlevliet J.E. (2002) Durability of resistance against fungal, bacterial and viral pathogens; present situation. 124, 147-156
- Patocchi A., Gianfranceschi L. & Gessler C. (1999a) Towards the map-based cloning of *Vf*: fine and physical mapping of the *Vf* region. *Theoretical and Applied Genetics*, 99, 1012-1017
- Patocchi A., Vinatzer B.A., Gianfranceschi L., Tartarini S., Zhang H.B., Sansavini S. & Gessler C. (1999b) Construction of a 550kb BAC contig spanning the genomic region containing the apple scab resistance gene *Vf*. *Molecular and General Genetics*, 262
- Patton J.L. (1993) Hybridization and hybrid zones in pocket gophers. In: *Hybrid zones and the Evolutionary Process* (ed. Harrison RG), pp. 290-308. Oxford University Press,

Oxford

- Payseur B.A. (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Mol Ecol Resour*, 10, 806-20
- Peressotti E., Wiedemann-Merdinoglu S., Delmotte F., Bellin D., Di Gaspero G., Testolin R., Merdinoglu D. & Mestre P. (2010) Breakdown of resistance to grapevine downy mildew upon limited deployment of a resistant variety. *BMC Plant Biol*, 10, 147
- Perez-Figueroa A., Garcia-Pereira M.J., Saura M., Rolan-Alvarez E. & Caballero A. (2010) Comparing three different methods to detect selective loci using dominant markers. *J Evol Biol*, 23, 2267-76
- Phadnis N. & Orr H.A. (2009) A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science*, 323, 376-9
- Phipps J.B., Robertson K.R., Rohrer J.R. & Smith P.G. (1991) Origins and evolution of subfam. *Maloideae* (Rosaceae). *Systematic Botany*, 16, 303-332
- Planade B., Lena J.P., Li H., Plenet S., Guegan J.F., Thomas F., Hurtrez-Bousses S., Renaud F. & Joly P. (2009) Tracking a heterosis effect in the field: tadpole resistance to parasites in the water frog hybridogenetic complex. *Parasitology*, 136, 1003-13
- Powell A.J., Conant G.C., Brown D.E., Carbone I. & Dean R.A. (2008) Altered patterns of gene duplication and differential gene gain and loss in fungal pathogens. *BMC Genomics*, 9, 147
- Presgraves D.C., Balagopalan L., Abmayr S.M. & Orr H.A. (2003) Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature*, 423, 715-9
- Pritchard J.K., Stephens M. & Donnelly P. (2000a) Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959
- Pritchard J.K., Stephens M., Rosenberg N.A. & Donnelly P. (2000b) Association mapping in structured populations. *American Journal of Human Genetics*, 67, 170-181
- Qiu W. & Moyer J.W. (1999) Tomato Spotted Wilt Tospovirus Adapts to the TSWV N Gene-Derived Resistance by Genome Reassortment. *Phytopathology*, 89, 575-82
- R Development Core Team. R (2011) A language and environment for statistical computing. URL <http://www.R-project.org>.
- Raabe R.D. & Gardner N.W. (1972) Scab of Pyracantha, Loquat, Toyon, and Kageneckia. *Phytopathology*, 62, 914-916
- Raffaele S. & Kamoun S. (2012) Genome evolution in filamentous plant pathogens: why bigger can be better. *Nat Rev Microbiol*, 10, 417-30
- Redon R., Ishikawa S., Fitch K.R., Feuk L., Perry G.H., Andrews T.D., Fiegler H., Shapero M.H., Carson A.R., Chen W., Cho E.K., Dallaire S., Freeman J.L., Gonzalez J.R., Gratacos M., Huang J., Kalaitzopoulos D., Komura D., MacDonald J.R., Marshall C.R., Mei R., Montgomery L., Nishimura K., Okamura K., Shen F., Somerville M.J., Tchinda J., Valsesia A., Woodwark C., Yang F., Zhang J., Zerjal T., Zhang J., Armengol L., Conrad D.F., Estivill X., Tyler-Smith C., Carter N.P., Aburatani H., Lee C., Jones K.W., Scherer S.W. & Hurles M.E. (2006) Global variation in copy number in the human genome. *Nature*, 444, 444-54
- Rep M., van der Does H.C., Meijer M., van Wijk R., Houterman P.M., Dekker H.L., de Koster C.G. & Cornelissen B.J. (2004) A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for I-3-mediated resistance in tomato. *Mol Microbiol*, 53, 1373-83
- Rice W.R. (1984) Disruptive selection on habitat preference and the evolution of reproductive isolation: a simulation study. *Evolution*, 38, 1251-1260
- Rieseberg L.H. (2001) Chromosomal rearrangements and speciation. *Trends Ecol Evol*, 16, 351-358

- Rieseberg L.H., Linder C.R. & Seiler G.J. (1995) Chromosomal and genic barriers to introgression in *Helianthus*. *Genetics*, 141, 1163-71
- Rieseberg L.H., Whitton J. & Gardner K. (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, 152, 713-27
- Roberts A.L. & Crute I.R. (1994) A scab resistance from *Malus floribunda* 821 (Vf) is rendered ineffective by isolates of *Venturia inaequalis* from *Malus floribunda*. *Norwegian Journal of Agricultural Science*, 17, 403-406
- Robertson A. (1975a) Gene frequency distributions as a test of selective neutrality. *Genetics*, 81, 775-85
- Robertson A. (1975b) Letters to the editors: Remarks on the Lewontin-Krakauer test. *Genetics*, 80, 396
- Robinson J.P., Harris S.A. & Juniper B.E. (2001) Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant Systematic and Evolution*, 226, 35-58
- Rodin S.N., Parkhomchuk D.V. & Riggs A.D. (2005) Epigenetic changes and repositioning determine the evolutionary fate of duplicated genes. *Biochemistry (Mosc)*, 70, 559-67
- Rodin S.N. & Riggs A.D. (2003) Epigenetic silencing may aid evolution by gene duplication. *J Mol Evol*, 56, 718-29
- Roux C. (2010) Effets de la sélection naturelle et de l'histoire démographique sur les patrons de polymorphisme nucléaire. 135 pp. Université de Lille I, Lille
- Roux C., Castric V., Pauwels M., Wright S.I., Saumitou-Laprade P. & Vekemans X. (2011) Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? *PLoS One*, 6, e26872
- Rouxel T., Grandaubert J., Hane J.K., Hoede C., van de Wouw A.P., Couloux A., Dominguez V., Anthouard V., Bally P., Bourras S., Cozijnsen A.J., Ciuffetti L.M., Degraeve A., Dilmaghani A., Duret L., Fudal I., Goodwin S.B., Gout L., Glaser N., Linglin J., Kema G.H., Lapalu N., Lawrence C.B., May K., Meyer M., Ollivier B., Poulain J., Schoch C.L., Simon A., Spatafora J.W., Stachowiak A., Turgeon B.G., Tyler B.M., Vincent D., Weissenbach J., Amselem J., Quesneville H., Oliver R.P., Wincker P., Balesdent M.H. & Howlett B.J. (2011) Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by Repeat-Induced Point mutations. *Nat Commun*, 2, 202
- Rundle H.D. (2002) A test of ecologically dependent postmating isolation between sympatric sticklebacks. *Evolution*, 56, 322-9
- Rundle H.D., Nagel L., Wenrick Boughman J. & Schluter D. (2000) Natural selection and parallel speciation in sympatric sticklebacks. *Science*, 287, 306-8
- Rundle H.D. & Nosil P. (2005) Ecological speciation. *Ecology Letters*, 8, 336-352
- Sapoukhina N., Durel C.E. & Le Cam B. (2009) Spatial deployment of gene-for-gene resistance governs evolution and spread of pathogen populations. *Theoretical Ecology*, 2, 229-238
- Sawamura K. & Yamamoto M.T. (1997) Characterization of a reproductive isolation gene, Zygotic hybrid rescue, of *Drosophila melanogaster* by using minichromosomes. *Heredity*, 79, 97-103
- Scannell D.R., Byrne K.P., Gordon J.L., Wong S. & Wolfe K.H. (2006) Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature*, 440, 341-5
- Schardl C.L. & Craven K.D. (2003) Interspecific hybridization in plant-associated fungi and oomycetes: a review. *Mol Ecol*, 12, 2861-73
- Schemske D.W. (2010) Adaptation and the origin of species. *Am Nat*, 176 Suppl 1, S4-S25
- Schlieffen U.K., Tautz D. & Paabo S. (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature*, 368, 629-32

- Schluter D. (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.
- Schluter D. (2001) Ecology and the origin of species. *Trends Ecol Evol*, 16, 372-380
- Schluter D. (2009) Evidence for ecological speciation and its alternative. *Science*, 323, 737-41
- Schluter D. & Conte G.L. (2009) Genetics and ecological speciation. *Proc Natl Acad Sci U S A*, 106 Suppl 1, 9955-62
- Schmidt P.S., Zhu C.T., Das J., Batavia M., Yang L. & Eanes W.F. (2008) An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*, 105, 16207-11
- Schnabel G., Schnabel E.L. & Jones A.L. (1999) Characterization of Ribosomal DNA from *Venturia inaequalis* and Its Phylogenetic Relationship to rDNA from Other Tree-Fruit *Venturia* Species. *Phytopathology*, 89, 100-8
- Schwarz-Sommer Z., de Andrade Silva E., Berndtgen R., Lonng W.E., Muller A., Nindl I., Stuber K., Wunder J., Saedler H., Gubitza T., Borking A., Golz J.F., Ritter E. & Hudson A. (2003) A linkage map of an F2 hybrid population of *Antirrhinum majus* and *A. molle*. *Genetics*, 163, 699-710
- Selker E.U., Cambareri E.B., Jensen B.C. & Haack K.R. (1987) Rearrangement of duplicated DNA in specialized cells of *Neurospora*. *Cell*, 51, 741-52
- Servedio M.R. & Noor M.A.F. (2003) The role of reinforcement in speciation: Theory and data. *Annu Rev Ecol Syst*, 34, 339-364
- Servedio M.R., Van Doorn G.S., Kopp M., Frame A.M. & Nosil P. (2011) Magic traits in speciation: 'magic' but not rare? *Trends Ecol Evol*, 26, 389-97
- Sites J.W., Barton N.H. & Reed K.M. (1995) The genetic-structure of a hybrid zone between 2 chromosome races of the *sceloporus grammicus* complex (sauria, phrynosomatidae) in central mexico. *Evolution*, 49, 9-36
- Sivanesan A. (1977) *The taxonomy and pathology of Venturia species*. J. Cramer, Vaduz.
- Slatkin M. (1975) Gene flow and selection in a two-locus system. *Genetics*, 81, 787-802
- Slatkin M. (1987) Gene flow and the geographic structure of natural populations. *Science*, 236, 787-92
- Sloan D.B., Giraud T. & Hood M.E. (2008) Maximized virulence in a sterilizing pathogen: the anther-smut fungus and its co-evolved hosts. *J Evol Biol*, 21, 1544-54
- Slusarenko A. (2001) *Annual Plant Reviews Vol. 4*. Sheffield Academic Press, Sheffield.
- Smereka K.J., MacHardy W.E. & Kausch A.P. (1987) Cellular differentiation in *Venturia inaequalis* ascospores during germination and penetration of apple leaves. *Canadian Journal of Botany*, 65, 2549-2561
- Solignac M., Vautrin D., Baudry E., Mougél F., Loiseau A. & Cornuet J.M. (2004) A microsatellite-based linkage map of the honeybee, *Apis mellifera* L. *Genetics*, 167, 253-62
- Soufflet-Freslon V., Gianfranceschi L., Patocchi A. & Durel C.E. (2008) Inheritance studies of apple scab resistance and identification of *Rvi14*, a new major gene that acts together with other broad-spectrum QTL. *Genome*, 51, 657-67
- Spanu P.D., Abbott J.C., Amselem J., Burgis T.A., Soanes D.M., Stuber K., Ver Loren van Themaat E., Brown J.K., Butcher S.A., Gurr S.J., Lebrun M.H., Ridout C.J., Schulze-Lefert P., Talbot N.J., Ahmadinejad N., Ametz C., Barton G.R., Benjdia M., Bidzinski P., Bindschedler L.V., Both M., Brewer M.T., Cadle-Davidson L., Cadle-Davidson M.M., Collemare J., Cramer R., Frenkel O., Godfrey D., Harriman J., Hoede C., King B.C., Klages S., Kleemann J., Knoll D., Koti P.S., Kreplak J., Lopez-Ruiz F.J., Lu X., Maekawa T., Mahanil S., Micali C., Milgroom M.G., Montana G., Noir S., O'Connell R.J., Oberhaensli S., Parlange F., Pedersen C., Quesneville H., Reinhardt R., Rott M., Sacristan S., Schmidt S.M., Schon M., Skamnioti P., Sommer H., Stephens A., Takahara H., Thordal-Christensen H., Vigouroux M., Wessling R., Wicker T. &

- Panstruga R. (2010) Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science*, 330, 1543-6
- Spies I.B., Anderson E.C., Naish K. & Bentzen P. (2007) Evidence for the existence of a native population of sockeye salmon (*Oncorhynchus nerka*) and subsequent introgression with introduced populations in a Pacific Northwest watershed. *Canadian Journal of Fisheries and Aquatic Sciences*, 64, 1209-1221
- Spirito F. (1992) The exact values of the probability of fixation of underdominant chromosomal rearrangements. *Theor Popul Biol*, 41, 111-20
- Staats M., van Baarlen P. & van Kan J.A. (2005) Molecular phylogeny of the plant pathogenic genus *Botrytis* and the evolution of host specificity. *Mol Biol Evol*, 22, 333-46
- Stergiopoulos I., De Kock M.J., Lindhout P. & De Wit P.J. (2007) Allelic variation in the effector genes of the tomato pathogen *Cladosporium fulvum* reveals different modes of adaptive evolution. *Mol Plant Microbe Interact*, 20, 1271-83
- Storz J.F. & Wheat C.W. (2010) Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution*, 64, 2489-509
- Stukenbrock E.H., Banke S., Javan-Nikkhah M. & McDonald B.A. (2007) Origin and domestication of the fungal wheat pathogen *Mycosphaerella graminicola* via sympatric speciation. *Mol Biol Evol*, 24, 398-411
- Stukenbrock E.H., Christiansen F.B., Hansen T.T., Dutheil J.Y. & Schierup M.H. (2012) Fusion of two divergent fungal individuals led to the recent emergence of a unique widespread pathogen species. *Proc Natl Acad Sci U S A*, 109, 10954-9
- Stukenbrock E.H. & McDonald B.A. (2008) The origins of plant pathogens in agroecosystems. *Annual Review of Phytopathology*, 46, 75-100
- Szymura J.M. (1993) Analysis of hybrid zones within *Bombina*. In: *Hybrid zones and the evolutionary process* (ed. Harrison RG), pp. 261-287. Oxford University Press, New York
- Szymura J.M. & Barton N.H. (1986) Genetic-Analysis of a Hybrid Zone between the Fire-Bellied Toads, *Bombina bombina* and *Bombina variegata*, near Cracow in Southern Poland. *Evolution*, 40, 1141-1159
- Tang S. & Presgraves D.C. (2009) Evolution of the *Drosophila* nuclear pore complex results in multiple hybrid incompatibilities. *Science*, 323, 779-82
- Tavaré S., Balding D.J., Griffiths R.C. & Donnelly P. (1997) Inferring coalescence times from DNA sequence data. *Genetics*, 145, 505-18
- Taylor D.R. & Ingvarsson P.K. (2003) Common features of segregation distortion in plants and animals. *Genetica*, 117, 27-35
- Taylor J., Pharis R., Loveys B., Notodimedjo S. & Edwards G. (1984) Changes in endogenous hormones in apple during bud burst induced by defoliation. *Plant Growth Regulation*, 2, 117-134
- Tenzer I., Degli Ivanissevich S., Morgante M. & Gessler C. (1999) Identification of Microsatellite Markers and Their Application to Population Genetics of *Venturia inaequalis*. *Phytopathology*, 89, 748-53
- Tenzer I. & Gessler C. (1997) Subdivision and genetic structure of four populations of *Venturia inaequalis* in Switzerland. *European Journal of Plant Pathology*, 103, 565-571
- Tenzer I. & Gessler C. (1999) Genetic diversity of *Venturia inaequalis* across Europe. *European Journal of Plant Pathology*, 105
- Tenzer I.G., C. (1997) Comparison of population structure of *Venturia inaequalis* in four orchards and one single tree. In: *Integrated Control of pome fruit diseases* (eds. Berrie AX, Berrie XM, Harris DC, Roberts AL, Evans KM & Barbara DJ), pp. 221-228. IOBC/wprs Bulletin

- Thrall P.H. & Burdon J.J. (2003) Evolution of virulence in a plant host-pathogen metapopulation. *Science*, 299, 1735-7
- Torriani S.F., Brunner P.C. & McDonald B.A. (2011) Evolutionary history of the mitochondrial genome in *Mycosphaerella* populations infecting bread wheat, durum wheat and wild grasses. *Mol Phylogenet Evol*, 58, 192-7
- Tsakas S. & Krimbas C.B. (1976) Testing the heterogeneity of F values: a suggestion and a correction. *Genetics*, 84, 399-401
- Turelli M., Barton N.H. & Coyne J.A. (2001) Theory and speciation. *Trends Ecol Evol*, 16, 330-343
- Turner J.R.G. (1967) Why does the genotype not congeal? *Evolution*, 21, 645-656
- Turner T.L., Bourne E.C., Von Wettberg E.J., Hu T.T. & Nuzhdin S.V. (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat Genet*, 42, 260-3
- Turner T.L., Hahn M.W. & Nuzhdin S.V. (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol*, 3, e285
- Van der Biezen E.A. & Jones J.D. (1998) Plant disease-resistance proteins and the gene-for-gene concept. *Trends Biochem Sci*, 23, 454-6
- Van der Hoorn R.A., De Wit P.J. & Joosten M.H. (2002) Balancing selection favors guarding resistance proteins. *Trends Plant Sci*, 7, 67-71
- Van der Plank J.E. (1963) *Plant diseases: epidemics and control*. Academic Press, New York.
- Van Valen L. (1973) A new evolutionary law. *Evolutionary Theory*, 1, 1-30
- Velasco R., Zharkikh A., Affourtit J., Dhingra A., Cestaro A., Kalyanaraman A., Fontana P., Bhatnagar S.K., Troggo M., Pruss D., Salvi S., Pindo M., Baldi P., Castelletti S., Cavaiuolo M., Coppola G., Costa F., Cova V., Dal Ri A., Goremykin V., Komjanc M., Longhi S., Magnago P., Malacarne G., Malnoy M., Micheletti D., Moretto M., Perazzolli M., Si-Ammour A., Vezzulli S., Zini E., Eldredge G., Fitzgerald L.M., Gutin N., Lanchbury J., Macalma T., Mitchell J.T., Reid J., Wardell B., Kodira C., Chen Z., Desany B., Niazi F., Palmer M., Koepke T., Jiwan D., Schaeffer S., Krishnan V., Wu C., Chu V.T., King S.T., Vick J., Tao Q., Mraz A., Stormo A., Stormo K., Bogden R., Ederle D., Stella A., Vecchietti A., Kater M.M., Masiero S., Lasserre P., Lespinasse Y., Allan A.C., Bus V., Chagne D., Crowhurst R.N., Gleave A.P., Lavezzo E., Fawcett J.A., Proost S., Rouze P., Sterck L., Toppo S., Lazzari B., Hellens R.P., Durel C.E., Gutin A., Bumgarner R.E., Gardiner S.E., Skolnick M., Egholm M., Van de Peer Y., Salamini F. & Viola R. (2010) The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat Genet*, 42, 833-9
- Vera Cruz C.M., Bai J., Ona I., Leung H., Nelson R.J., Mew T.W. & Leach J.E. (2000) Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proc Natl Acad Sci U S A*, 97, 13500-5
- Via S. (2001) Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology and Evolution*, 16, 381-390
- Vinatzer B.A., Patocchi A., Gianfranceschi L., Tartarini S., Zhang H.B., Gessler C. & Sansavini S. (2001) Apple contains receptor-like genes homologous to the *Cladosporium fulvum* resistance gene family of tomato with a cluster of genes cosegregating with Vf apple scab resistance. *Molecular Plant-Microbe Interactions*, 14, 508-515
- Vinatzer B.A., Zhang H.B. & Sansavini S. (1998) Construction and characterization of a bacterial artificial chromosome library of apple. *Theoretical and Applied Genetics*, 97, 1183-1190
- Vines T.H., Kohler S.C., Thiel M., Ghira I., Sands T.R., MacCallum C.J., Barton N.H. &

- Nurnberger B. (2003) The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B. variegata*. *Evolution*, 57, 1876-88
- Vines T.H. & Schluter D. (2006) Strong assortative mating between allopatric sticklebacks as a by-product of adaptation to different environments. *Proc Biol Sci*, 273, 911-6
- Vitalis R., Dawson K. & Boursot P. (2001) Interpretation of variation across marker loci as evidence of selection. *Genetics*, 158, 1811-1823
- Wagner I. & Weeden N.F. (2000) Isozyme in *Malus sylvestris*, *Malus x domestica* and in related *Malus* species. *Acta Horticulturae*, 538, 51-56
- Walton J.D. (2000) Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genet Biol*, 30, 167-71
- Watterson G.A. (1983) On the time for gene silencing at duplicate loci. *Genetics*, 105, 745-766
- Weiblen G.D. & Bush G.L. (2002) Speciation in fig pollinators and parasites. *Molecular Ecology*, 11, 1573-1578
- Werth C.R. & Windham M.D. (1991) A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. 137, 515-526
- Wheat C.W., Watt W.B., Pollock D.D. & Schulte P.M. (2006) From DNA to fitness differences: sequences and structures of adaptive variants of *Colias* phosphoglucose isomerase (PGI). *Mol Biol Evol*, 23, 499-512
- Whitkus R. (1998) Genetics of adaptive radiation in Hawaiian and Cook Islands species of *Tetramolopium* (Asteraceae). II. Genetic linkage map and its implications for interspecific breeding barriers. *Genetics*, 150, 1209-16
- Whitlock R., Hipperson H., Mannarelli M., Butlin R.K. & Burke T. (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Mol Ecol Resour*, 8, 725-35
- Wickham H. (2009) *ggplot2: elegant graphics for data analysis*. Springer, New York.
- Williams E.B. & Shay J.R. (1957) The relationship of genes for pathogenicity and certain others characters in *Venturia inaequalis*. *Genetics*, 42, 704-711
- Winton L.M., Stone J.K., Hansen E.M. & Shoemaker R.A. (2007) The systematic position of *Phaeocryptopus gaeumannii*. *Mycologia*, 99, 240-52
- Wright S. (1931) Evolution in Mendelian Populations. *Genetics*, 16, 97-159
- Wu C.-I. & Davis A.W. (1993) Evolution of postmating reproductive isolation: the composite nature of Haldane's rule and its genetic bases. *The American Naturalist* 142, 187-212
- Xu M. & Korban S.S. (2002) AFLP-derived SCARs facilitate construction of a 1.1 Mb sequence-ready map of a region that spans the *Vf* locus in the apple genome. *Plant Mol Biol*, 50, 803-18
- Yamagata Y., Yamamoto E., Aya K., Win K.T., Doi K., Sobrizal, Ito T., Kanamori H., Wu J., Matsumoto T., Matsuoka M., Ashikari M. & Yoshimura A. (2010) Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. *Proc Natl Acad Sci U S A*, 107, 1494-9
- Zaffarano P.L., McDonald B.A. & Linde C.C. (2008) Rapid speciation following recent host shifts in the plant pathogenic fungus *Rhynchosporium*. *Evolution*, 62, 1418-36
- Zaffarano P.L., McDonald B.A., Zala M. & Linde C.C. (2006) Global Hierarchical Gene Diversity Analysis Suggests the Fertile Crescent Is Not the Center of Origin of the Barley Scald Pathogen *Rhynchosporium secalis*. *Phytopathology*, 96, 941-50
- Zamir D. & Tadmor Y. (1986) Unequal segregation of nuclear genes in plants. *Bot. Gaz.*, 147, 355-358
- Zhang G., Salesses G. & Lespinasse Y. (1988) Etude cytologique de quelques clones diploïdes et polyplloïdes du pommier. Considérations quant à l'origine du pommier. *Cytologia*,

53, 739-758

- Zhang Y., Crous P.W., Schoch C.L., Bahkali A.H., Guo L.D. & Hyde K.D. (2011) A molecular, morphological and ecological re-appraisal of Venturiales-a new order of Dothideomycetes. *Fungal Divers*, 51, 249-277
- Zhang Y., Wang H.K., Fournier J., Crous P.W., Jeewon R., Pointing S.B. & Hyde K.D. (2009) Towards a phylogenetic clarification of Lophiostoma/Massarina and morphologically similar genera the Pleosporales. *Fungal Divers*, 38, 225–251
- Zohary D. & Hopf M. (2000) *Domestication of plants in the Old World*. Oxford University press, Oxford.

ANNEXE

The genetic structure of a *Venturia inaequalis* population in a heterogeneous host population composed of different *Malus* species

Thibault Leroy^{1,2,3}, Christophe Lemaire^{1,2,3*}, Frank Dunemann⁴ and Bruno Le Cam^{1,2,3*}

(article en révision pour BMC Evolutionary Biology)

Abstract

Background:

Adaptation, which induces differentiation between populations in relation to environmental conditions, can initiate divergence. The balance between gene flow and selection determines the maintenance of such a structure in sympatry. Studying these two antagonistic forces in plant pathogens is made possible because of the high ability of pathogens to disperse and of the strong selective pressures exerted by their hosts. In this article, we analysed the genetic structure of the population of the apple scab fungus, *Venturia inaequalis*, in a heterogeneous environment composed of various *Malus* species. Inferences were drawn from microsatellite and AFLP data obtained from 114 strains sampled in a single orchard on nine different *Malus* species to determine the forces that shape the genetic structure of the pathogen.

Results:

Using clustering methods, we first identified two specialist subpopulations: (i) a virulent subpopulation sampled on *Malus* trees carrying the *Rvi6* resistance gene; and (ii) an avirulent subpopulation infecting only *Malus* trees that did not carry this resistance gene. A genome scan of loci on these two subpopulations did not detect any locus under selection. Additionally, we did not detect any other particular substructure linked to different hosts. However, an isolation-by-distance (IBD) pattern at the orchard scale revealed free gene flow within each subpopulation.

Conclusions:

Our work shows a rare example of a very strong effect of a resistance gene on pathogen populations. Despite the high diversity of *Malus* hosts, the presence of *Rvi6* seems sufficient to explain the observed genetic structure. Moreover, detection of an IBD pattern at the orchard scale revealed a very low average dispersal distance that is particularly significant for epidemiologists and landscape managers for the design of scab control strategies

Keywords

Gene flow, isolation-by-distance (IBD), apple scab, adaptation, spatial genetic structure

Background

The diversification of pathogens on different hosts is commonly thought to arise via new adaptations in response to the disruptive selection exerted by host defences or non-host resistances. The “Red Queen’s race” hypothesis [1] describes the coevolution of hosts and pathogens in natural ecosystems. In agro-ecosystems, higher homogeneity and density of hosts, coupled with a low host species diversity, are expected to induce quick adaptive changes in pathogens [2]. Indeed, in pathogens, higher fecundity associated with large effective sizes and shorter generation times enhances the rise of new mutants able to settle on new hosts [3-6]. Adaptive mutations toward virulence can lead to founder effects that will remain detectable over the years only if gene flow with the resident population is prevented. In sympatry, high selection pressure experienced by the pathogen or strong assortative mating are required to maintain population structure over time by limiting the homogenising effects of gene flow between strains adapted to different hosts [7, 8]. In pathogens that mate within their hosts, adaptation can induce a strong and stable reproductive isolation between populations that facilitates the maintenance of genetic differentiation [7, 8]. Since an agricultural landscape can be seen as a mosaic of crops with multiple cultivars carrying different resistance traits, both adaptation and gene flow processes are expected to shape the genetic structure of pathogen populations. We can therefore ask if, as a result of these processes, a landscape constituted by different hosts will produce a mosaic of different pathogen populations, each specialised on a host species.

Short distances between plants in fields permit the rapid spread of a pathogen and large population sizes [2]. Since most pathogens produce large amounts of descendants through highly efficient sexual and/or asexual offspring productions (e.g., [9, 10]), detectable gene flow between populations on different neighbour hosts is expected [11], provided that appropriate tools are used. The two-dimensional stepping stone model [12], a system in which individuals stochastically diffuse within a lattice, should fit the spatial structure encountered in fields. In such a model, the probability of gene flow is dependent on the connectivity between demes. The relatedness between strains therefore decreases with geographical distance. No occurrence of an Isolation-By-Distance (IBD) pattern at a field scale (several hundreds of metres) has been reported to date in fungal pathogens since evidence of IBD implies short distance dispersal, which is not the general outcome assumed for pathogens able to produce small sexual propagules [11, 13]. An IBD pattern over a heterogeneous landscape of host species would indicate that strains are generalists. Under this hypothesis, two strains isolated from neighbour demes that infect different host species are expected to be more closely related than two strains isolated from two distant demes, even if they infect the same host. We could then question whether it remains possible to maintain pathogen differentiations related to hosts in a heterogeneous environment.

In summary, it is assumed that two types of patterns of population genetic structure exist for pathogen populations: IBD or/and hierarchical structure. First, is it possible to observe a structure related to the hosts? And second, can an IBD pattern be detected, suggesting that gene flow is only driven by free migration?

Venturia inaequalis is an ascomycete fungus responsible for scab, a major apple disease in most areas of the world. The interaction between *Malus x domestica* and *V. inaequalis* fits the gene-for-gene model. In a gene-for-gene interaction, the product of a resistance gene in the host recognises an “effector” gene product in the pathogen and activates a defence reaction that completely prevents infection [14]. Numerous resistance genes providing resistance against *V. inaequalis* have been identified within *Malus* species [15]. The life cycle of *V. inaequalis* comprises both sexual and asexual phases of reproduction. Sexual mating occurs during the winter inside dead leaves in the litter layer, between strains of opposite mating types that have infected the same leaf. Zygotes undergo immediate meiosis and yield haploid ascospores that are released to initiate new infections in spring that subsequently disseminate via asexually produced conidia. The pathosystem *Malus* spp.-*V. inaequalis* is suitable to study both gene flow and host adaptation. Epidemiological studies suggest that both types of propagules are dispersed over short distances (< 50 metres) [16, 17], which makes an IBD pattern possible at the orchard scale. Additionally, *V. inaequalis* is described as a good model for host adaptation in plant pathogens [7] because mating only occurs between strains that are able to infect identical hosts, thus facilitating the maintenance of adaptations to a host when host ranges do not overlap [18, 19]. *V. inaequalis* is a pathogen that is well-known for overcoming resistance genes introgressed into cultivars from *M. x domestica* germplasm and from wild genetic resources of *Malus* [15, 20]. Moreover, population genetic studies have shown that apple resistance genes might induce specialisation in *V. inaequalis* populations, thus favouring host-related adaptations [18, 20]. For example, in agro-ecosystems, the presence of the resistance gene *Rvi6* in apple divides *V. inaequalis* into two populations: one emerging - *virRvi6* – that infects *Rvi6* cultivars, and another - *avrRvi6* - that infects cultivars without this resistance gene [20]. To date, the population structure between *avrRvi6* and *virRvi6* is maintained in agro-ecosystems, even when *Rvi6* and non-*Rvi6* cultivars are planted in the same orchards [18].

The aim of this study was to evaluate which antagonistic force, host-related adaptation (e.g., virulence toward the *Rvi6* resistance gene) or gene flow, shapes the *V. inaequalis* genetic structure within a genetically heterogeneous orchard consisting of different species of *Malus*. Our prediction was the following: in the absence of adaptation, an IBD pattern was expected at the orchard scale that would indicate free gene flow between strains, regardless of the host. Using polymorphism of microsatellite and AFLP markers, we infer the genetic structure of *V. inaequalis* populations, allowing us to address the following question: do host adaptations or gene flow shape the structure of *V. inaequalis* populations?

Methods

Fungal sampling and genotyping

Samples were isolated in a *Malus* orchard located in Dresden-Pillnitz (Saxony, Germany). This orchard has been free from fungicide treatment since it was planted in 1997. Within the orchard, *Malus* trees expressed a wide range of disease severity, from *highly sensitive to fully resistant* to *V. inaequalis*. A total of 114 strains derived from monoconidial isolates of *V. inaequalis* were sampled on 57 trees classified into eight groups: five from different species of *Malus* (*M. sieversii*, *M. sylvestris*, *M. baccata*, *M. ioensis* and *M.*

coronaria) and three from hybrids of *Malus* (*M. x floribunda*, *M. x purpurea* and *M. x zumi* (supplementary data, Table 1).

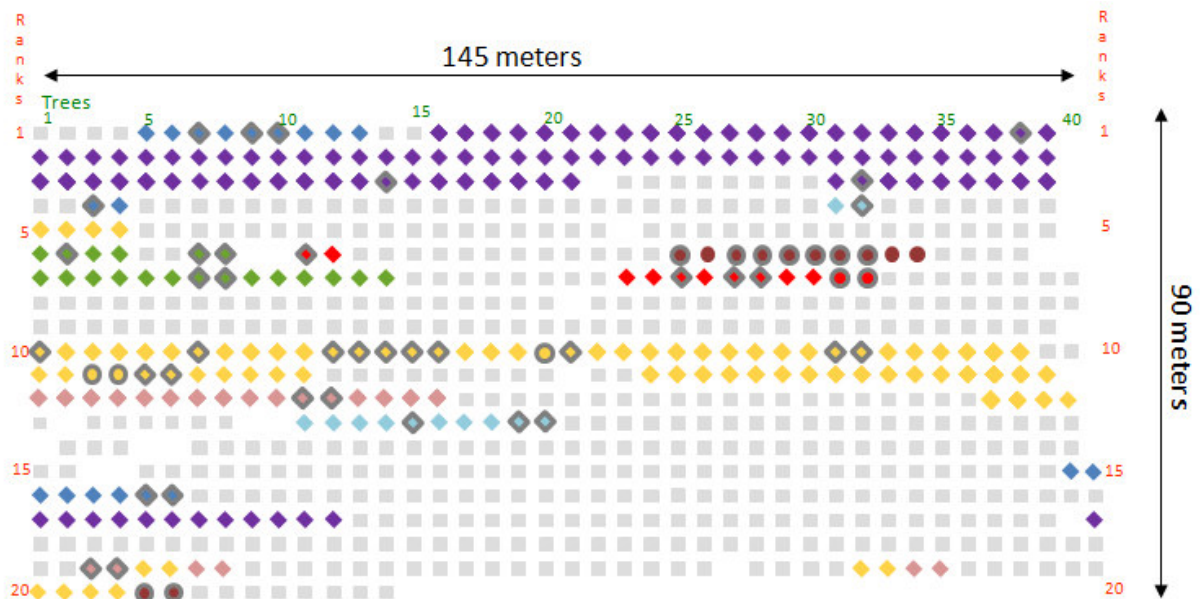


Figure 1: The orchard map of *Malus* hosts where *V. inaequalis* strains were sampled. Non-Rvi6 host genotypes are represented by a diamond: *M. sieversii* (blue), *M. sylvestris* (purple), *M. coronaria* (green), *M. x purpurea* (light blue), *M. baccata* (yellow), *M. ioensis* (red) and *M. x zumi* (pink). Rvi6 host genotypes are represented by a circle: *M. x floribunda* (brown), *M. baccata* (yellow) and *M. ioensis* (red). Other *Malus* species or hybrids not sampled (infected or not) are represented by grey squares.

Fungal DNA was extracted from monoconidial individuals using a phenol/chloroform protocol [21]. The individuals were genotyped at 11 microsatellite loci: *Itc1a*, *Itc1b*, *Itc1g*, *Iaac3b* [22], *Vitca7/P*, *Vitg11/70*, *Vicacg8/42*, *Viga7/116*, *Vica 9/152*, *Vica9/X* [23] and *M42* [19]. Individuals were also genotyped with an AFLP combination using the following selective primers: *Pst*I primer + A (5'-GACTGAGTACGTGCAGA-3') and *Msp*I primer + AA (5'-CGATGAGTCCTGAGCGGAA3'-). Both microsatellite and AFLP amplifications were performed according to a previously described protocol [20, 24]. PCR products were manually scored against a fluorescently labelled size standard (400HD-rox for microsatellite and GS500-TAMRA for AFLP) in an ABI 3130 automated sequencer using Genemapper 4.0 software (Applied Biosystems, Foster City, CA, USA). Only fragments between 70 and 395 bp were scored. The presence or absence of a polymorphic marker was binary coded. A clone-corrected dataset composed of 11 SSR and 79 AFLP polymorphisms was used for all of the following tests. The dataset is available through the DRYAD data archive system (see the availability of supporting data).

Plant material

The sampling strategy was established to select *V. inaequalis* strains from the greatest diversity of infected hosts available within the core orchard. Since the scab resistance gene

Rvi6 (previously named as *Vf*), introgressed from *M. x floribunda*, is known to exert a strong selective pressure on *V. inaequalis* populations in commercial orchards [18], we therefore investigated the presence/absence of the *Rvi6* locus within the genome of the sampled *Malus* accessions. All accessions were genotyped using three molecular markers tightly linked to the *Rvi6* locus on the apple linkage group 1 (LG 1) [25-28], except for the *M. ioensis* accession MAL330, whose *Rvi6* allele sequences were already known (Dunemann, pers. comm.). An apple accession was classified as carrying the *Rvi6* locus when all three diagnostic alleles were detected (CH-Vf1: allele 159, AL07-SCAR: 480 bp fragment, Vfa2: 550 bp fragment) (supplementary data, Table 1).

Data analysis

Standard diversity indices, AMOVA and PCoA

Because samples were collected during the asexual stages of the *V. inaequalis* life cycle, isolates with exactly the same alleles at all loci were removed (clone-corrected dataset). A genetic distance matrix was created from the dataset under Genalex 6.1 [29]. An analysis of molecular variance (AMOVA) and a principal coordinates analysis (PCoA) were performed using this genetic distance matrix. The average gene diversity (H_d ; [30]) and the average number of alleles (A) were estimated from clone-corrected datasets using Arlequin software, v. 3.11 [31].

Selection of neutral AFLP markers

In order to eliminate markers under selection for the subsequent genetic analyses, we performed an outlier detection using BayeScan software [32]. Using an estimation of locus-population specific F_{ST} , this method determines a posterior probability of a locus to be under selection. We used the “decisive” threshold included on the user-friendly interface, which is equivalent to a 95% confidence interval.

Individual assignments

Individual assignments were performed using the Bayesian clustering method implemented in STRUCTURE v. 2.2.3 [33-35]. Admixture mode was used, and the Monte Carlo Markov Chain scheme was run for 500,000 iterations after an initial burn-in period of 50,000. We ran STRUCTURE for K clusters ranging from 1 to 8, and performed five repetitions to check for the convergence of likelihood values for each value of K . Evanno’s method, ΔK , was used to best estimate K [36, 37]. This method was computed using the STRUCTURE HARVESTER programme, v. 0.56.3 [37].

Additional cluster analyses were performed using TESS, v. 2.3.1 [38-40]. TESS implements an individual-based spatially-explicit Bayesian algorithm, and uses a hidden Markov random field model to compute the proportion of individual genomes originating in K populations. The hidden Markov random field accounts for spatial connectivity and incorporates a decay of membership coefficient correlation with distance, which is a property similar to IBD. The algorithm was run with a burn-in period of 50,000 cycles and the estimation was performed using 100,000 additional cycles. We increased the maximum

number of clusters from $K_{\max}=2$ to $K_{\max}=8$ (100 replicates for each value). We used 10% of simulations that minimised the Deviance Information Criterion (DIC) to obtain simulations that best fit the model for each K , and discarded 90 simulations. The ten retained simulations were subsequently analysed under CLUMPP, v. 1.1.2 [41], to average the estimated admixture coefficients. The averaged admixture coefficients were used as inputs for the spatial interpolation. Interpolations were performed using the R script [42] available with TESS software (see the TESS users' manual).

Testing for IBD patterns

The detection of IBD was performed using two methods. The first method is implemented in SPAGeDI software, v. 1.2 [43], and the second one uses variograms of gene diversity [44]. In an IBD pattern, kinship is expected to linearly decrease with geographic distance when the demes are connected according to a one-dimensional stepping stone (1D-SS) model, and to linearly decrease with the natural logarithm of distance according to a two-dimensional stepping stone (2D-SS) model [45, 46]. Under an IBD model, an autocorrelogram shows a decreasing curve of the mean kinship between strains when the distance increases, whereas in a variogram of gene diversity, the curve represents the strain differences and is therefore expected to increase with distance. The gradual decrease of the relationship with distance between strains can be explained by limited migration abilities.

Using SPAGeDI, pairwise Loiselle estimators (F_{ij}) of kinship [47] between individuals were estimated from all microsatellite and neutral AFLP loci. $F_{(d)}$ statistics were calculated from the means of F_{ij} between pairs of individuals inside distance classes. Linear regression and pairwise Euclidian spatial distances between individuals were tested using a permutation procedure. Permutations were performed 20,000 times on localisations and 20,000 times on individuals. The autocorrelograms produced represent $F_{(d)}$ plotted against the natural logarithm of distance (i.e., in a field, the 2D-SS model, a model where individuals move at random within a lattice [46], appears closer to reality).

Confirmations of the IBD tests were performed using variograms of gene diversity, as described by Wagner *et al.* [44].

$$\hat{Y}_l(r) = \sum_k \sum_{a < b} \frac{\chi_{ab}^{(r)}}{2n_r} (z_{lka} - z_{lkb})^2 \chi_{ab}^{(r)} \widehat{H}_r \hat{Y}_l(r)$$

Ninety-five percent confidence intervals were estimated by manually analysing 100 different permuted datasets and observing the variation range of \hat{Y}_l gene diversity for each r class. For each of the 100 datasets, we performed permutations of individuals: 500 within the *virRvi6* subpopulation and 1000 within the *avrRvi6* subpopulation.

Classic and partial Mantel tests

The Meirmans procedure [48] was performed to test whether population structure detected by STRUCTURE was due to an IBD or to hierarchical clustering. Such an approach is based on computation of several independent Mantel tests using two different matrices and possibly a third matrix as a covariate for partial Mantel tests. The three different matrices are: (1) a genetic distance matrix; (2) a Euclidian geographic distance matrix; and (3) a matrix of cluster membership. The third matrix describes whether comparisons were made

between individuals mainly assigned to the same cluster by STRUCTURE (1) or to different clusters (2). Comparing r values and the significance of the Mantel test (e.g., clusters and genetic distance vs. geographic and genetic distance) reveals whether the structure is mainly spatial or hierarchical. For details on the procedure, see [48]. These tests were performed using the VEGAN package [49] in R [42].

Results

The presence of the *Rvi6* resistance gene in *Malus* accessions divides the *V. inaequalis* population into two subpopulations

Among the 114 strains analysed, we observed 106 unique haplotypes based on 11 microsatellite loci and 79 AFLP markers. The number of alleles at each microsatellite locus ranged from two, at *Iaac3b*, to 17, at *Itc1g*, with an average value of 7.8 (± 4.6 SD).

PCoA analysis separated *V. inaequalis* strains into three distinct groups with the first and second axes representing 27.9% and 18.7% of total inertia, respectively. The first axis obviously separated strains into two main groups: group 1 containing strains expressing a low score (< -0.4), and group 2 (> -0.4). The second principal coordinate also divided group 2 into two subgroups (“subgroup 2a” for negative values and “subgroup 2b” for positive values). Because the major resistance gene *Rvi6*, introgressed from the clone *Malus x floribunda* 821, is known to exert a high selective pressure on *V. inaequalis* populations [18], we evaluated the presence of this gene in all hosts sampled in this study (supplementary data, Table 1). We showed that all accessions sampled on non-*Rvi6* hosts, *M. sieversii*, *M. sylvestris*, *M. baccata*, *M. coronaria*, *M. x zumi* (except for one strain) and *M. x purpurea*, were grouped into the single and close subgroup 2b (diamonds; Fig. 2a). All strains sampled on the *Rvi6* trees, *M. x floribunda*, *M. ioensis* and *M. baccata*, are grouped into two different clusters, group 1 and subgroup 2a (circles; Fig. 2a).

Traces of host specialisation within the *V. inaequalis* population were also assessed with the assignment method implemented in STRUCTURE. The clustering algorithm supported two clusters. Admixture coefficients for the best partitioning (K=2) of all haplotypes were reported (Fig. 2a). Strains assigned to one cluster (red; Fig 2b) were sampled on hosts carrying the *Rvi6* resistance gene, except for one strain isolated on *M. x zumi* (MAL0964), which would very likely be a first-generation migrant (supplementary data, Table 3). Moreover, STRUCTURE detected 15 individuals with a Q admixture proportion to the first cluster of between 0.2 and 0.8 (14% of the strains), suggesting a substantial level of gene flow between the two clusters.

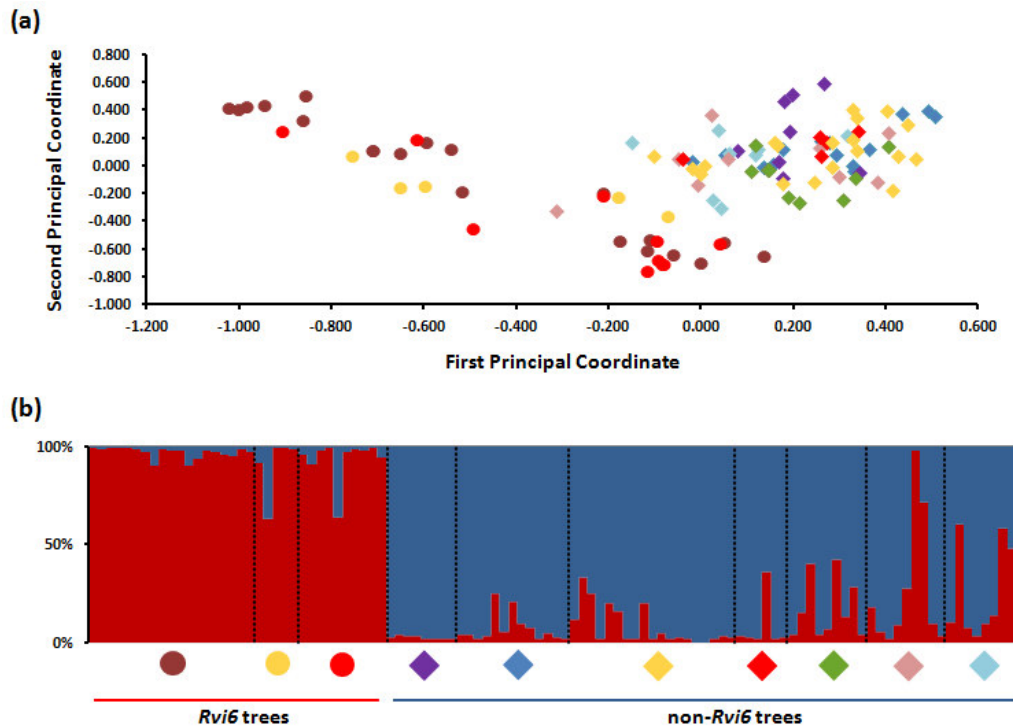


Figure 2: The principal coordinates and STRUCTURE assignment analyses for the 106 *V. inaequalis* haplotypes. (a) The first and second axes of the PCoA represent 27.9% and 18.7% of the total inertia. The strains collected from non-*Rvi6* hosts are represented by a diamond: *M. sieversii* (blue), *M. sylvestris* (purple), *M. coronaria* (green), *M. x purpurea* (light blue), *M. baccata* (yellow), *M. ioensis* (red) and *M. x zumi* (pink). The strains collected from *Rvi6* hosts are represented by a circle: *M. x floribunda* (brown), *M. baccata* (yellow) and *M. ioensis* (red). (b) Haplotypes are represented by a bar partitioned into $K=2$ segments that represent the haplotype's estimated membership fractions calculated by STRUCTURE for each of the two clusters.

Additionally, AMOVA reveals that the two groups were significantly different ($\Phi_{PT}=0.137$; $p=0.01$; Φ_{PT} is an analogue of F_{ST} that represents genetic diversity within and among populations [29]). Taken together, the analyses revealed two distinct subpopulations with some putative hybrids, which is indicative of gene flow between the two subpopulations.

For subsequent analyses, strains that were collected from hosts carrying the *Rvi6* gene were labelled “vir*Rvi6*”, whereas strains collected from non-*Rvi6* hosts were labelled “avr*Rvi6*” (supplementary data, Table 3).

No selection signature was detected within each differentiated subpopulation.

To check whether the loci are neutral or targeted by natural selection, the distribution of F_{ST} against the probability to be under selective pressure was simulated with BayeScan software [32]. Considering the two previously detected subpopulations, the posterior probability revealed no marker plotted outside the 95% confidence interval constructed (data not shown). No SSR and AFLP loci were subject to selection, but instead exhibited a moderate to high level of neutral differentiation (mean $F_{ST} \pm SE = 0.179 \pm 0.057$). All markers were then considered as neutral and were used for all subsequent analyses.

No evidence of additional genetic structure related to the host

Considering these two subpopulations, molecular diversity indices revealed a higher

diversity in the *avrRvi6* than in the *virRvi6* subpopulation (supplementary data, Table 4). Each dataset was then reanalysed under a clustering programme (STRUCTURE) to assess host-specificity within each subpopulation. Concerning the *avrRvi6* subpopulation, application of Evanno's method indicated two clusters. However, this optimal number of clusters was not strongly supported ($\Delta k=34.9$), which is consistent with a low genetic structure. In addition, individual assignments to the two clusters indicated that this genetic structure was not correlated to the host species (supplementary data, Fig. 1a). Subsequently, the *avrRvi6* subpopulation was analysed by another clustering method (TESS), taking the spatial distribution of samples within the orchard into account. Simulations with $K_{max}=2$ for the *avrRvi6* subpopulation indicated a spatial structure (Fig. 3a). The posterior probability of an individual belonging to cluster 1 gradually increased with distance (Fig. 3a). Sampling of strains on *M. baccata* trees along a line transect (the yellow diamonds in ranks 10 and 11; Fig. 1) allowed us to validate the observed pattern of spatial structure (Fig. 3a). Strain membership to cluster 1 gradually decreased from the left side of the rank to the right one along this transect, thus strengthening the idea that the main structure was spatial.

Concerning the *virRvi6* subpopulation, applications of a clustering method (STRUCTURE) and Evanno's procedure revealed three clusters with a clear-cut Δk peak observed for $K=3$. However, the three clusters in the *virRvi6* subpopulation were not correlated to the host species (supplementary data, Fig 1b). For example, strains sampled on *M. floribunda* were mainly assigned to each of the three clusters. Furthermore, the posterior probability of *virRvi6* samples belonging to cluster 1 inferred by TESS gradually decreased with distance. Genetic variation of strains sampled on *Rvi6* trees along a line transect also reflected a gradual decrease of membership to cluster 1 with distance (Fig. 3b). The assignment of individuals to a higher value of K clusters (TESS assignment to three or four different clusters; supplementary data, Fig. 2) did not reveal any additional substructure linked to a particular host species.

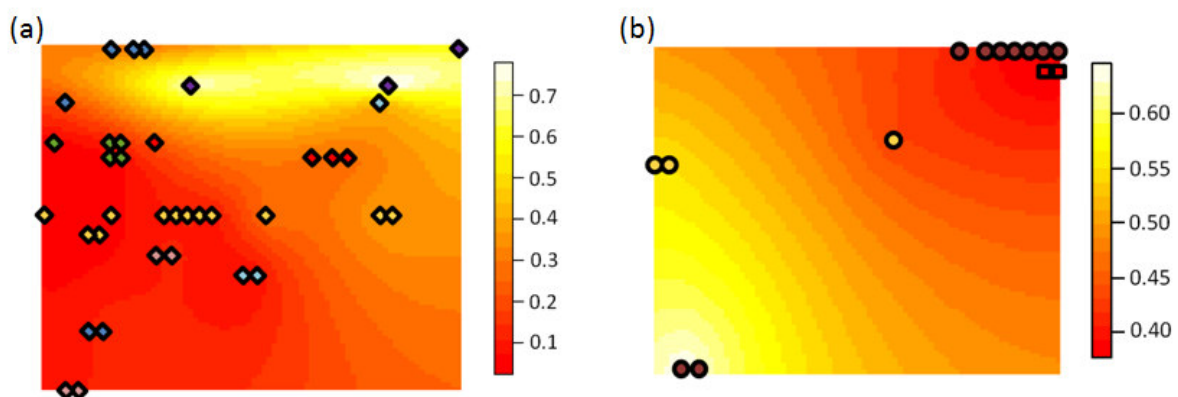


Figure 3: The spatial interpolation map showing differences in the admixture population derived from the TESS assignment to two different clusters. Given the two assumed clusters, the results for the *avrRvi6* dataset (a) and the *virRvi6* dataset (b) are only shown for the first cluster assignment. The strains collected from non-*Rvi6* hosts are represented by a diamond: *M. sieversii* (blue), *M. sylvestris* (purple), *M. coronaria* (green), *M. x purpurea* (light blue), *M. baccata* (yellow), *M. ioensis* (red) and *M. x zumi* (pink). The strains collected from *Rvi6* hosts are represented by a circle: *M. x floribunda* (brown), *M. baccata* (yellow) and *M. ioensis* (red).

Uses of Mantel tests using the Meirmans procedure [48] reinforced the interpretation that population genetic structure within each subpopulation is spatial rather than hierarchical. Indeed, Mantel statistics (r) are high and significant when spatial autocorrelation is explicitly tested, i.e., when testing the association between a matrix of genetic distance and a matrix of geographic distance (*avrRvi6*: $r=0.14$, $p<0.01$; *virRvi6*: $r=0.15$, $p<0.001$; first three rows in supplementary data, Table 5). On the contrary, testing the association between the matrix of genetic distances and a model of the matrix of cluster membership with the matrix of geographical distance as a covariate leads to lower r values and non-significant tests (*avrRvi6*: $r=0.058$, $p=0.051$; *virRvi6*: $r=0.032$, $p=0.342$; last row, in supplementary data, Table 5). This considerably strengthens our finding that the genetic structure within each subpopulation was mainly spatial.

Detection of an IBD pattern within each subpopulation.

The presence of an IBD pattern was then tested within each subpopulation using two methods based on autocorrelograms [45] and variograms of gene diversity [44]. Correlograms based on kinship analyses showed a significant IBD pattern across the *avrRvi6* subpopulation ($\bar{F}(d) = -0.0265 \pm 0.0053$, $r^2=0.0222$, $p=0$) (Fig. 4a). A non-significant IBD pattern ($p=0.085$) was detected among the *virRvi6* strains ($\bar{F}(d) = -0.0307 \pm 0.0080$, $r^2=0.0157$) (Fig. 4b) and between *avrRvi6* vs. *virRvi6* individual pair comparisons ($p=0.99$; Fig. 4, c) Moreover, we observed a global increase of gene diversity with distance (supplementary data, Figs. 3a and b). For the two subpopulations, gene diversity significantly increased between the first and second distance classes. The low gene diversity observed within each first class was never observed after random permutations of the dataset, which indicated a strong and fine-scale IBD pattern. These signals were lost after 20 metres in both cases (supplementary data, Figs. 3a and b), while no IBD pattern was detected between *avrRvi6* vs. *virRvi6* individual pair comparisons (supplementary data, Fig. 3c).

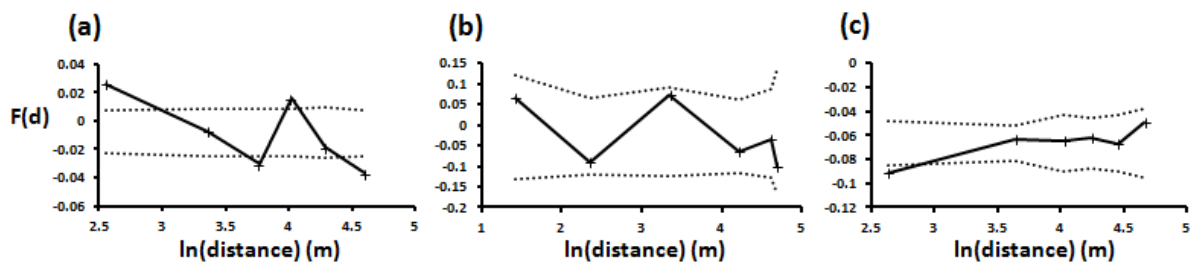


Figure 4: Correlograms of mean kinship coefficients $F(d)$ among pairs of isolates as a function of natural logarithm distance. The analyses were performed for the *avrRvi6* dataset (a), *virRvi6* dataset (b), and for the mean kinship in pairwise comparisons between *avrRvi6* vs. *virRvi6* strains (c). For each correlogram, mean kinship coefficients $F(d)$ were estimated for each of the six different distance classes.

Discussion

Balance between selection and migration is crucial for the maintenance of genetic structure over time. On the one hand, Giraud *et al.* [7, 8] reported that in some cases, host plants induce disruptive selective pressures on pathogen populations that are so high that gene flow is impeded, even if resident and new adapted populations are in sympatry. On the other hand, gene flow is particularly efficient in pathogens because of large population sizes and high fecundity, which is expected to facilitate population homogenisation. In this article, we studied the counteracting effects of the two forces by analysing the genetic structure of apple scab populations in a single orchard constituted of different *Malus* species. We showed that the presence of the *Rvi6* resistance gene in *Malus* accessions divides the *V. inaequalis* population into two subpopulations. Within each subpopulation, an IBD pattern was detected at the orchard scale.

Adaptations to resistance genes

Since wild apple species carry many scab resistance genes [15, 50], we assumed that several host selective pressures exerted on *V. inaequalis* populations would be revealed in this study, as already demonstrated with the *Rvi6* gene [18]. PCoA and a clustering method (STRUCTURE) revealed only two subpopulations: one subpopulation sampled from trees carrying the *Rvi6* gene and a second one infecting non-*Rvi6* trees. This population split was previously described in commercial orchards between infected *M. x domestica* cultivars carrying or not carrying the *Rvi6* gene [18-20]. Surprisingly, this study revealed that the sole presence of the *Rvi6* gene in the different wild *Malus* sp. induced an identical split in *V. inaequalis* populations, despite the very likely presence of other resistance genes in the diverse genetic background of the sampled accessions.

Evidence of admixture between the two subpopulations suggested that strains from the *virRvi6* subpopulation were able to infect non-*Rvi6* cultivars. However, admixture between the two subpopulations remained quite low, suggesting the existence of pre-zygotic and/or post-zygotic barriers to gene flow. Two different hypotheses can therefore be proposed: (i) the virulence cost (i.e., fitness cost on susceptible cultivars associated with a mutation to virulence) was strong enough to generate a strong pre-zygotic genetic barrier sufficient to isolate the *virRvi6* subpopulation from *avrRvi6* by itself [7, 18]; and (ii) *avrRvi6* and *virRvi6* subpopulations were much more divergent than previously suspected and accumulated genetic barriers to gene flow other than the sole mutation at the avirulent locus. Estimating the divergence time of these two subpopulations, associated with an investigation of divergent loci that impede free gene flow, is currently underway in the laboratory to test this last hypothesis.

No significant population genetic structure was detected within each subpopulation. Although different clusters were detected by STRUCTURE within each subpopulation, we did not find evidence of correlations between inferred clusters and host species. Given that *Rvi6* is the main resistance gene used in apple breeding programmes [51, 52] and that no resistance gene other than *Rvi6* has been cloned in apple to date, genotyping other *Malus* resistance genes was not possible. On the one hand, we cannot exclude the existence of other structuring resistance factors shared by accessions belonging to different species. On the other

hand, TESS assignments and map interpolations did reveal that the clusters detected by STRUCTURE in each subpopulation were more likely due to geographic discontinuities in the sampling scheme along the isolation gradient by distance than to other structuring factors related to host resistance genes. Furthermore, applications of partial Mantel tests using the Meirmans procedure [48] within each subpopulation reinforced our finding that structure was mainly spatial. Given the IBD, the sampling artefact is probably more consistent with detected clusters than other structuring factors.

IBD and dispersal abilities

The SGS analyses performed within each *avrRvi6* and *virRvi6* subpopulation highlighted a decrease of mean kinship between strains when the distance increased. First, if host adaptations other than *Rvi6* existed in this orchard, restricted gene flow between structured populations would be detected. Conversely, IBD detection and the gradual assignment to a TESS cluster associated with distance suggested that strains easily shifted from one tree to another, regardless of the host species. Gene flow between individuals of each subpopulation was free, even if the neighbour host belonged to another species. Second, IBD detection across this orchard highlighted restricted *V. inaequalis* dispersal over space. Holb *et al.* [16] showed that the mean daily ascospore count in traps at 21 metres and 45 metres was approximately one-third and one-tenth of the sexual spores trapped at the source, respectively, highlighting that a substantial number of ascospores could travel at least 45 metres from the inoculum source. Lower dispersal abilities were previously reported by Kaplan [53] who observed that 99% of the spores are not able to spread more than 5 or 6 metres. Many factors can explain these dispersal distance differences, including orchard conditions and differences in wind direction and velocity (see [16]). In this study, we confirmed that dispersal of the greater part of *V. inaequalis* spores in orchards was intrinsically very restricted, which is in agreement with several other authors (e.g., [54-56]), who demonstrated that no scab lesion development was detected at 15 to 60 metres beyond an inoculum source. We do not dismiss the sporadic events of Long Distance Dispersal (LDD). However, based on a simulation study, Aylor [57] has shown that *V. inaequalis* ascospores cannot disperse in the air over more than 5 km, even under favourable weather conditions for the pathogen.

Detection of signatures of natural selection

Comparing *virRvi6* and *avrRvi6* subpopulations, none of the markers (AFLP, microsatellites) were detected as outliers using the Foll & Gaggiotti method [32]. Several hypotheses can support this lack of detection. First, if the number and the size of the genomic regions affected by selection were low, the sampling effort was probably insufficient to detect loci affected by divergent natural selection. Such a hypothesis is likely if recombination between the two subpopulations was efficient enough to reduce the size of the genomic region of high F_{ST} , thus limiting high differentiation to the immediate genomic neighbourhood of the *virRvi6* locus. Second, our genome scan might not be powerful enough to detect loci under directional selection. Indeed, the genome scan was performed on a majority of AFLP markers (i.e., bi-allelic), by comparing only two subpopulations that, moreover, exhibited a significant

population structure. A low number of alleles, a reduced number of populations and a strong population genetic structure are known to strongly reduce the power of outlier detection [32].

Conclusions

In conclusion, we detected two genetic patterns of population structure in *V. inaequalis* in an orchard with numerous *Malus* species. The highly structuring effect of the presence or absence of a single resistance gene led to a split of the pathogen population into two subpopulations. Additionally, an IBD pattern was detected within each of these two subpopulations. Our work represents a rare example in pathogens where the dispersal evaluated by a fine scale IBD pattern appropriately fits with previously reported empirical field data [16, 53]. Because many pathogen dispersal capabilities remain unclear, our validation of the IBD approach is particularly significant for epidemiologists and also has practical implications for plant breeders and landscape managers for the design of disease control strategies.

Availability of supporting data

The dataset is available through the DRYAD data archive system (XXXX).

Competing interests

The authors declare no competing interests.

Authors' contributions

The work presented here was carried out jointly with all of the authors. All of the authors defined the research theme. TL carried out the *V. inaequalis* laboratory experiments, analysed the data, interpreted the results and drafted the paper. FD performed the genotyping of *Malus* trees at the *Rvi6* locus. All of the authors participated in the critical revision of the manuscript and gave final approval of the article.

Acknowledgements

The authors wish to thank Sabine Bartsch and Jürgen Egerer for technical assistance on *Rvi6* marker analysis, and Cyril Dutech and Fabien Halkett for their helpful suggestions. We are grateful to two anonymous reviewers for their detailed comments on earlier versions of this paper. We acknowledge Laurence Hibrand-Saint Oyant, responsible for the ANAN platform of the SFR QUASAV, where genotyping was performed. Thibault Leroy was supported by a PhD fellowship from the General Council of the Maine & Loire Department, and the COSAVE programme, funded by the Regional Council of the Pays de La Loire.

References

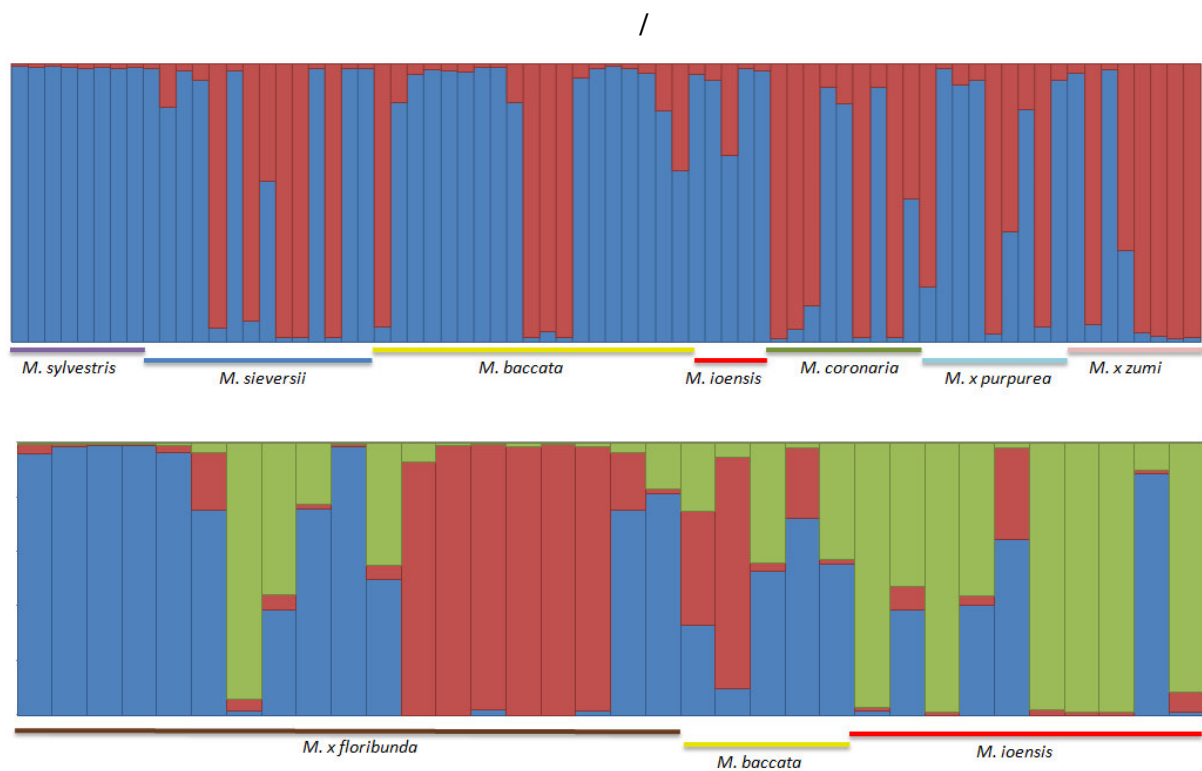
1. Bell G: **The Masterpiece of Nature: The Evolution and Genetics of Sexuality.** Berkeley, California, USA; 1982.
2. Stukenbrock EH, McDonald BA: **The origins of plant pathogens in agro-**

- ecosystems.** *Annual Review of Phytopathology* 2008, **46**:75-100.
3. Gandon S, Michalakis Y: **Local adaptation, evolutionary potential and host-parasite coevolution: interactions between migration, mutation, population size and generation time.** *Journal of Evolutionary Biology* 2002, **15**:451-462.
 4. Greischar MA, Koskella B: **A synthesis of experimental work on parasite local adaptation.** *Ecology Letters* 2007, **10**(5):418-434.
 5. Hamilton WD, Axelrod R, Tanese R: **Sexual reproduction as an adaptation to resist parasites.** *Proceedings of the National Academy of Sciences* 1990, **87**(9):3566-3573.
 6. Karasov T, Messer PW, Petrov DA: **Evidence that adaptation in *Drosophila* is not limited by mutation at single sites.** *PLoS Genet* 2010, **6**(6):e1000924.
 7. Giraud T, Gladieux P, Gavrillets S: **Linking the emergence of fungal plant diseases with ecological speciation.** *Trends in Ecology and Evolution* 2010, **25**(7):387-395.
 8. Giraud T, Villareal L, Austerlitz F, Le Gac M, Lavigne C: **Importance of the life cycle in sympatric host race formation and speciation of pathogens.** *Phytopathology* 2006, **96**(3):280-287.
 9. Alexopoulos CJ, Mims CW, Blackwell M: **Introductory mycology** 4th Edition edn. New York, USA: Wiley; 1996.
 10. Andrivon D, Pilet F, Montarry J, Hafidi M, Corbiere R, Achbani el H, Pelle R, Ellisseche D: **Adaptation of *Phytophthora infestans* to Partial Resistance in Potato: Evidence from French and Moroccan Populations.** *Phytopathology* 2007, **97**(3):338-343.
 11. McDonald BA, Linde C: **Pathogen population genetics, evolutionary potential, and durable resistance.** *Annual Review of Phytopathology* 2002, **40**:349-379.
 12. Kimura M, Weiss GH: **The stepping stone model of population structure and the decrease of genetic correlation with distance.** *Genetics* 1964, **49**(4):561-576.
 13. Brown JK, Hovmøller MS: **Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease.** *Science* 2002, **297**(5581):537-541.
 14. Jones JD, Dangl JL: **The plant immune system.** *Nature* 2006, **444**(7117):323-329.
 15. Bus VG, Rikkerink EH, Caffier V, Durel CE, Plummer KM: **Revision of the nomenclature of the differential host-pathogen interactions of *Venturia inaequalis* and *Malus*.** *Annu Rev Phytopathol* 2011, **49**:391-413.
 16. Holb IJ, Heijne B, Withagen JCM, Jeger MJ: **Dispersal of *Venturia inaequalis* ascospores and disease gradients from a defined inoculum source.** *Journal of phytopathology* 2004, **152**:639-646.
 17. MacHardy WE: **Apple scab: Biology, Epidemiology and Management.** St. Paul, MN: APS Press; 1996.
 18. Gladieux P, Guerin F, Giraud T, Caffier V, Lemaire C, Parisi L, Didelot F, B LEC: **Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants.** *Mol Ecol* 2011, **20**(21):4521-4532.
 19. Guérin F, Gladieux P, Le Cam B: **Origin and colonization history of newly virulent strains of the phytopathogenic fungus *Venturia inaequalis*.** *Fungal Genetics and Biology* 2007, **44**(4):284-292.
 20. Guérin F, Le Cam B: **Breakdown of the scab resistance gene *Vf* in apple leads to a**

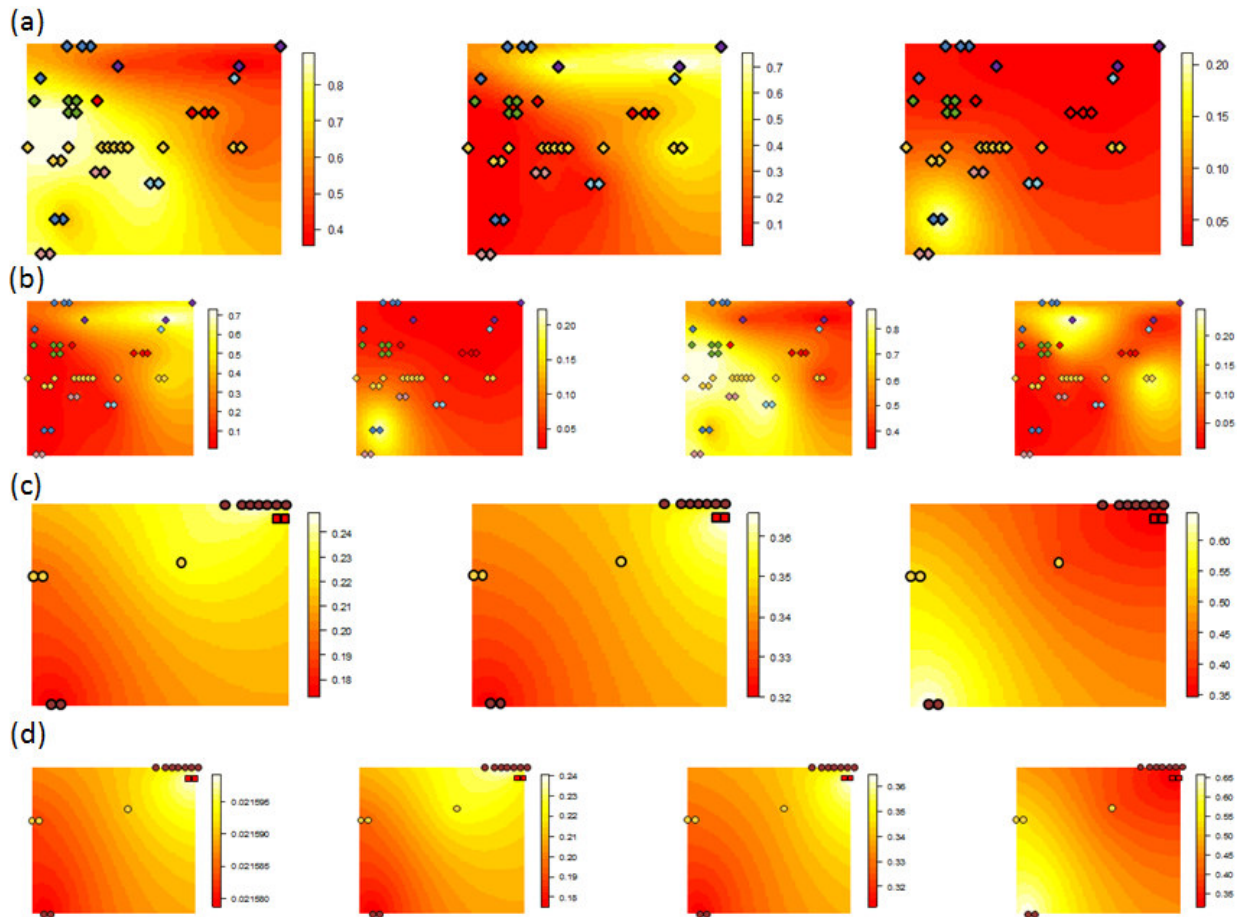
- founder effect in populations of the fungal pathogen *Venturia inaequalis*.** *Phytopathology* 2004, **94**(4):364-369.
21. Le Cam B, Devaux M, Parisi L: **Specific polymerase chain reaction identification of *Venturia nashicola* using internally transcribed spacer region in the ribosomal DNA.** *Phytopathology* 2001, **91**(9):900-904.
 22. Tenzer I, Degli Ivanissevich S, Morgante M, Gessler C: **Identification of Microsatellite Markers and Their Application to Population Genetics of *Venturia inaequalis*.** *Phytopathology* 1999, **89**(9):748-753.
 23. Guérin F, Franck P, Loiseau A, Devaux M, Le Cam B: **Isolation of 21 new polymorphic microsatellite loci in the phytopathogenic fungus *Venturia inaequalis*.** *Molecular Ecology Notes* 2004, **4**(2):268-270.
 24. Gladieux P, Zhang XG, Afoufa-Bastien D, Sanhueza RMV, Sbaghi M, Le Cam B: **On the origin and spread of the scab disease of apple: out of central Asia.** *Plos One* 2008, **3**(1).
 25. Dunemann F, Gläss R, Bartsch S, Saad Eldin M, Peil A, Bus V: **Molecular cloning and analysis of apple HcrVf resistance gene paralogs in a collection of related species.** *Tree Genetics & Genomes* 2012:1-15.
 26. Tartarini S, Gianfranceschi L, Sansavini S, Gessler C: **Development of reliable PCR markers for the selection of the Vf gene conferring scab resistance in apple.** *Plant Breeding* 1999, **118**:183–186.
 27. Vinatzer BA, Patocchi A, Tartarini S, Gianfranceschi L, Sansavini S, Gessler C: **Isolation of two microsatellite markers from BAC clones of the Vf scab resistance region and molecular characterization of scab-resistant accessions in *Malus* germplasm.** *Plant Breeding* 2004, **123**:321–326.
 28. Xu M, Korban SS: **A cluster of four receptor-like genes resides in the Vf locus that confers resistance to apple scab disease.** *Genetics* 2002, **162**(4):1995-2006.
 29. Peakall R, Smouse PE: **GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research.** *Molecular Ecology Notes* 2006, **6**:288-295.
 30. Nei M: **Molecular Evolutionary Genetics**, 1st Edn edn. New York, USA: Columbia University Press; 1987.
 31. Excoffier L, Laval G, Schneider S: **Arlequin ver. 3.0: An integrated software package for population genetics data analysis.** *Evolutionary Bioinformatics Online* 2005, **1**:47-50.
 32. Foll M, Gaggiotti O: **A genome-scan method to identify selected loci appropriate for both dominant and codominant Markers: a bayesian perspective.** *Genetics* 2008, **180**(2):977-993.
 33. Falush D, Stephens M, Pritchard JK: **Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies.** *Genetics* 2003, **164**(4):1567-1587.
 34. Falush D, Stephens M, Pritchard JK: **Inference of population structure using multilocus genotype data: dominant markers and null alleles.** *Molecular Ecology Notes* 2007, **7**(4):574-578.
 35. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations.** *American Journal of Human Genetics* 2000, **67**(1):170-181.

36. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.** *Molecular Ecology* 2005, **14**(8):2611-2620.
37. Earl DA: **Structure harvester** In., 0.6.1 edn; 2011.
38. Chen C, Durand E, Forbes F, François O: **Bayesian clustering algorithms ascertaining spatial population structure: A new computer program and a comparison study.** *Molecular Ecology Notes* 2007, **7**:747-756.
39. Durand E, Jay F, Gaggiotti OE, Francois O: **Spatial inference of admixture proportions and secondary contact zones.** *Mol Biol Evol* 2009, **26**(9):1963-1973.
40. Francois O, Ancelet S, Guillot G: **Bayesian clustering using hidden Markov random fields in spatial population genetics.** *Genetics* 2006, **174**(2):805-816.
41. Jakobsson M, Rosenberg NA: **CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure.** *Bioinformatics* 2007, **23**(14):1801-1806.
42. **A language and environment for statistical computing** [[http://www.R-project.org.](http://www.R-project.org)]
43. Hardy OJ, Vekemans X: **SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels.** *Molecular Ecology Notes* 2002:618-620.
44. Wagner HH, Holderegger R, Werth S, Gugerli F, Hoebee SE, Scheidegger C: **Variogram analysis of the spatial genetic structure of continuous populations using multilocus microsatellite data.** *Genetics* 2005, **169**(3):1739-1752.
45. Hardy OJ, Vekemans X: **Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models.** *Heredity* 1999, **83** (Pt 2):145-154.
46. Rousset F: **Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance.** *Genetics* 1997, **145**(4):1219-1228.
47. Loiselle BA, Sork VL, Nason J, Graham C: **Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae).** *American Journal of Botany* 1995, **82**:1420-1425.
48. Meirmans PG: **The trouble with isolation by distance.** *Mol Ecol* 2012, **21**(12):2839-2846.
49. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagne H: **Vegan: Community ecology package. R package version 2.0-5.** In.; 2012.
50. Gardiner SE, Bus VGM, Rusholme RL, Chagné D, Rikkerink EHA: **Apple.** In: *Fruits and Nuts Genome Mapping and Molecular Breeding in Plants.* vol. 4. Berlin: Springer; 2007: 1-62.
51. Crosby JA, Janick J, Pecknold PC, Korban SS, O'Connon PA, Ries SM, Goffreda J, Voordeckers A: **Breeding apples for scab resistance: 1945–1990.** *Fruit Varieties Journal* 1992, **46**:145–166.
52. Gessler C, Pertot I: **Vf scab resistance of Malus.** *Trees-structure and function* 2012, **26**:95-108.
53. Kaplan JD: **Dispersal gradients and deposition efficiency of *Venturia inaequalis* ascospores and their relationship to lesion densities.** Durham, NH: University New

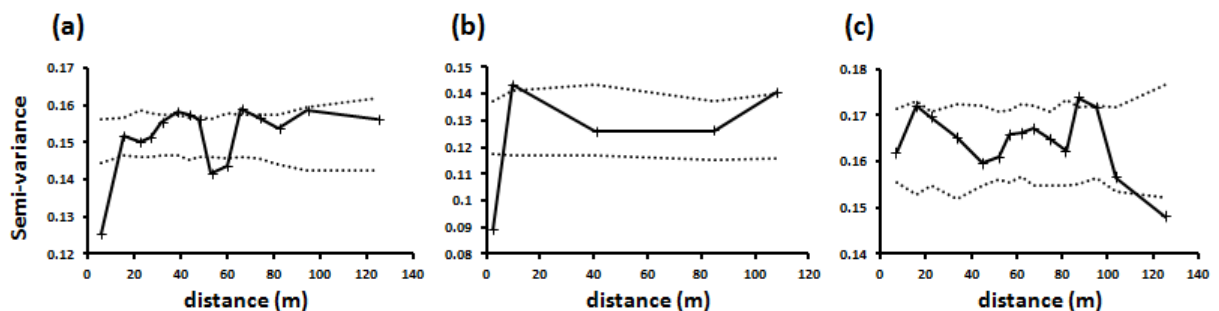
- Hampshire; 1986.
54. Burchill RT, Hutton KE: **The suppression of ascospore production to facilitate the control of apple scab (*Venturia inaequalis* [Cooke] Winter).** *Ann Appl Biol* 1965, **56**:285–292.
 55. Hirst JM, Stedman OJ: **The epidemiology of apple scab (*Venturia inaequalis* [Cke.] Wint.): III. The supply of ascospores.** *Ann Appl Biol* 1962, **50**:551–567.
 56. Palmiter DH: **Ground treatments as an aid in apple scab control.** *N Y Agric Exp St Bull* 1946, **714**:1–27.
 57. Aylor DE: **The aerobiology of apple scab.** *Plant Disease* 1998 **82**:838-849.



Supplementary data, Fig. 1: STRUCTURE individual assignments of K=2 or K=3 clusters inferred for *avrRvi6* (a) and *virRvi6* (b) subpopulations. Each haplotype is represented by a bar partitioned into K=2 or K=3 segments that represent the haplotype's estimated membership fractions in each of the two or three clusters. For each fungal haplotype, the *Malus* species where strains were sampled is indicated below each chart.



Supplementary data, Fig 2: The spatial interpolation map showing differences in the admixture population derived from the TESS assignment to three or four different clusters within each subpopulation. The interpolations for the avrRvi6 dataset, assuming $K_{max}=3$ (a) and $K_{max}=4$ (b), and for virRvi6 dataset with $K_{max}=3$ (c) and $K_{max}=4$ (d), are shown. The strains collected from non-Rvi6 hosts are represented by a diamond: *M. sieversii* (blue), *M. sylvestris* (purple), *M. coronaria* (green), *M. x purpurea* (light blue), *M. baccata* (yellow), *M. ioensis* (red) and *M. x zumi* (pink). The strains collected from Rvi6 hosts are represented by a circle: *M. x floribunda* (brown), *M. baccata* (yellow) and *M. ioensis* (red).



Supplementary data, Fig. 3: The variograms of gene diversity computed for the avrRvi6 subpopulation (a), the virRvi6 subpopulation (b) and avrRvi6 vs. virRvi6 pairwise comparisons (c). Ninety-five percent confidence interval limits were plotted as dotted lines. Gene diversity was estimated for different distance classes for each variogram: 15 for the avrRvi6 subpopulation, five for the virRvi6 subpopulation, and 15 for avrRvi6 vs. virRvi6 subpopulation pairwise comparisons.

Supplementary data, Table 1: The geographical and plant origins of the samples used in this study. The host species and accession number of the sampled tree, host localisation (rank, localisation inside the rank, X and Y coordinates) and alleles at the *Rvi6* resistance locus were reported. The following information concerns the number of sampled strains (n). * An apple genotype was classified as carrying the *Rvi6* locus (“1”) when all three alleles were detected (CH-Vf1: allele 159, AL07-SCAR: 480 bp fragment, Vfa2: 550 bp fragment), or when the *Rvi6* allele sequence was known (Dunemann, pers. comm.), (“0”) if all three allele assays were negative, (“?”) if only one or two alleles were detected, or (“nd”) for missing data.

Host	Trees	Acc Numb.	X coordinate	Y coordinate	<i>Rvi6</i> tree*	n
<i>M. sieversii</i>	1	MAL0946	21	85.5	0	1
	2	MAL0948	28	85.5	0	3
	3	MAL0949	31.5	85.5	0	3
	4	MAL0729	7	72	0	1
	5	MAL0789	14	18	0	2
	6	MAL0789	17.5	18	0	4
All <i>M. sieversii</i>						14
<i>M. sylvestris</i>	7	MAL0907	129.5	85.5	0	1
	8	MAL0925	45.5	76.5	0	3
	9	MAL0927	108.5	76.5	0	5
All <i>M. sylvestris</i>						9
<i>M. baccata</i>	10	MAL0043	0	45	0	2
	11	MAL0373	21	45	0	1
	12	MAL0328	38.5	45	0	1
	13	MAL0377	42	45	0	2
	14	MAL0377	45.5	45	0	2
	15	MAL0396	49	45	0	2
	16	MAL0396	52.5	45	0	2
	17	MAL0324	66.5	45	1	2
	18	MAL0458	70	45	0	1
	19	MAL0780	105	45	0	2
	20	MAL0780	108.5	45	0	1
	21	MAL0421	17.5	40.5	0	1
	22	MAL0421	14	40.5	0	2
	23	MAL0156	10.5	40.5	1	1
24	MAL0156	7	40.5	1	2	
All <i>M. baccata</i>						24
<i>M. ioensis</i>	25	MAL0298	35	63	0	2
	26	MAL0134	84	58.5	0	1
	27	MAL0343	91	58.5	0	1
	28	MAL0343	94.5	58.5	0	1
	29	MAL0330	105	58.5	0	5
	30	MAL0330	108.5	58.5	0	5
All <i>M. ioensis</i>						15
<i>M. coronaria</i>	45	MAL0350	3.5	63	0	2
	46	MAL0734	21	63	nd	1
	47	MAL0734	24.5	63	nd	1
	48	MAL0340	21	58.5	0	3
	49	MAL0340	24.5	58.5	0	3
All <i>M. coronaria</i>						10

Supplementary data Table 1 (continued)

	50	MAL0364	35	36	0	1
<i>M. x zumi</i>	51	MAL0364	38.5	36	0	3
	52	MAL0964	7	4.5	?	2
	53	MAL0964	10.5	4.5	?	3
<hr/>						
<i>All M. x zumi</i>						9
<hr/>						
	54	MAL0268	105	72	0	4
<i>M. x purpurea</i>	55	MAL0796	63	31.5	?	5
	56	MAL0796	66.5	31.5	?	2
	57	MAL0728	49	31.5	0	1
<hr/>						
<i>All M. x purpurea</i>						12
<hr/>						
	31	MAL0989	14	0	1	4
	32	MAL0989	17.5	0	1	5
	33	MAL0054	84	63	1	1
	34	MAL0401	91	63	1	3
<i>M. x floribunda</i>	35	MAL0401	94.5	63	1	2
	36	MAL0359	98	63	1	1
	37	MAL0359	101.5	63	1	3
	38	MAL0012	105	63	1	1
	39	MAL0012	108.5	63	1	1
<hr/>						
<i>All M. x floribunda</i>						21
<hr/>						
All trees studied						114
<hr/>						

Supplementary data, Table 2: The genetic diversity for strains sampled on each tree species. *n* represents the number of strains collected from each host, *K_a* the number of haplotypes, and *H_d* the average gene diversity calculated according to Nei (1987) and estimated from clone-corrected datasets.

Strains collected on	<i>n</i>	<i>K_a</i>	<i>H_d</i>
<i>M. x floribunda</i> 821	9	7	0.17
Others <i>M. x floribunda</i>	12	12	0.26
<i>M. sylvestris</i>	9	8	0.22
<i>M. sieversii</i>	14	14	0.33
<i>M. baccata</i>	24	24	0.33
<i>M. coronaria</i>	10	9	0.28
<i>M. x purpurea</i>	12	8	0.28
<i>M. x zumi</i>	9	9	0.32
<i>M. ioensis</i>	15	15	0.23
All	114	106	0.27

Supplementary data, Table 3: A list of strains grouped into each subpopulation. In both subpopulations, the origin of samples (tree species and genotype at the *Rvi6* locus) is reported. For each strain, main membership to one of the two clusters inferred by STRUCTURE was also reported (1 for red and 2 for blue in Fig. 2b). * An apple genotype was classified as carrying the *Rvi6* locus (“yes”) when all three alleles were detected (CH-Vf1: allele 159, AL07-SCAR: 480 bp fragment, Vfa2: 550 bp fragment), or when the *Rvi6* allele sequence was known (**), (“no”) if all three allele assays were negative, (“?”) if only one or two alleles were detected, or (“nd”) for missing data.

Subpop.	Strain	Host species	<i>Rvi6</i> tree*	Structure cluster
vir <i>Rvi6</i>	08M209	<i>M. x floribunda</i> 821	yes	1
	08M395a	<i>M. x floribunda</i> 821	yes	1
	08M395b	<i>M. x floribunda</i> 821	yes	1
	08M396	<i>M. x floribunda</i> 821	yes	1
	08M397	<i>M. x floribunda</i> 821	yes	1
	08M211	<i>M. x floribunda</i> 821	yes	1
	08M237	<i>M. x floribunda</i> 821	yes	1
	08M239	Others <i>M. x floribunda</i>	yes	1
	08M243	Others <i>M. x floribunda</i>	yes	1
	08M401b	Others <i>M. x floribunda</i>	yes	1
	08M406	Others <i>M. x floribunda</i>	yes	1
	08M407	Others <i>M. x floribunda</i>	yes	1
	08M408	Others <i>M. x floribunda</i>	yes	1
	08M409	Others <i>M. x floribunda</i>	yes	1
	08M410	Others <i>M. x floribunda</i>	yes	1
	08M411	Others <i>M. x floribunda</i>	yes	1
	08M412	Others <i>M. x floribunda</i>	yes	1
	08M214	Others <i>M. x floribunda</i>	yes	1
	08M233	Others <i>M. x floribunda</i>	yes	1
	08M162a	<i>M. baccata</i>	yes	1
	08M162b	<i>M. baccata</i>	yes	1
	08M175	<i>M. baccata</i>	yes	1
	08M177	<i>M. baccata</i>	yes	1
	08M178	<i>M. baccata</i>	yes	1
	08M199a	<i>M. ioensis</i>	yes**	1
	08M199b	<i>M. ioensis</i>	yes**	1
	08M200a	<i>M. ioensis</i>	yes**	1
	08M200b	<i>M. ioensis</i>	yes**	1
	08M2010	<i>M. ioensis</i>	yes**	1
	08M202b	<i>M. ioensis</i>	yes**	1
	08M203	<i>M. ioensis</i>	yes**	1
	08M204	<i>M. ioensis</i>	yes**	1
	08M205	<i>M. ioensis</i>	yes**	1
	08M206	<i>M. ioensis</i>	yes**	1

Supplementary data, Table 3: A list of strains grouped into each subpopulation (continued).

Subpop.	Strain	Host species	Rvi6 tree*	Structure cluster
avrRvi6	08M116	<i>M. sylvestris</i>	no	2
	08M128	<i>M. sylvestris</i>	no	2
	08M129	<i>M. sylvestris</i>	no	2
	08M130b	<i>M. sylvestris</i>	no	2
	08M133a	<i>M. sylvestris</i>	no	2
	08M133c	<i>M. sylvestris</i>	no	2
	08M134a	<i>M. sylvestris</i>	no	2
	08M135	<i>M. sylvestris</i>	no	2
	08M91	<i>M. sieversii</i>	no	2
	08M92	<i>M. sieversii</i>	no	2
	08M94a	<i>M. sieversii</i>	no	2
	08M94b	<i>M. sieversii</i>	no	2
	08M95	<i>M. sieversii</i>	no	2
	08M96	<i>M. sieversii</i>	no	2
	08M97	<i>M. sieversii</i>	no	2
	08M98a	<i>M. sieversii</i>	no	2
	08M108	<i>M. sieversii</i>	no	2
	08M110	<i>M. sieversii</i>	no	2
	08M111	<i>M. sieversii</i>	no	2
	08M112	<i>M. sieversii</i>	no	2
	08M113	<i>M. sieversii</i>	no	2
	08M114	<i>M. sieversii</i>	no	2
	08M136	<i>M. baccata</i>	no	2
	08M137	<i>M. baccata</i>	no	2
	08M146	<i>M. baccata</i>	no	2
	08M153	<i>M. baccata</i>	no	2
	08M154	<i>M. baccata</i>	no	2
	08M155	<i>M. baccata</i>	no	2
	08M156	<i>M. baccata</i>	no	2
	08M157	<i>M. baccata</i>	no	2
	08M158	<i>M. baccata</i>	no	2
	08M159	<i>M. baccata</i>	no	2
	08M160	<i>M. baccata</i>	no	2
	08M161	<i>M. baccata</i>	no	2
	08M163	<i>M. baccata</i>	no	2
	08M167	<i>M. baccata</i>	no	2
	08M168	<i>M. baccata</i>	no	2
	08M170	<i>M. baccata</i>	no	2
	08M172	<i>M. baccata</i>	no	2
	08M173	<i>M. baccata</i>	no	2
	08M174	<i>M. baccata</i>	no	2
	08M180b	<i>M. ioensis</i>	no	2
	08M181a	<i>M. ioensis</i>	no	2
	08M185	<i>M. ioensis</i>	no	2
08M192	<i>M. ioensis</i>	no	2	
08M195	<i>M. ioensis</i>	no	2	
08M248b	<i>M. coronaria</i>	no	2	
08M249b	<i>M. coronaria</i>	no	2	
08M294a	<i>M. coronaria</i>	nd	2	
08M297b	<i>M. coronaria</i>	nd	2	

Supplementary data, Table 3: A list of strains grouped into each subpopulation (continued).

Subpop.	Strain	Host species	Rvi6 tree*	Structure cluster
	08M251b	<i>M. coronaria</i>	no	2
	08M251c	<i>M. coronaria</i>	no	2
	08M256a	<i>M. coronaria</i>	no	2
	08M257b	<i>M. coronaria</i>	no	2
	08M258	<i>M. coronaria</i>	no	2
	08M261	<i>M. x zumi</i>	no	2
	08M267	<i>M. x zumi</i>	no	2
	08M268	<i>M. x zumi</i>	no	2
	08M269	<i>M. x zumi</i>	no	2
	08M271	<i>M. x zumi</i>	no	2
	08M273	<i>M. x zumi</i>	?	1
	08M275	<i>M. x zumi</i>	?	2
	08M278	<i>M. x zumi</i>	?	2
	08M279	<i>M. x zumi</i>	?	2
	08M306	<i>M. x purpurea</i>	no	2
	08M307	<i>M. x purpurea</i>	no	2
	08M309a	<i>M. x purpurea</i>	no	2
	08M309b	<i>M. x purpurea</i>	no	2
	08M315	<i>M. x purpurea</i>	?	2
	08M321	<i>M. x purpurea</i>	?	2
	08M327	<i>M. x purpurea</i>	?	2
	08M329	<i>M. x purpurea</i>	?	2

Supplementary data, Table 4: The genetic diversity for each subpopulation. *n* represents the number of strains in each subpopulation, *K_a* the number of unique haplotypes estimated for non-clone-corrected datasets, *H_d* the average gene diversity, and *A* the average number of alleles estimated from clone-corrected datasets. *H_d* was calculated according to Nei (1987).

Subpop.	<i>n</i>	<i>K_a</i>	<i>H_d</i>	<i>A</i>
avrRvi6	36	34	0.26	1.84
virRvi6	78	72	0.33	2.69
All	114	106	0.30	2.27

Supplementary data, Table 5: Results of standard, stratified, and partial Mantel tests for avrRvi6 and virRvi6 subpopulations. Stars indicate the significance level: * *p*<0.05, ***p*< 0.01, and ****p*<0.001. See Materials and Method section for explanation on Mantel tests and [48] for more details on the procedure.

Matrix A	Matrix B	Adjustment	Mantel's <i>r</i>	
			avrRvi6 subpop.	virRvi6 subpop.
Genetic	Geographic	None	0.146***	0.150**
Genetic	Geographic	Stratified: permuted within clusters	0.146***	0.150**
Genetic	Geographic	Partial: clusters as covariate	0.145***	0.153**
Genetic	Clusters	Partial: geography as covariate	0.058	0.032

Thèse de Doctorat

Thibault Leroy

Conséquences évolutives d'un contact secondaire chez les pathogènes: une alternative aux hypothèses écologiques?

Mise en évidence d'un contact secondaire avec flux de gènes chez *V. inaequalis*: conséquences génétiques et épidémiologiques.

Résumé

Les programmes d'amélioration variétale puisent souvent leurs sources de résistance dans les ressources génétiques sauvages. Après déploiement des variétés, les résistances de type monogénique sont souvent rapidement contournées dans les agrosystèmes par des souches virulentes. L'émergence de la virulence peut avoir deux origines : soit elle est acquise à partir d'une nouvelle mutation à l'intérieur des agrosystèmes, soit elle préexiste dans le compartiment non-cultivé. Sous ce second scénario, l'émergence conduit à des contacts secondaires potentiellement suivis de flux de gènes dans les agrosystèmes. Dans cette thèse, nous étudions les conséquences génétiques et épidémiologiques associées à ces contacts secondaires entre des populations avirulentes et virulentes, en utilisant comme modèle l'émergence des populations virulentes du champignon *Venturia inaequalis* sur les pommiers porteurs du gène de résistance *Rvi6*.

Notre approche de génomique des populations démontre que le contact secondaire révèle l'existence d'un flux de gènes hétérogène entre la population invasive virulente et la population résidente avirulente dû à des barrières génétiques endogènes (DMI) accumulées pendant la période de divergence allopatric. L'analyse d'une descendance a permis de cartographier ces barrières et de caractériser leur nature post-zygotique. Nous montrons que la recombinaison au verger entre les deux génomes divergents génère de nouvelles combinaisons génétiques plus agressives sur des variétés commerciales sensibles par cumul de facteurs d'agressivité provenant des deux génomes. Ces travaux pointent le risque épidémiologique d'émergence de nouvelles souches recombinantes favorisée par contact secondaire dans les agrosystèmes suite au déploiement de nouveaux gènes de résistance.

Mots clés

Barrières génétiques, incompatibilités de Dobzhansky-Muller, virulence, agressivité, gène de résistance *Rvi6* (*Vf*), hybridation intraspécifique, introgression, invasions, tavelure du pommier

Abstract

Plant breeding programs often use resistances from wild genetic resources. Once spatially deployed, these monogenic resistances are often quickly broken down in agro-ecosystems by virulent strains. The emergence of this virulence may have two origins: either it is acquired by mutation, either it migrates from the wild where it preexists. Under this second scenario, a secondary contact potentially followed by gene flow between these two populations is expected in agro-ecosystems. During this thesis, genetic and epidemiological consequences of this secondary contact were studied, using as a model the emergence of virulent populations of the fungus *Venturia inaequalis* on apple carrying the *Rvi6* resistance gene.

Using a population genomics approach, we demonstrated that secondary contact highlighted the occurrence of a heterogeneous gene flow between invading virulent and resident avirulent population due to endogenous genetic barriers (DMI) accumulated during allopatric divergence. Analysis of progenies permitted to map these barriers and to reveal their post-zygotic trait. We showed that recombination in orchards between the two divergent genomes may produce particularly aggressive strains on susceptible commercial cultivars. Overall, our study points out the epidemiological risk of emergence of new recombinant strains favored by secondary contacts in agro-ecosystems by deployment of new resistance genes.

Keywords:

Genetic barriers, Dobzhansky-Muller incompatibilities, virulence, aggressiveness, *Rvi6* resistance gene (*Vf*), intraspecific hybridization, introgression, invasions, apple scab