Hybridations inter-spécifiques chez le pommier et co-évolution hôte-pathogène

Alice Feurtey

▶ **To cite this version:**

Alice Feurtey. Hybridations inter-spécifiques chez le pommier et co-évolution hôte-pathogène. Génétique des populations [q-bio.PE]. Université Paris Saclay (COmUE), 2016. Français. NNT : 2016SACLS446 . tel-01941395

HAL Id: tel-01941395
https://theses.hal.science/tel-01941395

Submitted on 1 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NNT : 2016SACLS446

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS-SACLAY, préparée à l'Université Paris-Sud

ÉCOLE DOCTORALE N° 567
Sciences du Végétal : du Gène à l'Ecosystème

Biologie

Par

Madame Alice Feurtey

Hybridations inter-spécifiques chez le pommier et co-évolution hôte-pathogène

Thèse présentée et soutenue à Orsay, le 29 novembre 2016 :

Composition du Jury :

M. Dominique de Vienne	Professeur, Université Paris-Sud	Président du jury
M^{me} Véronique Decroocq	DR, INRA	Rapporteur
M. Rémy Petit	DR, INRA	Rapporteur
M. Pascal Frey	DR, INRA	Examinateur
M^{me} Tatiana Giraud	DR, CNRS	Directrice de thèse

Seen in the light of evolution, biology is, perhaps, intellectually the most satisfying and inspiring science. Without that light it becomes a pile of sundry facts -- some of them interesting or curious but making no meaningful picture as a whole.

Theodosius Dobzhansky, 1973

Remerciements :

À la place d'honneur, je remercie Rémy Petit et Véronique Decroocq pour avoir accepté la lourde tache d'être rapporteurs de ma thèse ainsi que Pascal Frey et Dominique de Vienne pour avoir accepté d'être examinateurs. Je voudrais aussi remercier les membres de mon comité de thèse Maud Tenaillon, Etienne Klein, Yves Vigouroux et Christian Parisod. Vos conseils avisés ont été précieux au cours des réunions que nous avons eues !

Je remercie également, bien évidemment, Tatiana Giraud pour ces trois années (et demi avec le stage de M2). C'était un plaisir de travailler et d'apprendre avec toi. Vraiment merci d'avoir été toujours disponible quand j'avais des résultats à discuter ou une question à poser, malgré les 70 projets (au moins !) que tu suis en même temps ! Merci aussi d'être super efficace pour les relectures ou la réponse aux mails, ça nous facilite l'existence à tous au quotidien, dans l'équipe.

Je voudrais également remercier Christophe Lemaire qui m'a proposé de travailler avec lui sur le projet *Venturia*/pommier. Je suis ravie que tu ais pensé à moi comme le sujet m'intéresse depuis que tu nous as donné un cours dessus en L3 ! Petit secret : c'est d'ailleurs grâce à ce même cours que j'ai répondu à l'offre de stage de Tatiana et tu es donc un peu à l'origine de ma thèse ! Merci également à Bruno Le Cam pour m'avoir incluse dans votre projet sur lequel j'ai adoré travailler !

Merci aussi à Amandine Cornille qui m'a encadrée durant mon stage de M2 et qui m'a donc appris toutes les bases dont j'ai eu besoin au cours de ma thèse et qui aussi continué à suivre les projets et m'aider en cas de besoin, malgré tous les projets qu'elle menait de front pendant ses post-docs. Juste un exemple parmi d'autres, merci pour les mini-cours sur l'ABC !

Merci aussi à Pierre qui m'a donné l'occasion de travailler sur l'invasion de *Microbotryum lychnidis-dioicae* et qui m'a beaucoup aidée pour la rédaction de mon premier article. Parmi les gens avec qui j'ai collaboré, je ne peux pas non plus oublier de remercier Jacqui

Shykoff. Désolée et merci pour tous ces après-midis passées à s'arracher les cheveux sur la meilleure méthode statistique alors que tu devais m'expliquer plein de choses en même temps ! Je remercie aussi Antoine Branca pour ses conseils théoriques et techniques ainsi que pour les discussions sur les analyses de génomique. Merci aussi à Douae et William, qui ont fait leurs stages avec moi. J'espère que vous avez trouvé votre voie !

J'ai aussi eu la chance au cours de ma thèse d'interagir avec des professionnels et des amateurs tous passionnés de pommes et/ou de conservation des forêts. Parmi les (très nombreuses) personnes qui ont récolté des pommiers sauvages pour nous et que je remercie tous, je voudrais mentionner en particulier Yves Lespinasse et Françoise Le Berre. Parmi ceux qui ont participé au projet entourant la forêt de Dourdan et son verger à graine, merci à Didier Desseaux, Éric Collin et Brigitte Musch, ainsi que tous les autres de l'ONF et de l'IRSTEA. Du côté associatif, je veux aussi remercier Charly Chauveau et Séverin Lavoyer. Merci aussi à Christian Tailliez avec qui j'ai collaboré sur un projet pour les lycéens, c'est un plaisir de voir que certains enseignants s'investissent à ce point ! Il y a eu tellement de gens que je ne peux pas citer tout le monde, mais ce fût un vrai plaisir de travailler avec vous tous et de ne pas rester coincée dans ma « tour d'ivoire » de chercheur !

J'ai souvent l'impression que, dans le monde de la recherche, on traite un peu l'enseignement comme parasitant notre temps de recherche. N'ayant pas enseigné moi-même, je ne peux pas vraiment donner un avis informé là-dessus, mais je voudrais quand même souligner que certains de mes enseignants m'ont vraiment influencée positivement et que ce sont eux qui m'ont donné envie de me lancer dans la recherche. En master, ça a été en particulier Dominique de Vienne qui a organisé l'UE me donnant envie de faire de la génomique qui a été la plus difficile mais aussi et surtout la plus instructive et la plus utile ! En licence, ça a été Christophe, que j'ai déjà remercié, mais aussi Sabine Castanier-Perthuisot et Didier Peltier. Tous ces enseignants ont su partager leur passion et m'ont donné envie de faire de la génétique, de l'évolution et de la microbiologie des intérêts de toute une vie. Merci à tous !

J'aimerais aussi remercier les équipes de médiation de bio et du journal du Palais de la Découverte. L'esprit d'équipe qui règne là-bas et la gentillesse de chacun m'a vraiment motivée et touchée. Vous m'avez également donné envie de continuer à communiquer avec le public grâce à votre enthousiasme communicatif ! Merci à l'équipe du journal Découverte avec qui j'ai eu la chance de travailler. Petite note spéciale pour Marie, bien sûr, mais aussi Tanguy, Elodie, Aurélie, Stéphanie, Gaëlle et Jean-Philippe.

Parmi les gens du labo qui ont contribué à faire de ces trois années de thèses des années funs et agréables, je veux aussi remercier plein de gens ! Emilie, merci d'avoir passé toutes ces

heures perdues au milieu des scripts Python et des analyses ABC avec moi ! Alodie, grâce à toi, j'ai découvert plein de recettes plus ou moins bizarres et des produits naturels. Nos petites séances à toutes les trois mêlant travail, bavardages et potins ont bien aidé à faire passer ces trois années super vite ! Qui sait si on ne retournera pas faire du pilate toutes les trois quand on aura monté notre labo (où on ne tuera pas de souris, mais où il y aura pleins de micro-organismes) !

Hélène, merci pour toute ton aide ! Grâce à toi, Emilie et moi sommes maintenant des ratons-laveurs dans le monde de python au lieu d'être de petits colibris. Ça nous a fait un vide quand tu as fini ta thèse !

Jérôme, tu as beau n'avoir passé que quelques mois avec nous, tu as réussi à super bien t'intégrer dans notre groupe et à faire partie des « potes/collègues » ce qui t'a bien fait rire ! Ne nous oublie pas dans les contrées glacées de l'autre côté de l'océan !

Aurélien, c'était cool d'avoir un camarade de galère de thèse pendant ces trois années ! Ton humour incompréhensible pour certains a fait passer pas mal de situations pénibles ! Et tu imagines, très bientôt, on va devoir prendre l'habitude d'aller à la cantine séparément dans nos futurs labos, les filles vont être choquées !

Stéphanie, c'était un plaisir de te connaitre. Ta bonne humeur et ton sourire ont manqué à tout le labo quand tu es partie !

I also want to mention a few other people with whom I was very happy to spend time. Sara, your energy and no-nonsense attitude was very welcome and a good wake-up call! I wish you all the luck and happiness in your own adventure next year! Ying-Chu, please do not get too anxious about your PhD, it is not always easy, but it is also a great experience and not anything to be (too much) afraid of! Thanks also to Ricardo, Taia, Sylvie, Noppol, Qian, Shanerin & Stefan!

J'ai aussi bien sur des gens à remercier en dehors du labo.

Commençons par Amandine et Floriane. Je suis vraiment contente que vous ayez décidé de venir vous installer juste à côté de chez nous ! Ces trois ans n'auraient pas été les mêmes sans pouvoir passer toutes ces heures à se raconter nos commérages ! Nos soirées et les sorties de mamies au marché du dimanche matin vont bien me manquer !

Je veux aussi dire un grand merci à mes parents. Ils m'ont élevée en me laissant une complète liberté de choisir ma propre voie et de trouver ce qui me rend heureuse, sans préjugés. J'ai toujours su que je pourrais faire toutes les études que je voudrais et qu'ils me soutiendraient (financièrement et mentalement). J'ai bien conscience que c'est un luxe et une chance que tous n'ont pas. Merci, maman. Merci, mon père.

Merci aussi au reste de ma famille. Ils m'ont beaucoup soutenue, même si, dans l'ensemble, ce que j'ai fait pendant ma thèse reste très obscur pour la plupart d'entre eux ! En particulier, merci à Ophélie (= petite sœur) et Charline (=mini-sœur) avec qui je peux toujours me plaindre, rigoler, partager et me détendre ! Ce sont mes sœurs mais aussi mes amies ce qui est formidable !

Et pour finir, merci à Fred ! Je ne sais pas comment j'aurais fait ces trois dernières années sans toi. (Ou, accessoirement, pour faire la mise en page de ce manuscrit !) Merci de supporter mes obsessions/sautes d'humeur au quotidien, mon ami !

Prologue

Dans le cadre de ma thèse, je me suis intéressée à deux grands thèmes de la biologie évolutive. Dans ce manuscrit, je traiterais ces deux thèmes en deux sections distinctes bien que les organismes étudiés et les techniques utilisées relient partiellement les deux sections.

Dans une première section, je présente mon travail sur les hybridations interspécifiques et l'histoire évolutive des espèces de pommiers en Europe. Les questions y découlent de l'histoire évolutive du pommier cultivé, incluant une étape présumée de domestication à partir de *Malus sieversii*, une espèce de pommiers sauvages trouvée en Asie Centrale, puis un transport vers l'Europe qui a permis des échanges de gènes avec d'autres espèces de pommiers sauvages. Dans la première partie de cette section, j'ai étudié les conséquences de ces flux de gènes sur l'espèce sauvage apparentée *M. sylvestris*, trouvée disséminée dans les forêts et agrosystèmes européens. La seconde partie de cette section concerne l'implication des hybridations entre *M. domestica* et *M. sylvestris* dans la domestication de *M. domestica*.

Dans une seconde section, je me suis intéressée à la co-évolution entre espèces de plantes et espèces de champignons pathogènes dans deux systèmes différents. La première partie de cette section s'intéresse à l'histoire évolutive combinée des pommiers cultivés et sauvages d'Asie Centrale et du champignon pathogène *V. inaequalis* infectant ces deux espèces hôtes. Dans la seconde partie de la section, j'ai comparé les structures spatiales à l'échelle européenne de la plante *Silene latifolia* et de son champignon pathogène *Microbotryum lychnidis-dioicae.*

TABLES DES MATIÈRES

Remerciements : 5
Prologue 9
SECTION 1 : HYBRIDATIONS INTER-ESPÈCES 15

I. Introduction générale de la section 1 17

Espèces : concepts & définitions 17
Hybridations interspécifiques 18
Conséquences des flux de gènes et hybridations pour les espèces domestiquées 19
Conséquences des flux de gènes et hybridations pour les espèces sauvages 22

Comment la diversité est-elle distribuée à l'échelle mondiale ? 23
À quelles menaces fait face la diversité ? 24
Quelles actions peuvent être entreprises afin de réduire ces menaces ? 26

II. Introduction aux espèces étudiées : *Malus* sp. 27

Importance culturelle et économique de la pomme 27
Les hybridations inter-espèces au sein du genre Malus 29
Histoire du pommier cultivé 29

III. Articles 33

Article 1. Forces motrices anthropogéniques et naturelles des flux de gènes chez un fruitier sauvage des zones tempérées : une base pour la conservation et les programmes d'amélioration variétale chez les pommes 33
Article 2. Flux de gènes des espèces cultivées aux espèces sauvages et ses conséquences sur la valeur sélective d'un fruitier sauvage : vers une stratégie globale de conservation du pommier sauvage en Europe 73
Article 3. Une histoire génomique de la domestication du pommier 93

IV. Discussion de la section 1 111

Flux de gènes des espèces sauvages vers les espèces cultivées 111

Structure génétique et domestication 111

Détection des introgressions 112
Sous-structuration chez le pommier cultivé 115
Flux de gènes des espèces cultivées vers les espèces sauvages 117
Structure des populations : étude des climats passés et conservation 118
D'importants flux de gènes influencés par des facteurs anthropiques 119
Valeur sélective aux premiers stades de développement 119
Capacités de dispersion 120
Etude des filières existantes de vente de Malus sylvestris 121
Mise en pratique des conclusions obtenues 123
SECTION 2 : RELATIONS ENTRE HÔTES ET PATHOGÈNES 125
I. Introduction générale de la section 2 127
Interactions entre organismes et évolution 127
Différents types d'interactions entre organismes 127
Les pathogènes et leur impact évolutif 129
La coévolution entre hôte et pathogènes 130
De la coévolution à la différenciation chez les pathogènes 130
L'adaptation locale comme une mesure de la co-évolution 130
Bases moléculaires de la coévolution hôte-pathogène 131
Le modèle gène-pour-gène 132
Les pathogènes dans les agrosystèmes 133
Scénarios évolutifs d'émergence des pathogènes dans les agrosystèmes 133
Résistance chez les plantes cultivées 134
II. Les pommiers et leur pathogène *Venturia inaequalis* 137
Introduction aux espèces étudiées : *Venturia inaequalis* 137
Découverte de deux populations de V. inaequalis en Asie Centrale 140
Le phénotype de V. inaequalis a été affecté par le processus de « pestification » . 140
Article 4. Conséquences de l'introduction des pommiers cultivés dans leur zone d'origine : une histoire sur l'hôte et sur son pathogène 143

III. Les silènes et leur pathogène *Microbotryum violaceum* ... 163

Introduction aux espèces étudiées : *Microbotryum – Silene* ... 163

Invasion des Etats-Unis par Silene latifolia et Microbotryum lychnidis-dioicae ... 165

Structure génétique spatiale chez Silene latifolia et Microbotryum lychnidis-dioicae ... 166

Article 5. Costructure chez Silene latifolia et Microbotryum lychnidis-dioicae ... 169

IV. Discussion de la section 2 ... 201

Venturia - Malus ... 201

Conséquences du contact secondaire chez le pommier ... 202

Conséquences du contact secondaire chez V. inaequalis ... 203

Capacités d'infection de V. inaequalis ... 205

Microbotryum lychnidis-dioicae -Silene latifolia ... 206

Une structure phylogéographique commune au pathogène et à son hôte ... 206

Des flux de gènes limités entre groupes génétiques ... 207

Une sous-structuration présente chez le pathogène, mais pas chez son hôte ... 207

Conclusion ... 209

Bibliographie ... 213

Annexes ... 225

Annexe 1 : Article de vulgarisation « Domestication, une révolution à l'origine de la civilisation » ... 225

Annexe 2 : Copie des conseils mis à disposition par le ministère de l'agriculture concernant *Malus sylvestris* ... 226

Annexe 3 : Article sur l'invasion de Microbotryum ... 231

N.B. : Dans les parties en français, correspondant au texte rédigé spécifiquement pour cette thèse, la numérotation des figures est faite en fonction de la section. La numérotation des figures des articles, rédigés en anglais, est interne à chaque article. De même, les références citées dans les articles sont listées à la fin de chaque article et une bibliographie séparée se trouve à la fin du manuscrit correspondant aux références citées dans les parties en français.

SECTION 1

HYBRIDATIONS INTER-ESPÈCES

I. Introduction générale de la section 1

Espèces : concepts & définitions

Les débats sur la définition et les concepts d'espèces font rage régulièrement chez les biologistes de l'évolution et de la systématique. Ainsi, le soi-disant « concept » d'espèce phylogénétique rassemble les populations en une espèce lorsqu'elles partagent au moins un caractère dérivé absent de toutes les autres populations étudiées (Bergstrom & Dugatkin, 2012), tandis que le « concept » d'espèce écologique, lui, souligne plutôt l'importance de l'adaptation à une niche écologique particulière (Cai et al., 2011). Le « concept » d'espèce biologique de Mayr, délimitant les espèces comme des groupes de populations naturelles qui sont interfertiles entre elles mais pas avec d'autres groupes, a eu une grande influence sur les scientifiques, mais aussi dans la culture populaire. Il est fréquent, par exemple, d'illustrer le « concept » d'espèce avec l'exemple du cheval et de l'âne, inter-fertiles, mais dont la descendance est stérile. Mais cette « définition » pose régulièrement de nombreuses questions. Qu'en est-il d'une grande partie de la biodiversité de la planète, asexuée ? Ou encore comment prendre en compte les cas d'hybridations occasionnels entre individus de populations qui sont dans leur ensemble isolées génétiquement ? Ainsi comment nommer deux populations dont les hybridations sont peu fréquentes ou même assez fréquentes, tout en restant différenciées écologiquement et génétiquement ? Et que dire si les descendants de ces deux populations ne sont pas stériles mais ont potentiellement une fertilité réduite ?

Cependant, toutes ces contradictions et difficultés apparentes disparaissent quand on réalise que 1) ces soi-disant « concepts » ne sont pas des définitions de l'espèce mais plutôt des critères permettant de délimiter les espèces sur le terrain. Ces critères ne peuvent en outre pas être universels puisque, notamment, la spéciation n'est pas un phénomène instantané, mais un processus évolutif lent agissant sur des organismes aux traits d'histoire de vie variés (par exemple, le mode de reproduction cité plus haut). Le critère optimal permettant de délimiter les

espèces (interstérilité, marqueurs génétiques) dépend de ces traits d'histoire de vie et du mode de spéciation. Il est alors plus simple et plus unanime de définir les espèces comme des segments de lignées évolutives ayant évolué de façon indépendante les unes des autres (Queiroz, 2007). Cette définition implique aussi que, la spéciation étant un processus graduel et lent, le niveau de l'espèce n'est pas une limite « tout ou rien » et qu'il existe donc des entités biologiques plus ou moins isolées génétiquement les unes des autres, pour lesquelles fixer une limite d'espèce va être subjectif.

Hybridations interspécifiques

Lorsque l'on ne considère plus les espèces sous la seule lumière de l'un des « concepts » d'espèce cités ci-dessus, on peut dont reconnaître l'existence de lignées évolutives traitées comme des espèces et portant des noms latins d'espèces différents, puisque différenciées génétiquement et écologiquement, mais qui ne sont pas complètement isolées génétiquement. On peut alors observer des hybridations dites interspécifiques entre ces lignées. Celles-ci sont beaucoup plus fréquentes que les idées préconçues pourraient le faire penser et pourrait avoir lieu chez 10% des animaux et chez 25% des plantes (Mallet, 2005). Les hybridations interspécifiques peuvent avoir des conséquences importantes sur l'évolution des lignées impliquées, notamment chez les plantes (Ellstrand, 2014). Ces conséquences sont très variables et ont parfois même des effets contradictoires selon les cas considérés. Ainsi, les hybridations interspécifiques peuvent causer la disparition des barrières aux flux de gènes ou au contraire les renforcer, par une sélection contre les hybrides à faible valeur sélective (Abbott et al., 2013). Dans certains cas, elles peuvent même conduire à des spéciations, par création d'une troisième lignée d'ancestralité mixte (hybride) et se maintenant distincte des deux lignées parentales (Mallet, 2007). Lorsque les hybridations entre deux espèces produisent des hybrides qui sont eux-mêmes fertiles et capables de se croiser de nouveau avec les deux espèces parentes, la possibilité d'obtenir des introgressions apparait, c'est-à-dire que certains allèles de l'une des espèces peuvent être incorporés dans le génome de la seconde espèce à travers des back-cross répétés (Ellstrand, Prentice, & James, 1999).

L'existence de barrières semi-perméables entre les espèces a des conséquences importantes, par exemple en cas de contact secondaire provoqué par les Hommes. Ce type de contact peut être induit par des modifications de l'environnement, par des introductions accidentelles ou encore par des introductions volontaires (R. G. Harrison & Larson, 2014). Des espèces naturellement allopatriques peuvent alors avoir des aires de répartition devenant au

moins en partie chevauchantes, ce qui peut entraîner des flux de gènes entre elles. C'est notamment le cas lors d'introduction volontaire d'espèces domestiquées dans des territoires nouveaux pour étendre leur exploitation (voir annexe 1 pour un article de vulgarisation sur la domestication et ses implications pour la civilisation humaine). L'existence de flux de gènes entre espèces domestiquées et leurs apparentés sauvages a ainsi été démontrée dans de nombreux cas, y compris chez les animaux d'élevage (Feulner et al., 2013) et chez les plantes cultivées (Ellstrand et al., 2013).

Les flux de gènes causés par les hybridations entre espèces de plantes domestiquées et sauvages peuvent exister dans les deux sens, c'est-à-dire des espèces cultivées vers leurs apparentés sauvages, mais aussi des espèces sauvages vers les espèces cultivées. L'apparition et le maintien d'hybrides, que ce soit dans le milieu cultivé ou le milieu sauvage a, dans certains cas, des conséquences écologiques et/ou économiques qui ne sont pas négligeables (Ellstrand et al., 2013). Ces flux de gènes peuvent donc influencer le devenir des espèces sauvages comme des espèces cultivées et c'est cela que j'ai étudié dans la première section de ma thèse.

Conséquences des flux de gènes et hybridations pour les espèces domestiquées

Les hybridations entre espèces sauvages et cultivées peuvent avoir des conséquences économiques négatives pour les agriculteurs lorsque des plantes hybrides reçoivent d'un parent cultivé des résistances à certains pesticides ou herbicides et d'un parent sauvage des caractères morphologiques économiquement peu intéressants. Ces hybrides peuvent alors prospérer au milieu des cultures et réduire la production. Cela a été montré dans le cas notamment des betteraves et des carottes (Ellstrand et al., 2013).

A l'inverse, si les caractères transmis des espèces sauvages aux espèces domestiquées sont intéressants du point de vue humain, les flux de gènes peuvent avoir des conséquences économiques positives. Cela peut être le cas pour la transmission de caractères de résistances aux stress biotiques et abiotiques. Par exemple, les hybridations entre le maïs et la téosinte sauvage, *Zea mays spp. mexicana*, ont probablement contribué à la résistance de certaines variétés de maïs cultivées aux conditions environnementales des hauts plateaux mexicains (Hufford et al., 2013). Dans certains cas, les hybridations entre espèces domestiquées et sauvages sont mêmes encouragées et activement induites dans les programmes de sélection variétales afin de permettre l'intégration dans le génome d'espèces domestiquées de gènes conférant des phénotypes jugés intéressants. Ces introgressions de gènes sauvages peuvent avoir différentes

conséquences sur les variétés créées, notamment l'amélioration de la qualité du fruit, comme chez le pêcher (Quilot et al., 2004), la résistance à des stress abiotiques, comme dans le cas du maïs cité plus haut, et surtout la résistance à différents pathogènes, introduite à partir d'apparentés sauvages chez de très nombreuses espèces dont les bananes, le riz, les tomates et le blé (Hajjar & Hodgkin, 2007).

Chez certaines espèces, les hybridations inter-spécifiques ont eu un rôle crucial dès leur domestication en créant de nouvelles espèces de plantes cultivées. Ce fut le cas dans l'exemple classique de l'évolution et de la domestication du blé, plante annuelle. Ainsi, les espèces de blé résultent d'évènements d'hybridations suivies de polyploïdisation permettant aux hybrides de devenir fertiles. Le premier de ces évènements, entre 500 000 et 150 000 ans avant notre ère, a eu lieu entre *Triticum urartu* et une espèce sauvage proche d'*Aegilops speltoides,* et a permis l'apparition de l'espèce tétraploïde *Triticum turgidum* L. subsp. *diccocoides*. Cette espèce a ensuite subit un processus de domestication produisant l'amidonnier (*Triticum turgidum* subsp. *dicoccum* Schübl.), puis le blé dur (*Triticum turgidum* subsp. *durum* Husn.). Un second évènement de polyploïdisation a eu lieu beaucoup plus récemment (il y a 10 000 ans environ) impliquant un hybride entre l'une des espèces domestiquées tétraploïdes et d'une espèce sauvage

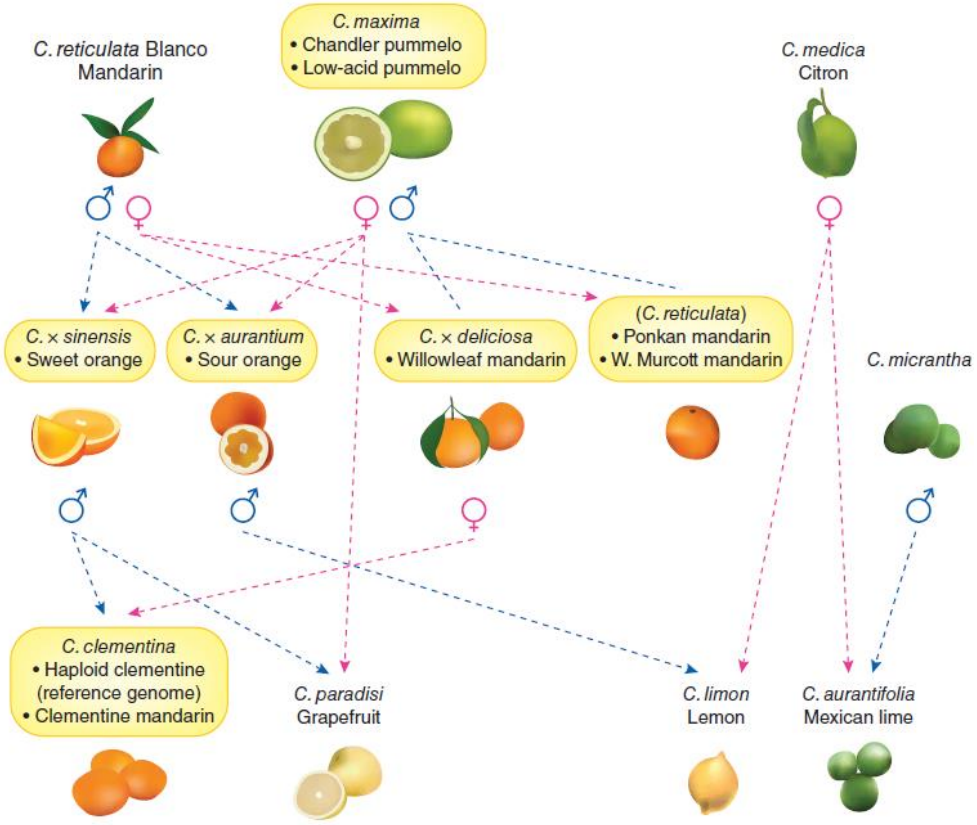


Figure 1.1 : Représentation schématique de l'origine de différentes espèces d'agrumes par hybridations entre espèces. Image tirée de Velasco and Licciardello 2014.

diploïde (*Triticum tauschiito*). C'est ce second évènement qui a permis la création du blé tendre utilisé pour le pain, *Triticum aestivum* L. (synthèse dans Charmet 2011).

Parmi les plantes cultivées pérennes, on peut également trouver des cas où les hybridations ont eu une importance cruciale, comme chez les espèces fruitières et chez les espèces ornementales. Chez les arbres fruitiers, l'un des cas les plus frappants est celui des agrumes pour lesquels une généalogie complexe comprenant de multiples hybridations a été révélée (Wu et al 2014, Fig. 1.1). Ainsi les hybrides entre les mandarines *Citrus reticulata* et les pommelos *Citrus maxima* sont soit des oranges, soit des mandarines selon le sexe des parents. Les citrons verts et jaunes ont tous les deux pour géniteur femelle *Citrus medica* mais le géniteur mâle est l'espèce sauvage *Citrus micrantha* pour les premiers, et *Citrus x auranthium* (le bigaradier, nommée aussi orange amère en anglais) pour les seconds.

Chez les roses, le processus d'obtention des variétés modernes (*Rosa x hybrida*) a compris encore plus d'étapes d'hybridations que dans le cas des agrumes, certaines impliquant des polyploïdisations, d'autres non. Les variétés modernes de roses semblent issues d'un évènement d'hybridations entre l'espèce sauvage européenne, *Rosa foetida* et les variétés plus anciennes dites « Hybrides de thé », elles-mêmes issues de nombreux croisements complexes entre diverses espèces sauvages (Bendahmane et al 2013, Fig. 1.2).

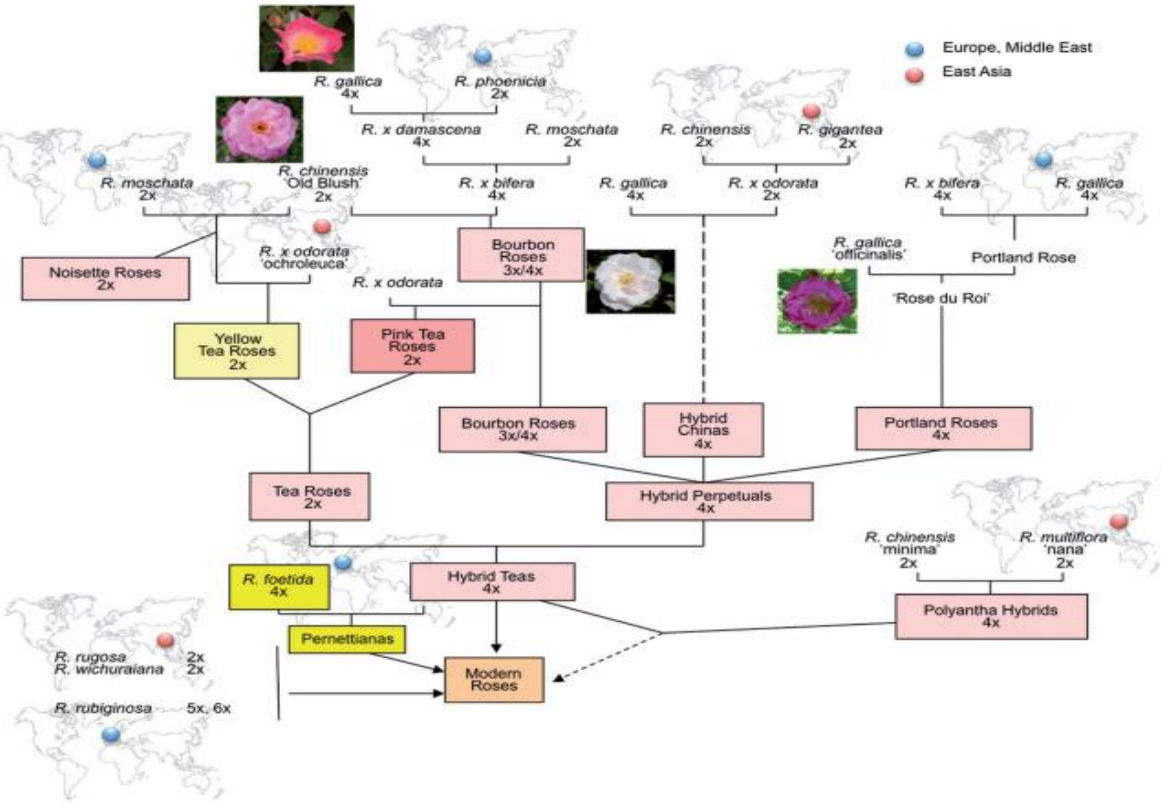


Figure 1.2 : Représentation schématiques de la généalogie des roses cultivées modernes issues d'hybridations entre roses cultivées anciennes et roses sauvages. Image tirée de Bendahmane 2013

Conséquences des flux de gènes et hybridations pour les espèces sauvages

Les hybridations entre espèces cultivées et espèces sauvages peuvent avoir des conséquences non seulement sur les espèces cultivées, mais aussi sur les espèces sauvages. Les conséquences causées par les hybridations entre espèces sauvages et domestiques sur les espèces sauvages ont longtemps été minimisées et donc trop peu étudiées. En effet, beaucoup supposaient que les hybrides entre des espèces cultivées et des espèces sauvages auraient une valeur sélective plus faible que leurs parents sauvages dans le milieu naturel. Dans certains cas, cette hypothèse a été confirmée, par exemple dans le cas des tournesols (Gutierrez, Cantamutto, & Poverene, 2011). Cependant, avec l'accumulation des études sur le sujet, il semblerait qu'une valeur sélective plus faible ne soit pas une règle absolue, ni même nécessairement la majorité des cas. Ainsi, chez le sorgo par exemple, aucune baisse de valeur sélective des hybrides n'a été observée (Arriola & Ellstrand, 1997). Certains hybrides montrent même une valeur sélective supérieure à celle des espèces parentales, comme pour la salade (Hooftman, Oostermeijer, Jacobs, & Den Nijs, 2005) et pour les radis (Hovick, Campbell, Snow, & Whitney, 2012). Les flux de gènes des espèces cultivées vers leurs apparentés sauvages, difficilement contrôlables, sont particulièrement intéressants à étudier dans un monde où des cultures génétiquement modifiées sont cultivées en plein champ. Ainsi, il existe déjà des observations de transgènes trouvés dans des populations sauvages proches de culture de plantes transgéniques, soulevant ainsi des questions légales et éthiques, notamment en ce qui concerne la conservation de l'intégrité génétique des espèces sauvages (Snow, 2012).

Mais les flux de gènes entre espèces sauvages et domestiques posent également des problèmes de conservation en dehors de toute considération de cultures transgéniques. En effet, lorsque les hybrides entre espèces cultivées et sauvages sont retrouvés en milieu naturel, les flux de gènes peuvent aller jusqu'à menacer le maintien des espèces sauvages dans leur propre milieu. La menace d'extinction causée par les hybridations peut résulter de deux phénomènes différents en fonction de la valeur sélective des hybrides. Dans le cas où la valeur sélective des hybrides est faible en milieu naturel, les hybridations ne vont pas résulter en l'implantation d'une nouvelle population hybride dans ce milieu. En revanche, l'abondance du pollen venant des cultures peut mener à une importante proportion des graines portées par l'espèce sauvage qui sera hybride. Puisque ces descendants hybrides ont un taux de survie et/ou de fertilité bas, les effectifs de la population sauvage risquent de diminuer à cause du remplacement des graines de type sauvages par des graines hybrides ne permettant pas le renouvellement de la population. C'est ce que l'on

appelle le « demographic swamping ». Dans le cas où, à l'inverse, la valeur sélective des hybrides est élevée ou même moyenne, mais suffisante pour permettre leur survie dans le milieu naturel, la population sauvage risque peu à peu de perdre son intégrité génétique et d'être remplacée par des génotypes hybrides, ce qui est alors désigné comme « genetic swamping » (Todesco et al., 2016; Wolf, Takebayashi, & Rieseberg, 2001).

Quels que soient les espèces et mécanismes génétiques et évolutifs étudiés, s'intéresser aux flux de gènes des espèces cultivées vers leurs apparentés sauvages peut donc mener à s'intéresser également à la biologie de la conservation. Cette discipline touche aux espèces, communautés et écosystèmes qui sont perturbés directement ou indirectement par des activités humaines ou d'autres causes (Soulé, 1985). Avec une visée appliquée, la biologie de la conservation se concentre sur la protection et la restauration de la biodiversité en utilisant à la fois les connaissances obtenues par la biologie mais aussi par les sciences telles que l'économie, la sociologie ou l'écologie. Elle repose donc sur l'analyse des processus de maintien de la biodiversité pour fournir des informations utiles pour une gestion durable des espèces et écosystèmes.

D'après la Société pour la Biologie de la Conservation (une communauté de professionnels de la conservation), afin de mener cette mission à bien, trois questions doivent obtenir réponse : 1. Comment la biodiversité est-elle distribuée à l'échelle mondiale ? 2. A quelles menaces fait face la biodiversité ? 3. Quelles actions peuvent être entreprises afin de réduire ou d'éliminer ces menaces et de restaurer la diversité biologique ?

Comment la diversité est-elle distribuée à l'échelle mondiale ?

La biologie de la conservation peut s'exercer à plusieurs niveaux, dont les écosystèmes, les communautés, les espèces ou les populations. On a vu en introduction que la catégorie de l'espèce n'est pas une limite nécessairement plus pertinente et facilement délimitable que la population. Malgré cela, le mot d'espèce est très utilisé en conservation où de nombreuses activités sont centrées sur la conservation d'espèces en particulier, probablement parce que cette unité de mesure est plus aisée à utiliser et plus pratique que des notions telles que l'écosystème, les populations ou les allèles. C'est aussi une notion évocatrice pour le public et qui satisfait l'esprit humain qui aime classifier les choses en cases bien séparées. L'UICN (Union Internationale pour la Conservation de la Nature), par exemple, utilise le niveau de l'espèce pour constituer leur « Liste rouge des espèces», un inventaire mondial de l'état de conservation global des espèces végétales et animales (www.uicn.fr).

Et pourtant, concentrer les efforts de conservation au seul niveau de l'espèce pourrait conduire à une perte de la diversité et structure génétiques naturelles des espèces. Afin notamment de prendre en compte cette variation intraspécifique, le concept d'« evolutionary significant unit » est utilisé en conservation, correspondant à ce qui est fréquemment désigné avec le terme de populations (Newton, Allnutt, Gillies, Lowe, & Ennos, 1999). Ces unités peuvent désigner des populations isolées géographiquement les unes des autres, mais elles peuvent très souvent aussi être identifiées avec des données d'ADN et via des analyses phylogéographiques. La création de ces populations peut provenir d'évènements démographiques et historiques. En particulier, de nombreuses espèces ont vu leur structure génétique sculptée par les oscillations climatiques du Quaternaire, notamment en Europe. La dernière grande glaciation (120 000 – 10 000 ans) a vu l'Europe du Nord se couvrir de glace et les espèces tempérées ont été chassées vers le sud, dans les péninsules ibérique, italienne et balkanique. A la fin de l'ère glaciaire, un phénomène inverse s'est produit et les espèces tempérées ont pu coloniser de nouveau les territoires redevenus accueillants pour elles vers le nord. Des populations s'étant différenciées génétiquement les unes des autres lors de leur isolement dans les refuges glaciaires ont été identifiées pour de nombreuses espèces tempérées après leur recolonisation de l'Europe (Hewitt, 1999; Stewart, Lister, Barnes, & Dalén, 2010). Il semble également qu'il existe pour certaines espèces des refuges dits « cryptiques » plus au Nord (Provan & Bennett, 2008). Les mouvements de contraction-expansion suivant les changements climatiques ont non seulement affecté la structure des populations des espèces tempérées mais aussi leur diversité génétique, avec d'une part des évènements de fondation successifs abaissant la diversité graduellement en s'éloignant des refuges (Hewitt, 1999), et d'autre part des zones de contact entre les différentes populations, augmentant la diversité localement (Provan & Bennett, 2008).

Dans le cadre de la biologie de la conservation, il est important de noter cependant que la variation génétique obtenue par ces analyses phylogéographiques est bien souvent neutre. Lorsque cela est possible, il est idéal de travailler plutôt avec la structure et diversité génétique adaptative afin d'identifier des « management units », par exemple liées à une adaptation locale des populations, plutôt que de travailler uniquement avec les « evolutionary significant units » (Newton et al., 1999).

À quelles menaces fait face la diversité ?

Le futur de nombreuses espèces et populations est intimement lié à l'usage des territoires de leur aire de répartition par les humains. En dehors des villes et agglomérations, les territoires ruraux sont des mosaïques d'espaces utilisés de façon variée : vergers et champs sont

entrecoupés d'habitations, de routes et de cours d'eau. On y trouve aussi des milieux naturels et semi-naturels que sont les bois et forêts ainsi que certains espaces de végétation éparse le long des cours d'eaux et entre les parcelles agricoles sous forme de haies. Bien que tous les paysages soient dynamiques, cela est particulièrement vrai des environnements agricoles en raison de la manipulation par les humains des différents éléments de ce paysage afin, notamment, d'augmenter la productivité agricole (Bennett, Radford, & Haslem, 2006). L'une des conséquences de cette recherche d'augmentation de productivité depuis quelques décennies est la disparition des bocages via la suppression des haies, notamment, mais aussi le comblement des points d'eau. Ces modifications anthropiques des paysages peuvent bouleverser les écosystèmes et affecter directement la survie des espèces, conduisant à diverses initiatives associatives et citoyennes de protection de ces milieux et de la richesse de leur faune et de leur flore. Parmi ces très nombreuses initiatives, on peut citer l'exemple de « si les mares m'étaient comptées » un inventaire participatif des mares d'Ile-de-France lancé par la société nationale de protection de la nature et par l'agence régionale Natureparif.

Ces modifications anthropiques des paysages ont également des effets plus indirects sur la conservation des espèces et populations naturelles puisqu'elles peuvent également entraîner des changements dans les flux de gènes entre espèces et entre populations. Par exemple, l'altération des paysages ruraux est susceptible de faciliter les contacts entre environnement agricoles et environnements naturels. Elle peut également toucher la distribution et l'abondance locale des espèces cultivées et de leurs apparentés sauvages. Tous ces changements peuvent parfois faciliter les flux de gènes entre espèces cultivées et espèces sauvages. De plus, la fragmentation de l'habitat naturel de certaines espèces limite bien souvent les échanges possibles entre populations sauvages devenues distantes les unes des autres. L'influence des activités humaines sur les flux de gènes peut également s'exercer de façon indirecte en affectant l'abondance ou la nature des pollinisateurs (Kremen et al., 2007). Chez certaines espèces, les flux de gènes sont encore favorisés par leurs traits d'histoire de vie comme des barrières interspécifiques faibles et des systèmes de reproduction auto-incompatibles qui rendent obligatoires les croisements entre individus éloignés génétiquement.

Or, comme nous l'avons vu précédemment, les flux de gènes entres espèces sauvages et leurs apparentés cultivés peuvent avoir des conséquences graves du point de vue de leur conservation. Il est donc important d'étudier la quantité des flux de gènes et les facteurs les facilitant ou les limitant, mais il est tout aussi crucial de s'intéresser aux conséquences de ces flux de gènes sur les populations naturelles. Une fois que les risques sont identifiés et quantifiés,

il peut alors être nécessaire de mettre en place des actions de conservation en relation avec les décideurs.

Quelles actions peuvent être entreprises afin de réduire ces menaces ?

Différentes techniques de conservation peuvent être utilisées pour contrer les menaces sur les espèces végétales. Les approches dites « *in situ* » ou sur site tentent de conserver les populations naturelles dans leur milieu d'origine. Ce mode de conservation dynamique permet aux populations d'être protégées tout en continuant d'évoluer le plus naturellement possible et de maintenir une certaine diversité (Koskela et al., 2013). Cela permet aussi de maintenir l'écosystème auquel participe l'espèce et dans lequel elle forme potentiellement des relations d'interdépendance avec d'autres espèces. Dans le cadre d'une conservation in situ, il peut arriver de réintroduire artificiellement des individus dans des milieux qui en était devenus dépourvus afin, notamment, de créer des ponts entre populations isolées. Ces approches, lorsqu'elles sont possibles, sont généralement préférables à tout autre type d'approches, puisqu'elles permettent de maintenir sur le long terme le potentiel d'adaptation de l'espèce menacée.

Cependant, dans certains cas, les approches *in situ* sont difficiles à mettre en place. Cela peut être le cas, par exemple, lorsque l'habitat naturel de l'espèce est lui-même en voie de disparition à cause de changements anthropiques ou de réchauffement climatique. Il peut alors être nécessaire de faire appel à des méthodes de conservation dites « *ex situ* », définies par la Convention sur la Diversité Biologique des Nations Unies comme la conservation d'éléments constitutifs de la diversité biologique en dehors de leur milieu naturel ce qui a des conséquences plus ou moins sévères pour l'espèce qui n'évolue plus dans son écosystème naturel. Ces méthodes peuvent être dynamiques ou statiques selon qu'une forme de brassage génétique soit permise ou non (Balsemin & Collin, 2004). Les méthodes de conservation *ex situ* dites dynamiques sont par exemple la création de vergers conservatoire pour les espèces d'arbres ou d'élevage conservatoire pour certains animaux. A condition que le verger ou l'élevage contiennent suffisamment de diversité, les nouvelles générations issues de ces programmes pourront être réintroduites dans des environnements naturels où la sélection pourra alors s'exercer de nouveau. Les approches de conservation *ex situ* statiques se bornent, elles, à la conservation « à l'identique » de la population ou des individus menacés. Cela peut aller de méthodes « traditionnelles », comme la constitution de jardins botaniques et de parcs animaliers à des méthodes plus récentes comme la création de la Réserve mondiale de semences du Svalbard (stockant des graines de variétés cultivées).

II. Introduction aux espèces étudiées : *Malus* sp.

Bien que les conséquences des hybridations entre espèces cultivées et sauvages aient été bien étudiées dans le cas de cultures annuelles comme dans le cas du blé (Arrigo et al., 2011) et du riz (Pusadee, Schaal, Rerkasem, & Jamjod, 2013), elles sont encore mal connues chez les espèces de plantes pérennes, comme les arbres fruitiers. Et pourtant les espèces d'arbres en général ont de fortes capacités de dispersion comparées aux espèces annuelles et des barrières interspécifiques faibles (Petit & Hampe, 2006), deux traits d'histoire de vie qui favorisent les hybridations. Ils sont donc des modèles intéressants pour étudier ces questions.

Importance culturelle et économique de la pomme

Le second arbre fruitier le plus cultivé dans le monde (après la banane) et le premier dans les régions tempérées est le pommier domestiqué, *Malus domestica* Borkh. (syn. *Pyrus malus* L., *Malus pumila* Mill.). La production mondiale de pommes a été estimée en 2013 (FAOSTAT) à plus de 80 millions de tonnes. Les premiers pays producteurs sont la Chine (qui représente environ la moitié de la production mondiale à elle seule), les Etats-Unis, la Turquie et la Pologne. La France, elle, arrive en septième position avec une production de 1,7 millions de tonnes. Les utilisations des pommiers sont principalement centrées sur le fruit, bien que le bois des pommiers, apprécié pour sa forte densité et son homogénéité, soit parfois utilisé en ébénisterie, comme le bois de poirier. Le fruit est utilisé pour être consommé frais ou cuit et son jus est utilisé notamment pour faire du cidre. L'un des intérêts principaux de ce fruit par rapport à d'autres fruits comme les pêches, les abricots ou les poires est sa grande capacité de stockage. Un autre atout est que les pommes cultivées sont très diverses. Il en existe plus de 6000 variétés qui peuvent varier en taille, en couleur, en goût, en texture… (Hancock, Luby, Brown, & Lobos, 2008). De nombreuses localités (pays, régions) ont ainsi des variétés qui y sont considérés comme

traditionnelles et locales comme la Reinette du Mans en Sarthe ou la de l'Estre dans le Limousin. Malgré cette très grande diversité, la production actuelle est dominée par un nombre très limité de variétés qui sont connues internationalement comme les pommes « Golden Delicious » ou « Fuji » (Sansavini, Donati, Costa, & Tartarini, 2004). Cependant, le marché est variable sur le long terme et des effets de mode accentués par du marketing permettent à de nouvelles variétés d'émerger comme l'« Ariane » (créée par l'INRA), la « Cripps Pink » (souvent vendue comme Pink Lady®) ou encore la « Delblush » (vendue en tant que Tentation®, Hancock et al 2008).

Maintenir ces variétés, qu'elles soient traditionnelles et locales ou mondialisées, repose sur la greffe. En effet, les pommiers sont en général auto-incompatibles et doivent être pollinisés par un génotype différent. La descendance d'un pommier ne possèdera donc pas le même phénotype que l'arbre mère et propager les caractères phénotypiques propres à chaque variété implique donc de propager le génotype de cette variété. La majorité des variétés de pommiers cultivés (à l'exception de certaines variétés à petits fruits d'Asie occidentale) ne peuvent pas non plus être reproduits à partir de boutures ou de surgeons, c'est-à-dire les rejetons sortant du pied d'un arbre et c'est donc bien par greffe que les variétés sont reproduites (Zohary & Hopf, 2000).

En plus de leur importance économique, les pommes ont également une grande importance culturelle, notamment en Europe. L'un des symboles les plus forts de la place cruciale de ce fruit dans la culture européenne se trouve dans l'interprétation qui a été faite de la Genèse de la Bible. Ainsi, dans le mythe d'Adam et Eve, les deux humains mangent un fruit défendu, ce qui cause la perte de leur immortalité et leur exil de l'Eden. Les textes latins ne précisent pas de quel fruit il s'agit précisément (« pomum »), mais ce fruit a généralement été interprété comme étant une pomme en Europe Occidentale (d'où, notamment, l'expression la pomme d'Adam pour désigner le larynx). Cette interprétation pourrait provenir d'un jeu de mot sur le terme latin « malus » qui désigne à la fois le pommier mais signifie également « mauvais ». Les pommes sont également omniprésentes dans les mythes gréco-romains (comme, par exemple, le Jugement de Paris, le Jardin des Hespérides ou Atalante), les contes (Blanche- Neige) ou les légendes (l'île légendaire du roi Arthur, Avalon, semble être nommé d'après le mot galois pour pomme ou pommier). De façon plus contemporaine, l'importance culturelle des pommes en France est reflétée dans la création d'associations centrées autour de la sauvegarde de variétés anciennes et traditionnelles (comme les Croqueurs de Pommes qui ont des antennes à travers toute la France).

Les hybridations inter-espèces au sein du genre Malus

Le pommier cultivé appartient à la famille des Rosaceae, qui est l'une des familles de plantes les plus diverses et les plus importantes économiquement. En effet, les plus de 90 genres inclus dans la famille des Rosaceae regroupent environ 3000 espèces dont de nombreuses espèces fruitières des régions tempérées comme les fraises, les pêches, les framboises, les abricots, les poires, et, bien sûr, les pommes.

Le genre *Malus* comprend environ 30 espèces et est réparti dans les régions tempérées d'Europe, d'Asie et d'Amérique du Nord (Zohary & Hopf, 2000). Plusieurs de ces espèces sauvages sont endémiques de régions où la culture des pommes est très répandue. En Europe, l'espèce endémique de pommier sauvage est *Malus sylvestris* Mill. Dans le Caucase, en Iran et en Anatolie, l'espèce est *M. orientalis* Uglitzkich (parfois considérée comme une sous-espèce de *Malus sylvestris* sous le nom de *M. sylvestris* subsp. *orientalis* Browicz). L'Asie abrite, elle, plusieurs groupes de pommiers sauvages dont *M. sieversii* Roem en Asie Centrale (parfois découpée en espèces locales comme *M. kirghizorum* et *M. turkmenorum*) et *M. prunifolia* en Sibérie orientale et dans le Nord de la Chine (Zohary & Hopf, 2000). Le flou entourant la taxonomie du genre *Malus* provient d'un manque de caractères morphologiques propres à chaque espèce, du chevauchement des aires de répartition des espèces sauvages comme des espèces cultivées et ornementales ainsi que des faibles barrières aux flux de gènes interspécifiques dans le genre *Malus* (Cornille et al 2014, Juniper and Mabberley 2006).

Histoire du pommier cultivé

Dans les années 1920, le scientifique russe Vavilov a exploré les forêts d'Asie Centrale et y a remarqué des zones de forêt de *M. sieversii* avec une très grande diversité de phénotypes. Il a également observé une remarquable ressemblance morphologique entre *M. sieversii* et *M. domestica*. Vavilov a donc proposé cette région du monde, incluant l'ouest de la Chine, le Kazakhstan, le Kirghizstan et le Tadjikistan comme centre d'origine du pommier cultivé (Harris, Robinson, & Juniper, 2002; Vavilov, 1926). Il est clair que le pommier *M. domestica* a ensuite été importé vers l'Europe, cependant il est difficle d'identifier le moment exact de son introduction dans ces régions. En effet, les fruits des espèces de pommiers sauvages endémiques d'Europe étaient certainement consommés durant le Néolithique et l'Age de Bronze. Il est donc possible que les « pommes » des premières traces écrites (issus de la litérature sumérienne) fassent plutôt référence à ces pommes sauvages, ce que semble confirmer la présence de pommes de petit

diamètre dans le complexe mortuaire de la reine Pu-abi (milieu du troisième millénaire avant J.C. ; Janick 2005, Miller 2013). Ce qui est sûr est qu'au cours des premiers siècles avant J.C., la culture du pommier *M. domestica* était répandue et déjà très perfectionnée au moins dans le monde greco-romains puisque Théophraste (-371 à -288) fait mention des variétés différentes de pommes en les distinguant des pommes sauvages et évoque la nécessité de la greffe pour les reproduire et que le romain Varro (-116 à -27) décrit la construction d'édifice de stockage spécifiques aux pommes (Hancock et al., 2008; Jackson, 2003).

Il est très probable que les premières pommes cultivées aient été transportées le long des grandes routes de commerces, dites routes de la Soie, reliant la Chine à l'Europe de l'Ouest et passant à travers l'ère de distribution de *M. sieversii* (Harris, Robinson, & Juniper, 2002). L'importation du pommier cultivé vers l'Ouest de l'Eurasie l'a introduit dans les aires de répartition d'autres pommiers sauvages avec lesquelles il est fertile, comme *M. orientalis* (dans le Caucase, en Turquie et en Iran) ou *M. sylvestris* (en Europe occidentale). L'existence d'hybridations et d'introgressions de ces pommiers sauvages vers les pommiers cultivé a été suggérée.

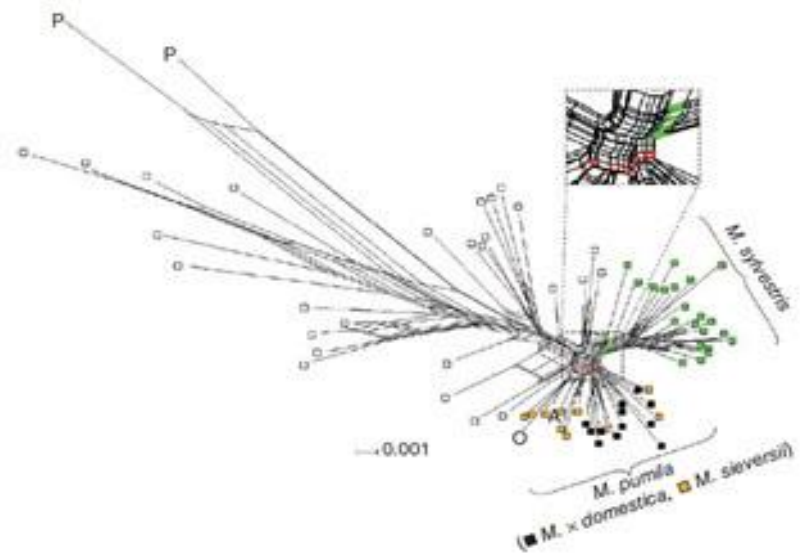


Figure 1.3 : Relations phylogénétiques entre espèces de pommiers incluant des variétés de Malus domestica *basées sur la séquence concaténée de 23 locus. En noir,* M. domestica *; en orange,* M. sieversii *et vert,* M. sylvestris*. La lettre P désigne les individus du genre* Pyrus*. Figure tirée de Velasco et al. 2010.*

Des données génétiques semblent confirmer cette histoire. Dans leur article rapportant le séquençage du premier génome de *M. domestica*, Velasco et al (2010) ont analysé 23 gènes nucléaires dans 12 *M. domestica*, 10 *M. sierversii*, 21 *M. sylvestris* et d'autres espèces de pommiers (comme *M. orientalis* ou *Malus floribunda*) dans le but d'identifier l'espèce de pommier sauvage la plus proche du pommier cultivé. L'arbre réticulé issu de cette analyse

montre les pommiers cultivés dans un même groupe avec *M. sieversii* et clairement séparés de *M. sylvestris* (Fig. 1.3). Les auteurs concluent que cette analyse résout le débat de l'origine de *M. domestica* en faveur de *M. sieversii* comme ancêtre unique et propose même que les deux groupes forment une espèce unique qu'il serait plus correct de nommer *M. pumila* Mill.

Cependant, une ré-analyse de ces mêmes données par Harrison and Harrison (2011) met en doute cette conclusion, indiquant que l'analyse phylogénétique des gènes pris indépendamment les uns des autres montre la possibilité de flux de gènes entre *M. domestica* et *M. sylvestris*. Une phylogénie basée sur les chloroplastes montre même une relation inverse avec les chloroplastes de nombreuses variétés (comme la Golden Delicious, la Gala et la Cox's Orange Pippin) se regroupant avec ceux de *M. sylvestris* alors que seule la Granny Smith forme un groupe avec *M. sieversii* (Nikiforova, Cavalieri, Velasco, & Goremykin, 2013). Ce résultat avait été préfiguré dans une comparaison d'haplotypes chloroplastiques antérieure (Coart, Van Glabeke, De Loose, Larsen, & Roldan-Ruiz, 2006). Des analyses impliquant des microsatellites nucléaires et des comparaisons de modèles de divergence indiquent que ces trois espèces forment des groupes génétiques distincts, avec une divergence initiale de *M. domestica* depuis *M. sieversii*, et des introgressions secondaires fréquentes de *M. sylvestris* dans le pool de gènes de *M. domestica*, au point que les individus cultivés apparaissent actuellement plus proches de ceux de l'espèce *M. sylvestris* que de ceux de l'espèce *M. sieversii* (Cornille et al 2012).

III. Articles

Article 1. Forces motrices anthropogéniques et naturelles des flux de gènes chez un fruitier sauvage des zones tempérées : une base pour la conservation et les programmes d'amélioration variétale chez les pommes

À l'origine du projet présenté dans ce premier article se trouvait une étude réalisée par Cornille et al. (2013b) montrant de forts taux d'introgressions de *M. domestica* dans les populations du pommier sauvage européen. Ce résultat, contrastant avec des taux plus faibles trouvés auparavant (Coart et al., 2003), soulevait la possibilité que les hybrides aient une meilleure valeur sélective que supposé précédemment et donc que l'intégrité génétique du pommier sauvage européen soit davantage en danger qu'envisagé jusque-là. En effet, les inquiétudes concernant cette espèce ont augmenté récemment en raison des hybridations avec le pommier cultivé mais aussi de la baisse de sa taille de population, particulièrement durant les vingt dernières années (Schnitzler, Arnold, Cornille, Bachmann, & Schnitzler, 2014). Le statut conservatoire de *M. sylvestris* n'était pas déterminé au début de ma thèse : bien que l'espèce soit considérée comme étant en danger dans certains pays d'Europe (comme la Belgique, Coart et al. 2003), la Liste Rouge de l'UICN la classe comme « data deficient » en raison des conséquences inconnues des hybridations avec *M. domestica*.

Nous avons tenté d'évaluer plus précisément l'intensité des hybridations avec *M. domestica* sur *M. sylvestris* sur l'ensemble de son aire de répartition (Fig. 1.4). Pour cela, un échantillonnage sans précédent de 1889 individus répartis dans 15 pays d'Europe a été effectué.

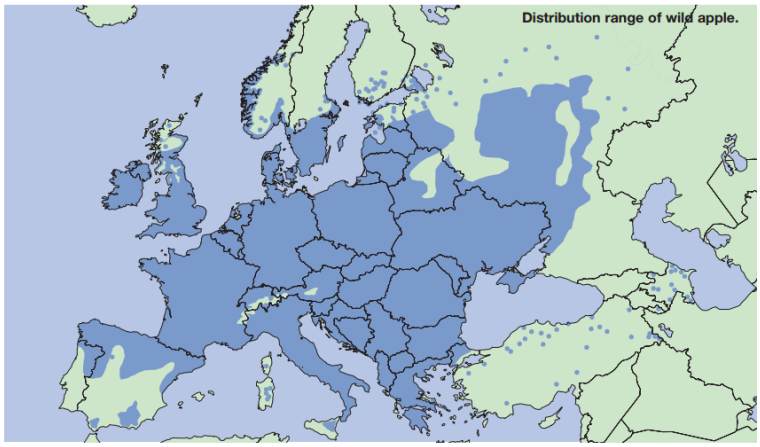


Figure 1.4 : Aire de repartition de Malus sylvestris. Illustration tirée de Stephan et al 2003.

Comme évoqué précédemment, la première étape de toute étude de conservation passe par l'évaluation des niveaux et de la distribution de la diversité pour l'espèce concernée. Nous avons donc essayé de comprendre la façon dont la diversité génétique de *M. sylvestris* est distribuée à l'échelle européenne. L'analyse de 26 marqueurs microsatellites nous a permis de compléter les résultats de Cornille et al. (2013a) en étudiant la répartition de la diversité génétique de l'espèce à travers l'Europe, ainsi que sa structure phylogéographique, afin d'identifier d'éventuelles populations différenciées, potentiellement liées à des changements climatiques lors de la dernière glaciation. Ce type d'analyse a révélé des schémas differents entre espèces et a été effectuée pour de nombreuses espèces d'abres comme pour le chêne ou le pin sylvestre (Cheddadi et al., 2006; Petit et al., 2002).

Après avoir étudié la diversité et la structure génétique de l'espèce sur son aire de répartition, nous avons également utilisé cet échantillonnage pour estimer les degrés d'introgressions de *M. domestica* dans les populations sauvages de pommiers en Europe et leur répartition à l'échelle européenne, et plus localement à l'échelle française. Nous avons également évalué l'influence indirecte des activités anthropiques sur les flux de gènes entre les deux espèces via les pollinisateurs.

Ma contribution à cet article a été de participer à l'extraction d'ADN et à l'amplification des marqueurs microsatellites, puis de constituer les génotypes. Parmi les analyses, j'ai réalisé celles permettant d'identifier les taux d'introgressions et d'identifier les représentants « purs » de chaque espèce. J'ai également étudié la diversité génétique chez *M. sylvestris* et sa structure de population, ainsi qu'estimé les flux de gènes historiques et l'autocorrélation spatiale. Cet article a été publié dans Evolutionary Applications en 2015.

Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: a basis for conservation and breeding programs in apples

Cornille Amandine*, Feurtey Alice, Gélin Uriel, Ropars Jeanne, Misvanderbrugge Kristine, Gladieux Pierre and Giraud Tatiana*

*co-directed the work

Introduction

Over the last century, researchers in evolutionary and applied biology have paid much attention to gene flow, which is considered to play a key role in species persistence, evolution, and diversification. Many studies have investigated the consequences of intra- and interspecies gene flow for evolution (e.g., Garant et al. 2007; Feder et al. 2012). However, the anthropogenic and natural determinants of gene flow remain largely unexplored (Manel et al. 2003; Manel and Holderegger 2013) because multiple data sources (genetic, environment, ecology) are required to unravel these issues. One central question concerns the extent to which climate changes and recent global changes due to human disturbances (e.g., land-use management and fragmentation) affect gene flow (Manel and Holderegger 2013). This question must be addressed, to make it possible to forecast the impact of these changes on the evolution of populations and to develop appropriate management, breeding, and conservation programs.

Within-species gene flow, which is mediated by pollen and seed dispersal, is a key factor underlying population persistence (Ellstrand 2014). Gene flow shapes the spatial genetic structure and can therefore be inferred from the amount of genetic differentiation between populations by the use of neutral marker loci (Vekemans and Hardy 2004). Over the last 20 years, an array of methods, including autocorrelation, parentage analyses, and assignment methods, has been developed for the identification of natural factors affecting gene flow (Pritchard et al. 2000; Manel et al. 2003; Smouse and Sork 2004; Storfer et al. 2010). Studies based on these methods have identified climate as a key driver of within-species historical gene flow in many temperate species in Europe during the Pleistocene (Hewitt 2004). In particular, climate-driven range fluctuations during the Pleistocene have continuously reshaped genetic structure within species. However, the impact of climate on gene flow remains largely unexplored in scattered species, because of the difficulties of large-scale sampling. Yet, the determination of genetic

diversity, differentiation, and admixture across the geographical range of such scattered species can be a practical basis for sustainable management and conservation programs.

Human activities also affect gene flow. In particular, the anthropization of landscapes and the cultivation of crops over extended areas have increased the opportunities for contact between cultivated plants and their wild relatives. Over the last 20 years, crop-to-wild gene flow has been considered a human disturbance and has received close attention, particularly in the context of the cultivation of transgenic crops and the ecological consequences of such gene flow for the fitness and genetic integrity of wild relatives (Ellstrand et al. 1999; Ellstrand 2003). Crop-to-wild gene flow has been investigated using neutral markers and shown to be widespread (Ellstrand et al. 1999; Coart et al. 2003; Delplancke et al. 2011; Aerts et al. 2013). However, only a few studies (Arnaud et al. 2003) have investigated the human disturbances and anthropogenic landscape drivers affecting crop-to-wild gene flow, despite the potentially large number of such drivers at work. Local human food crop production can affect crop-to-wild gene flow directly, by modifying the distribution and abundance of the crop. Humans can also have indirect effects on introgression rates, through their land-use practices. For instance, clear evidence has been obtained of a decline in the populations of both wild and domesticated pollinators associated with the intensification of cultivation and landscape management (Kremen et al. 2007). Pollinators are key vectors of gene flow and allele distribution in landscapes. Land-use changes can therefore have also an indirect impact on both crop-to-wild and within wild gene flow, through anthropogenic effects on the distribution and movements of the pollinator in the landscape.

Fruit trees, particularly scattered trees, are key species in ecosystems (Manning et al. 2006), and fruit tree crops are essential to human society, due to their contribution to worldwide food production (e.g., apple, grape, orange, coffee, cocoa, olive). However, despite the ecological and economic importance of these trees, much uncertainty remains concerning the impact of recent global changes on gene flow from crop-to-wild species and within wild species, particularly in temperate fruit trees. The cultivated apple (*Malus domestica*) is the one of the most important fruit tree crop in terms of worldwide production (http://faostat.fao.org/). It was introduced into Europe 1500 years ago and has been subject to the introgression of genes from the local wild European apple, *Malus sylvestris,* to the extent that *M. domestica* is now more closely related to *M. sylvestris* than to its primary progenitor in the Tian Shan, *M. sieversii* (Cornille et al. 2012). Reciprocally, crop-to-wild introgressions have been reported in the European wild apple *M. sylvestris* (Coart et al. 2006; Larsen et al. 2006; Gross et al. 2012; Cornille et al. 2013b). Investigations of the demographic history of the European wild apple have

revealed the existence of three main genetic groups shaped by past climate change during the Pleistocene (Cornille et al. 2013a). However, both a sampling gap in the distribution of *M. sylvestris* in France and a lack of information about the pollination ecology of this species have limited our understanding of the structure of the population and the identification of drivers of intra- and interspecies gene flow. Yet, cultivated and wild apples in Europe, with their high levels of gene flow from crop-to-wild species and within wild species, easily detectable due to relatively recent secondary contact, constitute an ideal model for investigations of the anthropogenic and natural drivers of gene flow in fruit trees at large geographic scales.

We report here an unprecedented dense sampling of *M. sylvestris* across Europe (1889 individuals), with a focus in France (1092 individuals), unique for a temperate fruit tree. We also sampled a large number of *M. domestica* varieties (339 varieties), and combined genetic (microsatellites), ecological, and environmental data, to investigate the natural and anthropogenic drivers of gene flow in apples. Using genetic data across Europe, we estimated the degree of crop-to-wild introgressions into the European wild apple, across its geographic distribution and investigated the impact of climate on population structure and diversity, through estimates of historical gene flow within the nonintrogressed wild apple gene pool. Furthermore, by estimating recent crop-to-wild introgressions in France, and using environmental and ecological data, we investigated the effects of apple production on these crop-to-wild introgressions. Finally, we determined whether human activities had indirectly impacted rates of gene flow through their effects on the flower visitor community.

Altogether, we highlighted the impact of previous natural climate change on historical gene flow across Europe in the wild apple and identify human activities as key drivers of gene flow in a wild temperate fruit tree. Our landscape genetics study provides useful information for *in situ* and *ex situ* conservation management for this emblematic fruit tree species in Europe, for agroforestry programs aiming at integrating wild apples in agrosystems and for future breeding programs, the European wild apple being a major contributor to the cultivated apple genome.

Materials and methods

Sampling, DNA extraction, and microsatellite genotyping

In addition to the 599 samples of *M. sylvestris* from Europe previously studied (Cornille et al. 2013a,b) (38 sites), we collected a further denser sample of 1290 *M. sylvestris* trees (28

sites), mostly in France (20 sites, $N = 1092$), but also in Denmark (6 sites, $N = 162$), Italy (1 site, $N = 5$) and Belgium (1 site, $N = 31$). In total, 1889 *M. sylvestris* trees from 66 sites were analyzed for this study. In addition to the 299 *M. domestica* previously analyzed (Cornille et al. 2012), we sampled additional *M. domestica* trees in the field in Denmark ($N = 24$) and in French apple orchards ($N = 16$). In total, we analyzed 339 *M. domestica* trees in this study. The dataset for this study thus corresponds to 2228 individuals (details of the samples are provided in Table S1 and Fig. S1).

DNA was extracted with the NucleoSpin® plant II DNA extraction kit (Macherey & Nagel, Düren, Germany). Multiplex microsatellite PCR amplifications were performed with a multiplex PCR kit (Qiagen Inc.) as previously described (Cornille et al. 2012; Patocchi et al. 2009), with 26 microsatellite markers. We retained only multilocus genotypes for which <20% of the data were missing. The suitability for population genetic analysis of the markers used was demonstrated in previous studies (Cornille et al. 2012, 2013a,b).

Crop-to-wild gene flow, population structure, and historical gene flow (large-scale analyses, Europe)

Bayesian inference of crop-to-wild gene flow, population structure, and admixture

We used the individual-based Bayesian clustering methods implemented in STRUCTURE 2.3.3 (Pritchard et al. 2000) to estimate recent introgression from *M. domestica* into *M. sylvestris*. STRUCTURE uses Markov chain Monte Carlo (MCMC) simulations to infer the proportion of ancestry of genotypes from K distinct clusters. The underlying algorithms attempt to minimize deviations from Hardy–Weinberg within clusters and linkage disequilibrium among loci. A lack of consideration of intraspecies genetic structure in STRUCTURE analyses can bias the interpretation of introgression rates (Kalinowski 2011; Cornille et al. 2013b). We ran STRUCTURE, assuming $K = 1$ to $K = 9$ clusters, on the whole dataset including *M. sylvestris* ($N = 1889$) and *M. domestica* ($N = 339$). Based on 10 repeated runs of MCMC sampling comprising 500 000 iterations after a burn-in of 50 000 steps, we determined the amount of additional information explained by increasing K, using the ΔK statistic (Evanno et al. 2005). As a genetic structure was observed within the *M. sylvestris* dataset (see below), we assumed that the total membership coefficient of *M. sylvestris* individuals in the *M. sylvestris* gene pool was the sum of their membership coefficients in each *M. sylvestris* cluster. All *M. sylvestris* individuals with a cumulative membership coefficient for the *M. sylvestris* gene pool >0.9 were considered to be pure *M. sylvestris* trees. Conversely, all

M. sylvestris individuals with a cumulative membership coefficient >0.1 in the *M. domestica* gene pool were considered to be admixed with *M. domestica*. *Malus sylvestris* individuals with a cumulative membership coefficient <0.1 in the *M. sylvestris* gene pool were considered to be misidentified individuals actually belonging to *M. domestica*. The proportion admixture of the *M. domestica* gene pool with the *M. sylvestris* gene pool is hereafter denoted $P_{DOM \rightarrow SYL}$ and referred as 'introgression rate'. In the absence of genomewide genotyping data, we probably failed to identify advanced generation backcrosses. However, our approach aimed to identify early generations of admixture and introgression, because our focus here was the effect of recent human disturbance on gene flow.

To investigate the population structure and admixture of *M. sylvestris*, we excluded hybrids previously identified as displaying introgression from *M. domestica* (i.e., individuals assigned to a wild gene pool with a membership coefficient <0.9; N = 1376 individuals, 62 sites) and used the individual-based spatially explicit Bayesian clustering method implemented in TESS 2.3.1 (Chen et al. 2007). TESS is similar to STRUCTURE, except that it incorporates a spatial component into the clustering procedure, such that genotypes from areas located close together geographically are considered more likely to belong to the same cluster. We used the conditional autoregressive (CAR) Gaussian model of admixture with linear trend surface, setting the spatial interaction parameter (ρ) to 0.6. These parameters (ρ and trend) affect the weight given to spatial distance when clustering genotypes. Ten independent analyses were carried out for each number of clusters K ($2 \leq K_{max} \leq 9$), using 500 000 MCMC iterations after a burn-in of 50 000 steps. We determined the amount of additional information explained by increasing K, using the rate of change of the deviation index criterion (DIC) with increasing K.

For both TESS and STRUCTURE analyses, outputs were processed with CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007), to identify distinct clustering solutions in the replicated runs for each K value.

Genetic variation within *M. sylvestris*

For the analysis of within-species genetic variation, we excluded hybrids identified as displaying introgression from *M. domestica*. Descriptive statistical analyses were performed for each site (i.e., geographical location) and each population (i.e., clusters inferred by TESS, including hybrids with a membership coefficient of up to 0.55 for the focal cluster).

Allelic richness (A_R) and private allelic richness (A_P) were calculated with ADZE (Szpiech et al. 2008), using standardized sample sizes of N = 2 (one individual x two chromosomes) and N = 100 (50 individuals × two chromosomes), respectively, corresponding to

the minimal number of observations across sites and populations. Heterozygosities (expected and observed), Weir and Cockerham *F*-statistics and deviations from Hardy–Weinberg equilibrium were calculated with GENEPOP 4.2 (Raymond and Rousset 1995; Rousset 2008). Only sampling sites for which at least four individuals were successfully genotyped were included in site-specific calculations (51 of the 62 sites). Spatial patterns of genetic variability were visualized by mapping variation (A_R), for 43 sites (i.e., geographical locations with at least four individuals and for which at least two individuals were successfully genotyped for each marker) and genetic diversity (H_O) for 51 sites across space, with the kriging and surface functions (*field* R package).

Spatial autocorrelation and historical gene flow

We estimated spatial genetic structure within the five populations previously identified with TESS. Analyses were performed with SPAGeDI 1.3 (Hardy & Vekemans 2002) for each population, using the estimator of kinship coefficient of Nason (Loiselle et al. 1995). Kinship coefficient values (F_{ij}) were regressed against the natural logarithm, $\ln(d_{ij})$, to obtain the regression slope, b. The spatial positions of individuals were permuted 9999 times to test for the existence of a spatial genetic structure. We compared the extent of spatial genetic structure between sites over the same spatial scale, by calculating b_{Ld}, the regression slope of F_{ij} against $\ln(d_{ij})$ for d_{ij} = 10 km (the maximum distance providing the best representation of pairs of individuals for each population). We then calculated the Sp statistic, defined as $Sp = -b_{(Ld10)}/(1-F_N)$, where F_N is the mean F_{ij} between neighboring individuals, which was approximated by $F_{(d)}$ for the first distance interval (Vekemans and Hardy 2004; Hardy et al. 2006).

Factors affecting crop-to-wild gene flow and the diversity of flower visitors (fine-scale analyses, France)

All statistical analyses were performed with R version 2.15.3 (R Foundation for Statistical Computing, Vienna, Austria).

Anthropogenic factors affecting crop-to-wild introgression

We expected the density of cultivated apple trees to be the principal determinant of recent crop-to-wild introgression rate detected by STRUCTURE (i.e., $P_{DOM \rightarrow SYL}$). We first investigated the effect of the density of cultivated apple trees (based on apple production in

tons/ha/year for the French administrative region concerned) on $P_{DOM \rightarrow SYL}$ in wild apple. We then included cultivated apple density as a fixed factor in the models 1 and 2 described below, as the source of the crop-to-wild gene flow. Because the number and area of apple orchards were highly correlated ($r = 0.91$, $P < 0.001$), we investigated the effects of the number and area of apple orchards on $P_{DOM \rightarrow SYL}$ independently in two distinct models (hereafter called models 1 and 2, respectively), controlling for cultivated apple tree density. With this approach, we aimed to determine the relative contributions of orchard number and orchard area to crop-to-wild gene flow in apples. For these analyses, we used the dense sample obtained in France ($N = 1028$ individuals from 20 sites, Fig. S1), and we ran a generalized linear model (GLM) with a quasi-Poisson distribution, as the data appeared to be overdispersed. Apple production (tonnes/year) and the number of and area covered by apple orchards were extracted, for each French administrative region, from the AGRESTE (http://agreste.agriculture.gouv.fr/).

Influence of cultivation and management on the distribution of taxa visiting apple flowers

Anthropogenic effects on introgression rates may also be indirectly mediated by effects on pollinators. An intensification of cultivation and management practices may affect the diversity of flower visitors and, thus, gene flow. We therefore explored the impact of the intensity of cultivation and management on the diversity of the taxa visiting flowers in France (hereafter called model 3). We investigated whether the frequency of apple taxa visiting flowers depended on the intensity of cultivation and management. We ran a GLM with a quasi-Poisson distribution, as the data appeared to be overdispersed. In total, 306 observations corresponding to different intensities of cultivation were recorded (Fig. S1 and Table S3).

The intensity of cultivation and management was obtained from the GLC2000 dataset (The Land Cover Map for Europe in the Year 2000, http://edit.csic.es/Soil-Vegetation-LandCover.html). We defined two classes of cultivation and management intensity, distinguishing between areas with higher and lower intensities of cultivation and management, corresponding to the 1–15 and 17–18 classes of the GLC2000 dataset, respectively. These two classes were defined qualitatively on the basis of the standard classification in the worldwide GLC database as 'natural' or 'cultivated and managed'. Individual geographical coordinates were assigned to these two classes with the extract function (*raster* R package).

The diversity of the insect taxa visiting apple flowers (i.e., the number of observed taxa per insect order) was obtained from a nationwide monitoring exercise based on observations by citizens. Using the database generated by this exercise, entomology experts used the photographs provided to check the preliminary classifications obtained with a simple online key by the

citizens participating in the project (for further information on the SPIPOLL project methodology, see Deguines et al. 2012). As it is difficult to differentiate between *M. sylvestris* and *M. domestica* morphologically, we considered that it would not be possible to assign the observations of citizens to one species or the other; these observations were therefore pooled.

Results

Crop-to-wild gene flow, population structure, and historical gene flow (large-scale analyses, Europe)

Large spatial scale of crop-to-wild gene flow in apple

Failing to take the population structure of the wild species into account can lead to spurious signals of introgression from crop species (Kalinowski 2011). We therefore analyzed the structure of *M. sylvestris* and identified five distinct clusters within the wild species, and in addition a cluster corresponding to *M. domestica* (Fig. S2). The use of *K* values *>6* uncovered no further structure within *M. sylvestris*, indicating that the *M. sylvestris* × *M. domestica* hybrids detected at $K = 6$ were not artifacts due to genetic structure within the wild apple gene pool. We therefore used cumulative membership coefficient in the five *M. sylvestris* clusters in subsequent analyses to identify crop-to-wild hybrid genotypes.

For the total dataset for individuals identified *a priori* as *M. sylvestris* ($N = 1889$), 436 genotypes (23.1% of the total dataset) showed signs of introgression from *M. domestica* (i.e., cumulative membership coefficients <0.9 and >0.1 in the *M. sylvestris* gene pool, $P_{DOM \rightarrow SYL}$) and 77 genotypes were found to correspond to misidentified *M. domestica* trees (i.e., cumulative membership coefficients <0.1 in the *M. sylvestris* gene pool; 4.1% of the total dataset). The mean introgression rates and numbers of hybrids are presented in Fig. 1, Tables S1 and S2. Our results indicated large-scale crop-to-wild gene flow in apple in Europe.

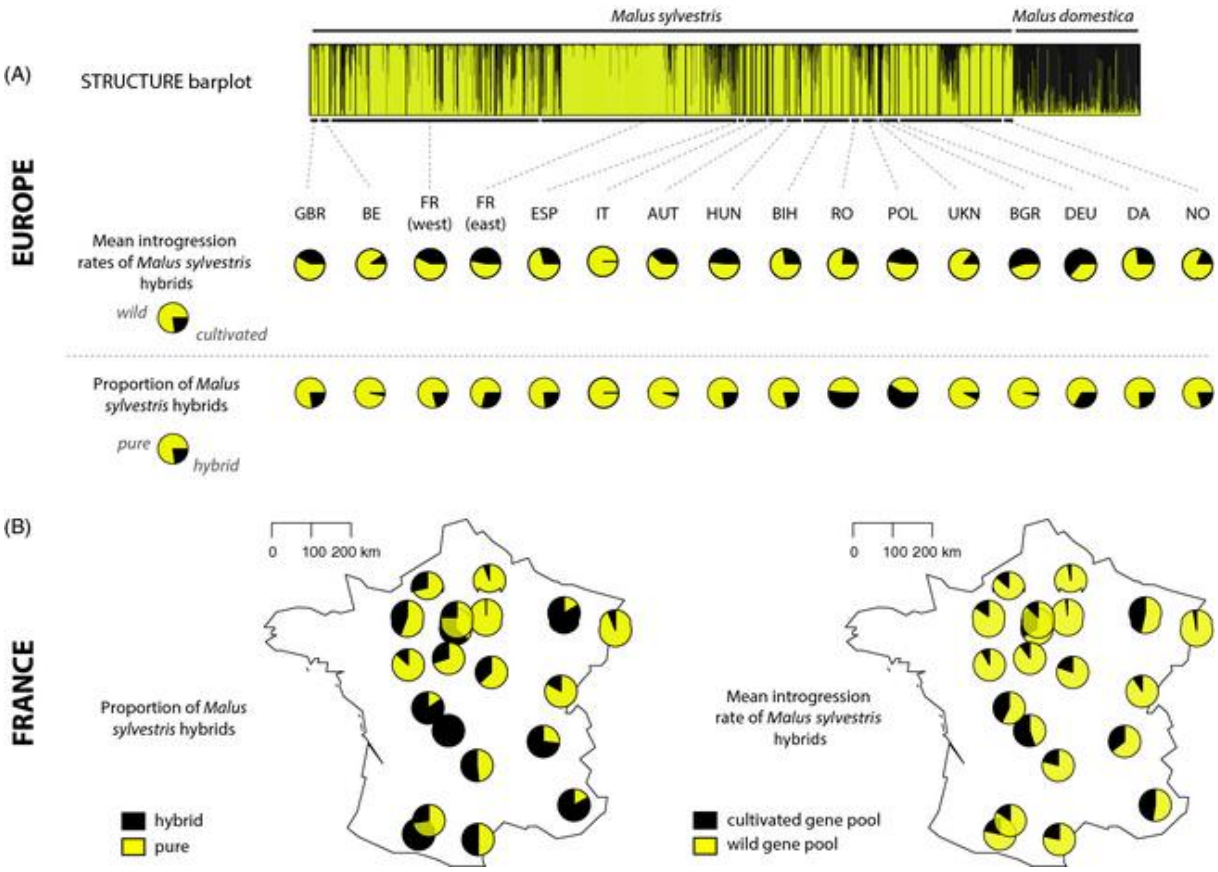


Figure 1. Geographical quantification of introgression rates and numbers of hybrids among European wild apple sites in Europe and, more specifically, in France. (A) STRUCTURE barplot for K *= 6, for the European wild apple gene pool (yellow) and the cultivated apple gene pool (black) at the European scale (*N *= 2228, 66 sites); below, associated country-averaged introgression rates ($P_{DOM \rightarrow SYL}$) and number of hybrids. GB: Great Britain, BE: Belgium, FR (west): Western France, FR (east): Eastern France, ESP: Spain, IT: Italy, AUT: Austria, HUN: Hungary, BIH: Bosnia Herzegovina, RO: Romania, POL: Poland, BGR: Bulgaria, UKN: Ukraine, DEU: Germany, DA: Denmark, NO: Norway. (B) Site-averaged introgression rates ($P_{DOM \rightarrow SYL}$) and number of hybrids in France (*N *= 1092, 20 sites).*

We also observed recent wild-to-crop introgressions into the *M. domestica* cultivars (i.e., cultivars with membership coefficients <0.9 and >0.1 in the *M. domestica* gene pool), as previously demonstrated (Cornille et al. 2012). The mean wild-to-crop introgression rate was 0.15 (Table S3).

Spatial genetic variation and population structure in *Malus sylvestris*

For analyzing the population structure of *M. sylvestris*, introgressed individuals (i.e., genotypes assigned to a wild gene pool with a membership coefficient <0.9) were removed. The final dataset included 1376 individuals (62 sites). For each site, the summary statistics for genetic diversity are shown in Table S4. On average, gene diversity was high across sites and across markers [H_E = 0.79 ± 0.07 (min–max: 0.57–0.89)] and F_{IS} values were low [F_{IS} = 0.10 ± 0.06 (0–0.21)], although there was a highly significant species-wide heterozygote deficit compared to Hardy–Weinberg expectations (P < 0.001). Mean F_{ST} across loci were small [F_{ST} = 0.025 ± 0.04

(0–0.56)], but significant for 1349/1831 pairs of sites ($P < 0.05$). Overall, these results indicate that spatial genetic structure is weak in *M. sylvestris*.

The map of interpolated A_R (Fig. 2C) showed that genetic variability decreased with increasing latitude, with the highest values found in Southern and Eastern Europe. This latitudinal trend was confirmed by the highly significant negative correlation observed between latitude and A_R ($r = -0.75$, $P < 0.001$). A significant positive correlation between longitude and A_R was also found ($r = 0.33$, $P = 0.02$). The map of interpolated H_O (Fig. 2D) showed a significant negative correlation between longitude and H_O ($r = -0.30$, $P = 0.03$); no correlation between latitude and H_O was also found ($r = 0.19$, $P = 0.16$). This latter map highlights the main admixture zones among *M. sylvestris* populations as regions of high diversity.

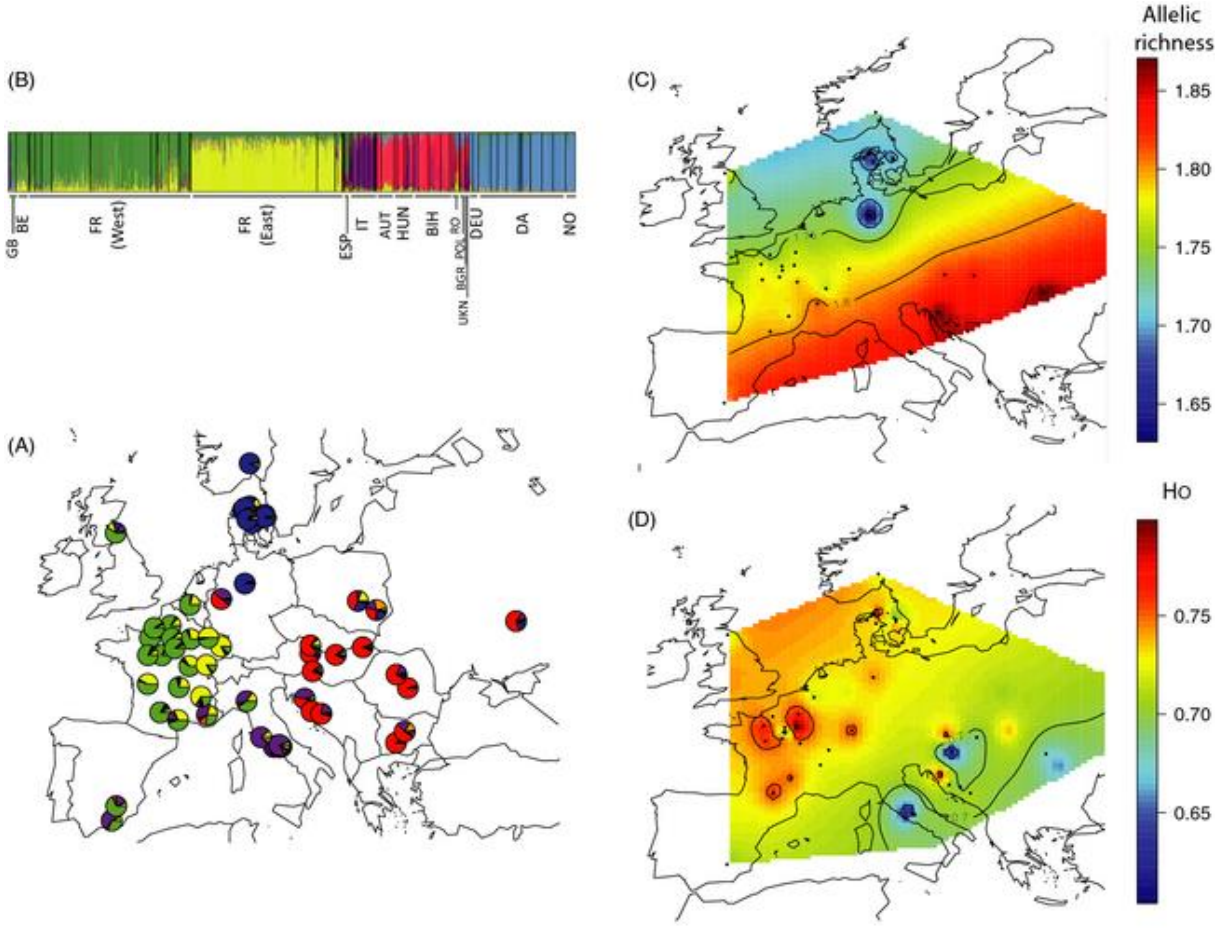


Figure 2. Geographical genetic diversity, genetic structure, and admixture of Malus sylvestris *in Europe (A) Population structure for the nonintrogressed* Malus sylvestris *dataset (*N *= 1376, 62 sites), inferred with TESS, for* $K_{max} = 8$*, showing five distinct genetic clusters. Each vertical line represents an individual. Individuals were grouped by country. Colors represent the inferred ancestry from* K *ancestral populations. For clarity, the 62 sites are grouped by country: GB: Great Britain, BE: Belgium, FR (West): Western France, FR (East): Eastern France, ESP: Spain, IT: Italy, AUT: Austria, HUN: Hungary, BIH: Bosnia Herzegovina, RO: Romania, POL: Poland, BGR: Bulgaria, UKN: Ukraine, DEU: Germany, DA: Denmark, NO: Norway; (B) Map representing the mean membership proportions inferred by TESS for the five detected clusters, for samples of* Malus sylvestris *collected from the same site (62 sites across Europe). (C) and (D) Maps of interpolated allelic richness* A_R *and observed heterozygosity* H_O*, respectively, across Europe.*

Spatially explicit clustering analyses carried out with TESS confirmed the existence of five well-defined clusters within *M. sylvestris* (Figs 2A,B and S3): two of these clusters spanned Western (green) and Central Europe (yellow); one spread over Eastern Europe and the Balkan Peninsula (red); one covered Northern Europe (blue); the last one was located in Italy (purple). DIC values decreased monotonically from K_{max} = 2 to K_{max} = 9, indicating that increases in the number of clusters continually improved the fit of the model to the data, with a sharper decrease in DIC was observed at K_{max} = 8 (Fig. S4) indicating that K_{max} = 8 was the best-fit model. However, at K_{max} = 8, only five clusters still appeared (Fig. S3), as in three of the eight clusters, the membership coefficients of all individuals were close to 0. We therefore concluded that the biologically most relevant number of clusters was five, as observed in STRUCTURE analyses; we nevertheless used the TESS membership coefficient inferred at K_{max} = 8 to define the five populations used in subsequent analyses, as it had the highest DIC value and therefore the best fit to the data. Individuals were assigned to populations for which their membership coefficients exceeded 0.55. The five populations are hereafter referred to as the Western (green, N = 421), Central (yellow, N = 362), Southern (purple, N = 83), Eastern (red, N = 177), and Northern (blue, N = 245) populations. In total, 88 genotypes could not be assigned to any population; N = 1288 individuals were thus retained for population-specific analyses.

For each population, summary statistics for genetic diversity are shown in Table 1. Genetic variability (H_E and H_O, A_R and A_P) differed significantly between all populations (Wilcoxon signed rank (WSR) test, $P < 0.001$) other than the Western and Central populations (WSR, $P > 0.60$), and the Eastern and Southern populations (WSR, $P > 0.44$). Overall, the Eastern and Southern populations displayed significantly higher levels of genetic variability, in terms of allelic richness (A_R and A_P), than the Northern, Western, and Central populations. Higher heterozygosities (H_E and H_O) were considered to be footprints of recent zones of admixture between different genetic clusters (i.e., the Western/Central populations and Central/Eastern populations).

Table 1. Genetic polymorphism, population structure, and the spatial pattern of genetic differentiation in *Malus sylvestris* (N = 1288 individuals)

	Northern	Eastern	Southern	Western	Central
N	245	177	83	421	362
H_O	0.66	0.79	0.84	0.72	0.74
H_E	0.72	0.89	0.89	0.78	0.79
F_{IS}	0.08***	0.11***	0.06***	0.07***	0.07***
A_R	1.72	1.89	1.89	1.78	1.79
A_P	0.67	1.17	1.19	0.73	0.79
Population structure (colors as in Fig. 2)					
Q_1 (blue)	0.92	0.02	0.02	0.01	0.04
Q_2 (red)	0.04	0.85	0.13	0.04	0.02
Q_3 (purple)	0.02	0.11	0.84	0.04	0.05
Q_4 (green)	0.00	0.01	0.01	0.84	0.04
Q_5 (yellow)	0.01	0.01	0.01	0.07	0.86
Spatial pattern					
Mean(Ln(dist))	0.53	0.34	1.17	0.77	1.55
b	−0.014***	−0.021***	−0.007***	−0.011**	−0.007**
$F_{(1)}$	0.04	0.06	0.06	0.06	0.008
r^2	0.08	0.15	0.03	0.05	0.005
Sp	0.014	0.022	0.007	0.011	0.007

N, *sample size of each population;* H_O *and* H_E, *observed and expected heterozygosity, respectively;* F_{IS}, *inbreeding coefficient;* A_R, *mean allelic richness, corrected by the rarefaction method, estimated for a sample size of 2;* A_P, *private allelic richness, corrected by the rarefaction method, estimated for a sample size of 100;* $Q_{1\text{-}5}$, *mean membership coefficient for each of the five clusters inferred by TESS analyses, including hybrids with membership coefficients of up to 0.55 for the cluster concerned; Sp, Sp parameter; mean(Ln(dist)), mean of the logarithm of the geographic distance between genotypes;* b, *slope of the regression between* F_{ij} *and the logarithm of geographic distance;* $F_{(1)}$, *mean* F_{ij} *between individuals from the first distance class;* r^2, *squared correlation coefficient for the relationship between the logarithm of geographic distance and* F_{ij}; **, $0.01 < P \leq 0.001$; ***, $P < 0.001$.

Proportions of admixture differed between the five populations (Fig. 2, Tables 1 and 2). Substantial admixture was detected between the Western, Southern, and Northern populations, with hybrid numbers highest for the Western and Central populations (N = 321 admixed individuals) and then for the Northern and Western populations (N = 41) (Table 2). The Eastern cluster mostly displayed admixture with the Southern (N = 31), Northern (N = 21), and Central

(N = 21) populations. The Southern population displayed admixture mostly with the Eastern (N = 31) population and the Western population (N = 23). The admixture detected is indicative of gene flow between these populations. Genetic differentiation (F_{ST}) was greatest between the Central/Western/Northern and Southern/Eastern populations (Table 2). These results indicate a strong west–east differentiation of allelic frequencies, that was detected at K = 2 by TESS and STRUCTURE (Figs S2 and S3).

Table 2. Genetic differentiation (upper triangle, F_{ST}) and total number of admixed genotypes (lower triangle, i.e. individuals assigned to the population with membership coefficients of 0.55–0.9) among the five populations of *Malus sylvestris*

	Northern	Western	Central	Southern	Eastern
Northern	–	0.03	0.06	0.13	0.11
Western	41	–	0.04	0.09	0.08
Central	14	321	–	0.09	0.08
Southern	2	23	11	–	0.04
Eastern	21	13	21	31	–

All F_{ST} *values were highly significant* ($P < 0.001$).

Spatial autocorrelation, isolation by distance (IBD), and historical gene flow

Within each of the five populations, genetic differentiation and geographical distance were significantly correlated, consistent with an IBD model (Table 1). The Sp statistic was low suggesting weak spatial genetic structure and high levels of historical gene flow and/or effective population size. Overall, the mean Sp value was 0.01.

Factors affecting crop-to-wild gene flow and flower visitors (finer scale analyses in France)

Anthropogenic factors affecting crop-to-wild introgression

For model 1, the proportion of admixture of the *M. domestica* gene pool into the *M. sylvestris* gene pool ($P_{DOM \rightarrow SYL}$) was found to be dependent on the interaction between cultivated apple density and the number of apple orchards (Table 3). Indeed, for small numbers

of apple orchards, $P_{DOM→SYL}$ were unaffected by cultivated apple density, whereas cultivated apple density had a positive effect on $P_{DOM→SYL}$ for large numbers of orchard apple farms. For model 2, $P_{DOM→SYL}$ appeared to depend on the interaction between cultivated apple density and the area of apple orchards (Table 3) with the same trend found in model 1

Table 3. Summary of minimal models 1, 2, and 3, which were tested for the identification of factors affecting crop-to-wild gene flow and the diversity of visitors to apple flowers

Model	Response variable	Explanatory variable	*F*-value (df)	*P*-value
Model 1	$P_{DOM→SYL}$	Cultivated apple density	0.67 (1,563)	0.387
		Number of apple orchards	46.64 (1,563)	<0.001
		Number of apple orchards * Cultivated apple density	31.53 (1,563)	<0.001
Model 2	$P_{DOM→SYL}$	Cultivated apple density	10.01 (1,563)	<0.001
		Area of apple orchards	27.20 (1,563)	<0.001
		Area of apple orchards * Cultivated apple density	25.92 (1,563)	<0.001
Model 3	Diversity of taxa	Intensity of cultivation and management	125.12 (1,306)	<0.001
		Order	41.97 (1,306)	<0.001

*$P_{DOM→SYL}$, proportion of admixture of the Malus domestica gene pool into the Malus sylvestris gene pool; df, degrees of freedom; x*y, interaction between explanatory variables x and y.*

Factors affecting the diversity of flower visitors

Apple species were visited by 86 taxa from six orders of insects (Fig. 3 and Table S5). In model 3, the frequency of taxa visiting apple flowers was significantly lower in intensively cultivated and managed areas (26 taxa, 59 observations) than in less intensively cultivated and managed areas (83 taxa, 247 observations) (Fig. 3 and Table 3), after controlling for the order to

which each taxon belonged. This relationship was also observed for all the insect orders considered (Fig. 3 and Table 3, $P < 0.0001$).

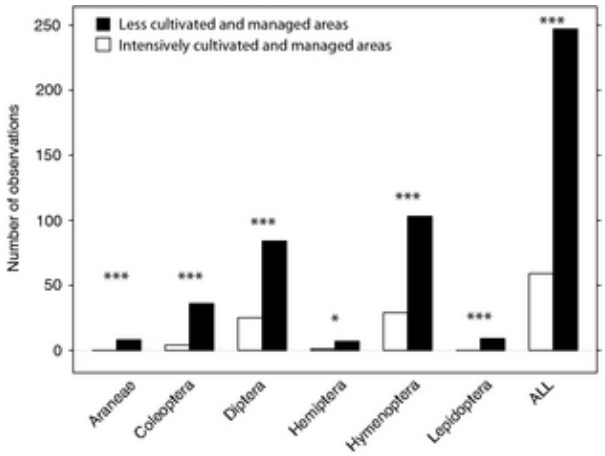


*Figure 3. Number of observations for the six orders of insect visitors to apple flowers in France (N = 306), in areas with more intensive (white) and less intensive (black) cultivation and management. ALL: number of observations for all orders considered together. ***: P < 0.0001; *: P < 0.01.*

Discussion

Previous studies investigating crop-to-wild gene flow in fruit trees have provided estimates of introgression rates, but did not identify the factors influencing these rates (Delplancke et al. 2011; Hoban et al. 2012; Aerts et al. 2013). This study provides new insight into the impact of human activities on the extent of recent crop-to-wild gene flow in fruit trees. We detected (i) large-scale recent introgressions from the cultivated apple to *M. sylvestris*, the extent of which was influenced by the intensity of apple production, and (ii) a decrease in the diversity of apple flower visitors with increasing intensity of cultivation and management practices, suggesting an indirect impact of land-use on pollen flow.

Our study also unraveled the genetic structure of a scattered temperate European fruit tree, through comprehensive sampling. It provides new insight into the history of the wild apple, by revealing the influence of past climatic changes and barriers to dispersal in shaping the genetic structure of this forest fruit tree, resulting in the formation of differentiated genetic clusters in Europe. We also estimated levels of inter- and intracluster gene flow, and we discuss the fundamental and applied impact and significance of these findings below.

Human impact on crop-to-wild gene flow in apples

Since its introduction into Europe by humans 1500 years ago, the cultivated apple has acted as a source of genes for introgression into the European wild apple (Cornille et al. 2013b). We found that the rates of crop-to-wild introgression varied across Europe and could be accounted for by direct and indirect anthropogenic factors. Interestingly, the density of cultivated apple trees, as estimated from apple production, was found to be positively related to recent introgression rates in wild apple populations, with the number of apple orchards surrounding wild populations and with the total orchard area. These results suggest that crop-to-wild introgression in apple is both dependent on the distribution of orchards within the landscape and on their area. It seems intuitive that several large apple orchards scattered across the landscape would act as a more efficient source of crop-to-wild gene flow than a smaller number of small orchards. Our results indicate that human management practices have also had an impact on the insect community visiting apple flowers in France. The numbers of insect orders visiting apple trees were smaller in managed and cultivated areas, suggesting that human disturbance has an indirect impact on gene flow, through negative effects on potential pollinator populations. Further investigations are required to determine the extent to which the changes in apple visitor communities according to land-use affect gene flow from crop-to-wild species and between populations in wild species.

New insights into the demographic history of the European wild apple

Past climate changes and spatial barriers to dispersal across Europe can explain the population structure of wild apples revealed here. Over the last 2.6 million years, trees species have experienced dynamic alternations between the contraction and expansion of their ranges in Europe, in response to climate changes (Hewitt 1996, 1999; Petit et al. 2003). Our investigations of the population structure and diversity of *M. sylvestris*, based on the use of unprecedented dense sampling in France, provide further insight into the distribution ranges of apples in Europe in the past, over and above that provided by previous studies (Coart et al. 2006; Larsen et al. 2006; Cornille et al. 2013a; Schnitzler et al. 2014). We identified five well-differentiated gene pools with specific spatial distributions, whereas our previous sampling campaign led to the identification of only three clusters (Cornille et al. 2013a). Furthermore, the denser sampling used in this study made it possible to identify hotspots of allelic richness located in the South of France, Italy, the Balkans, and the Carpathian Mountains. The South of France, the Balkans, and the Carpathian Mountains have already been identified as glacial refugia

(Cornille et al. 2013a), but our results suggest that Italy and the south-eastern Alps may also have served as glacial refugia for wild apple trees, as already demonstrated for other tree species (Tzedakis et al. 2013). We also found a longitudinal trend for allelic richness that could be potentially accounted for by an east–west gradient of climate severity in the Mediterranean region during the last glacial maximum (Conord et al. 2012). Spain would thus have been a minor glacial refugium for *M. sylvestris*, due to its harsher climate. Three main contact zones between clusters were identified, in the South-eastern Alps (between the green and purple clusters), in France (between the green and yellow clusters), and in the Carpathian Mountains (between the red and purple clusters). These results suggest that previous contractions of the distributional range of wild apple during the last glacial maximum in southern Europe were followed by a recolonization of Northern Europe via two main routes: a western front (South of France—Italy, up to Scandinavia) and an eastern front (Balkans—Carpathian Mountains, up to Scandinavia). In particular, the additional Italian cluster detected appears to have acted as a bridge for the recolonization of the western and eastern parts of Europe after glaciation. Barriers to dispersal during recolonization and bottlenecks may also have affected the existing population structure. For instance, populations from northern Germany, Denmark, and Norway have a low allelic richness, and a bottleneck might account for these populations belonging to a distinct gene pool.

Weak isolation by distance reveals high historical levels of gene flow during the recolonization of Europe

Our findings suggest that, despite the observed population structure, the level of natural intraspecies gene flow during the course of history was high for a wild fruit tree. Indeed, we detected a weak spatial genetic structure within each cluster, spanning large geographic distributions (with low Sp values), suggesting relatively high dispersal capacities and/or effective population sizes (Vekemans and Hardy 2004). Long-distance dispersal events have been reported for another temperate fruit tree, *Sorbus domestica* (Kamm et al. 2009). Nevertheless, we might have expected to find a stronger genetic structure in this zoochorous fruit tree, because fruits carry half-sib or full-sib seeds as dispersal units, rendering dispersal nonrandom. This would be expected to lead to a stronger genetic structure than is observed in wind-pollinated trees (Hardy et al. 2006). However, the spatial genetic structure detected in this study was as weak as that detected in wind-dispersed trees. This may reflect the small number of generations since European recolonization after the last glacial maximum (<10 000 years ago, corresponding to about 1000 generations). Too little time has thus elapsed for genetic differentiation to occur,

particularly given the long generation time, overlapping generations and high effective population sizes of trees (Austerlitz et al. 2000). We used genetic distances between individuals to estimate gene flow. This method principally reflects historical patterns of gene flow (Manel et al. 2003). Our estimates may thus not be indicative of local-scale pollen and seed gene flows in contemporary landscapes. Only genetic analyses of parentage (Smouse and Sork 2004) would provide an estimate of the current dispersal abilities of the wild European apple.

Conclusion and future prospects: a basis for future conservation and breeding programs

The consequences of crop-to-wild gene flow for the wild apple remain unclear, particularly as concerns the possibility of genetic swamping (Lenormand 2002). This phenomenon occurred in the wild ancestor of coconut (*Cocos nucifera*), in which introgression from domesticated populations led to the disappearance of all genetically pure wild individuals (Ellstrand 2014). Crop-to-wild gene flow in apple has been shown to be responsible for a decrease in fitness in populations of the European wild apple in Germany (Bleeker et al. 2007). The European wild apple may also be vulnerable to changes in the diversity of insect visitors to its flowers, resulting in less efficient pollination services, for instance. Given the emblematic nature of the wild apple in Europe and its usefulness as a resource for future breeding programs (Cornille et al. 2014), conservation measures should be urgently considered.

The implementation of conservation programs is challenging, in terms of the application of the recommended practices in the field. Landscape genetics can provide useful information for conservation management, in the form of fundamental knowledge for the identification of high-priority populations for the conservation of genetic variation. Our landscape genetic studies of within-species gene flow and crop-to-wild gene flow in apples in Europe should facilitate the allocation of priority zones for the conservation of the European wild apple, across its natural distribution. We detected five main genetic groups, within which it would be optimal to conserve populations. Southern Europe, housing the main glacial refugia with the highest levels of genetic diversity, should also be considered a priority area for conservation programs, together with France and the Carpathian Mountains, the main suture zones, with their original allelic combinations. The identification of highly introgressed European wild apple populations and of the anthropogenic drivers of crop-to-wild introgressions will also help to dynamically manage these priority conservation zones. This *in situ* conservation could also be combined with

the *ex situ* conservation of genetic resources through the creation of gene banks for a core collection of nonintrogressed *M. sylvestris* populations from diverse European regions.

Our landscape genetic investigations would also help to identify the optimal provenance for seeds or germplasm for the reintroduction of wild apples into agroforestry programs (Krauss et al. 2013). Wild apple trees are, indeed, currently being planted in France in the context of agroforestry programs, but the provenance and the degree of introgression of the wild apple germplasm used have never been considered.

In conclusion, the relatively recent introduction of the cultivated apple into Europe, only about 1500 years ago, has particularly facilitated the detection of wild-to-crop and crop-to-wild gene flow and the identification of its determinants. The cultivated-wild apple complex in Europe thus appears to be a good model in which to address both fundamental and applied issues relating to the impact of human landscape changes on gene flow.

Statement of authorship

AC, PG, and TG conceived and designed the experiments; AC, PG, and TG obtained funding; AC and AF performed the experiments; AC and TG supervised the Master's thesis of AF; AC, AF, and UG analyzed the data. KV and JR provided samples. The manuscript was written by AC with critical input from UG and TG.

Acknowledgements

We thank the *Région Ile de France* (PICRI and DIM R2DS), IDEEV (*Institut Diversité Ecologie et Evolution du Vivant*), SBF (*Société Botanique de France*), and laboratoire d'Excellence BASC (*Biodiversité, Agrosystèmes, Société et Climat*) for funding, and the *Plateforme de Génotypage* GENTYANE INRA UMR 1095 for assistance with genotyping. We thank all those who helped with sampling (Table S1). We thank Sébastien Rioux Paquette and Guillaume Larocq for fruitful discussions, Colin Fontaine for sharing SPIPOLL data on apples and anonymous referees for very useful suggestions. The authors declare no conflict of interest.

Data accessibility

Data for this study (genetic, visitor flower diversity and relative to apple production) are available at the Dryad Digital Repository: http://doi.org/10.5061/dryad.bm46p.

References

Aerts, R., G. Berecha, P. Gijbels, K. Hundera, S. Van Glabeke, K. Vandepitte, B. Muys et al. 2013. Genetic variation and risks of introgression in the wild Coffea arabica gene pool in south-western Ethiopian montane rainforests. *Evolutionary Applications* 6:243–252.

Arnaud, J. F., F. Viard, M. Delescluse, and J. Cuguen 2003. Evidence for gene flow via seed dispersal from crop to wild relatives in Beta vulgaris (Chenopodiaceae): consequences for the release of genetically modified crop species with weedy lineages. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270:1565–1571.

Austerlitz, F., S. Mariette, N. Machon, P. H. Gouyon, and B. Godelle 2000. Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics* 154:1309–1321.

Bleeker, W., U. Schmitz, and M. Ristow 2007. Interspecific hybridisation between alien and native plant species in Germany and its consequences for native biodiversity. *Biological Conservation* 137:248–253.

Chen, C., E. Durand, F. Forbes, and O. François 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Notes* 7:747–756.

Coart, E., X. Vekemans, M. J. M. Smulders, I. Wagner, J. Van Huylenbroeck, E. Van Bockstaele, and I. Roldán-Ruiz 2003. Genetic variation in the endangered wild apple (Malus sylvestris (L.) Mill.) in Belgium as revealed by amplified fragment length polymorphism and microsatellite markers. *Molecular Ecology* 12:845–857.

Coart, E., S. Van Glabeke, M. De Loose, A. S. Larsen, and I. Roldán-Ruiz 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.

Conord, C., J. Gurevitch, and B. Fady 2012. Large-scale longitudinal gradients of genetic diversity: a meta-analysis across six phyla in the Mediterranean basin. *Ecology and Evolution* 2:2600–2614.

Cornille, A., P. Gladieux, M. J. Smulders, I. Roldán-Ruiz, F. Laurens, B. Le Cam, A. Nersesyan et al. 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genetics* 8:e1002703.

Cornille, A., T. Giraud, C. Bellard, A. Tellier, B. Le Cam, M. J. M. Smulders, J. Kleinschmit et al. 2013a. Post-glacial recolonization history of the European crabapple (Malus sylvestris Mill.), a wild contributor to the domesticated apple. *Molecular Ecology* 22:2249–2263.

Cornille, A., P. Gladieux, and T. Giraud 2013b. Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *Evolutionary Applications* 6:737–748.

Cornille, A., T. Giraud, M. J. M. Smulders, I. Roldán-Ruiz, and P. Gladieux 2014. The domestication and evolutionary ecology of apples. *Trends in Genetics* 30:57–65.

Deguines, N., R. Julliard, M. de Flores, and C. Fontaine 2012. The whereabouts of flower visitors: contrasting land-use preferences revealed by a country-wide survey based on citizen science. *PLoS One* 7:e45822.

Delplancke, M., N. Alvarez, A. Espíndola, H. Joly, L. Benoit, E. Brouck, and N. Arrigo 2011. Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers. *Evolutionary Applications* 5:317–329.

Ellstrand, N. C. 2003. Current knowledge of gene flow in plants: implications for transgene flow. Philosophical Transactions of the Royal Society of London. Series B, *Biological sciences* 358:1163–1170.

Ellstrand, N. C. 2014. Is gene flow the most important evolutionary force in plants? *American Journal of Botany* 101:737–753.

Ellstrand, N. C., H. C. Prentice, and J. F. Hancock 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* 30:539–563.

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

Feder, J. L., S. P. Egan, and P. Nosil 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics* 28:342–350.

Garant, D., S. E. Forde, and A. P. Hendry 2007. The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology* 21:434–443.

Gross, B., A. Henk, P. Forsline, C. Richards, and G. Volk 2012. Identification of interspecific hybrids among domesticated apple and its wild relatives. *Tree Genetics and Genomes* 8:1223–1235.

Hardy, O. J., and X. Vekemans 2002. SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2:618–620.

Hardy, O. J., L. Maggia, E. Bandou, P. Breyne, H. Caron, M.-H. Chevallier, A. Doligez et al. 2006. Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology* 15:559–571.

Hewitt, G. M. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58:247–276.

Hewitt, G. M. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 68:87–112.

Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences* 359:183–195.

Hoban, S. M., T. S. McCleary, S. E. Schlarbaum, S. L. Anagnostakis, and J. Romero-Severson 2012. Human-impacted landscapes facilitate hybridization between a native and an introduced tree. *Evolutionary Applications* 5:720–731.

Jakobsson, M., and N. A. Rosenberg 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806.

Kalinowski, S. T. 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity* 106:625–632.

Kamm, U., P. Rotach, F. Gugerli, M. Siroky, P. Edwards, and R. Holderegger 2009. Frequent long-distance gene flow in a rare temperate forest tree (Sorbus domestica) at the landscape scale. *Heredity* 103:476–482.

Krauss, S. L., E. A. Sinclair, J. D. Bussell, and R. J. Hobbs 2013. An ecological genetic delineation of local seed-source provenance for ecological restoration. *Ecology and Evolution* 3:2138–2149.

Kremen, C., N. M. Williams, M. A. Aizen, B. Gemmill-Herren, G. LeBuhn, R. Minckley, L. Packer et al. 2007. Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change. *Ecology Letters* 10:299–314.

Larsen, A., C. Asmussen, E. Coart, D. Olrik, and E. Kjær 2006. Hybridization and genetic variation in Danish populations of European crab apple (Malus sylvestris). *Tree Genetics and Genomes* 2:86–97.

Lenormand, T. 2002. Gene flow and the limits to natural selection. *Trends in Ecology & Evolution* 17:183–189.

Loiselle, B. A., V. L. Sork, J. Nason, and C. Graham 1995. Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). *American Journal of Botany* 82:1420–1425.

Manel, S., and R. Holderegger 2013. Ten years of landscape genetics. *Trends in Ecology & Evolution* 28:614–621.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet 2003. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* 18:189–197.

Manning, A. D., J. Fischer, and D. B. Lindenmayer 2006. Scattered trees are keystone structures – Implications for conservation. *Biological Conservation* 132:311–321.

Patocchi, A., F. Fernández-Fernández, K. Evans, D. Gobbin, F. Rezzonico, A. Boudichevskaia, F. Dunemann, et al. 2009. Development and test of 21 multiplex PCRs composed of SSRs spanning most of the apple genome. *Tree Genetics & Genomes* 5:211–223.

Petit, R. J., I. Aguinagalde, J.-L. de Beaulieu, C. Bittkau, S. Brewer, R. Cheddadi, R. Ennos et al. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. *Science* 300:1563–1565.

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

Raymond, M., and F. Rousset 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86:248–249.

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

Schnitzler, A., C. Arnold, A. Cornille, O. Bachmann, and C. Schnitzler 2014. Wild European apple (Malus sylvestris (L.) Mill.) population dynamics: insight from genetics and ecology in the rhine valley. Priorities for a future conservation programme. *PLoS One* 9:e96596.

Smouse, P. E., and V. L. Sork 2004. Measuring pollen flow in forest trees: an exposition of alternative approaches. *Forest Ecology and Management* 197:21–38.

Storfer, A., M. A. Murphy, S. F. Spear, R. Holderegger, and L. P. Waits 2010. Landscape genetics: where are we now? *Molecular Ecology* 19:3496–3514.

Szpiech, Z. A., M. Jakobsson, and N. A. Rosenberg 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24:2498–2504.

Tzedakis, P. C., B. C. Emerson, and G. M. Hewitt 2013. Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution* 28:696–704.

Vekemans, X., and O. J. Hardy 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* 13:921–935.

Table S1 Description of the *Malus* accessions analyzed, with their geographical origins (locality, latitude and longitude) and providers.

Species and country of origin	Population Name	Locality	Latitude	Longitude	N	Providers
***Malus sylvestris* (N=1,812*)**						
Austria	Aut_hern	Hernstein	47.8843	16.1152	12	Konrad HEINO
	Aut_stock	Stockereu	48.3862	16.1300	32	Thomas KIRISITS
Belgium	fland	Blanden, Vorselaar, Bilzen, Wynendale	50.9948	4.4114	30	Kristine VANDER MIJNSBRUGGE
Bosnia and Herzegovina	Bih_jab	Jabuka	44.9711	17.1949	30	Dalibor BALIAN
	Bih_kuk	Kukavica	44.0044	16.2315	22	Dalibor BALIAN
	Bih_Liv	Livno	44.1594	17.1622	28	Dalibor BALIAN
	Bih_Olo	Olovo	43.9950	18.0492	30	Dalibor BALIAN
	Bih_Pre	Preyj	43.5203	24.4306	22	Dalibor BALIAN
Bulgaria	Bgr_Lov	Lovech	43.1348	24.4562	2	Petya GERCHEVA
	Bgr_Plo	plovdiv Zolravez	42.1438	25.2994	2	Petya GERCHEVA
	Bgr_tria	Triavna	42.8673	25.4927	1	Petya GERCHEVA
Denmark	Da_Aaby	Aaby Bjerg	57.1920	9.7596	30	Einar FLENSTED-JENSEN
	Da_he	Hestehave	56.2600	10.4500	29	Anders LARSEN
	Da_Hvolg	Hvolgaard/Oxholm Skov	57.0640	9.5620	34	Einar FLENSTED-JENSEN
	Da_k	Kostekov	56.7660	11.6830	30	Anders LARSEN
	Da_ka	Kaas	56.6167	11.6830	29	Anders LARSEN
	Da_no	Noerbjerg	57.1833	9.7500	30	Anders LARSEN
	Da_Norha	Nørhalne Grusgrav	57.1611	9.8551	6	Einar FLENSTED-JENSEN
	Da_Nymo	Nymølle	57.3051	10.2286	50	Einar FLENSTED-JENSEN
	Da_Oxholm	Oxholm Skov	57.0843	9.5830	19	Einar FLENSTED-JENSEN
	Da_Skeel	Skeelslund Skov	57.1245	9.7507	23	Einar FLENSTED-JENSEN
France	Alsace	Colmar, Gerstheim, Hautecour, Mollkirch, Offendorf, Hagueneau, Volksberg, Plaine, Fouchy, Breitenau, Mietesheim, Butten, Marmoutier, Keskatel, Eschbourg, Wimmenau, Kiffis, Lucelle, Winkel, Daubensand, Ernstein, Hardtl, Illkirch-Graffenstaden, Mackenheim, Rhinau, Hombourg, Rouffach	48.3182	7.4416	333	Pascal HEITLZER, Annik SCHNITZLER, Dominique OESTERLE, Philippe HUM, Hubert HOLVECK, Jean-Marc WILHELM, Gérard LAVAUPOT, Pierre BENDHIF-SYLLAS, Alexandra LOUIS, Jean BRAUD
	Aquitaine	Brocas les Forges, Chéraute	44.6977	0.2916	1	Roman CLAVEAU, Hubert FRANCAIS
	Auvergne	Saint-Rémy-en-Rollat, Droiturier, Salettes, Job, Charbonnières-les-Varennes, La Tour d'Auvergne,	45.7033	3.3449	56	Christian AMBLARD, GATHIER, Thomas DUMAS, BRUGHEAILLE, A. VALNET, Claire EDELIST, Jean-

	Tauves, Labessette, Trémouille, Saint Loup, Bagnols, Montmurier, Durdat, Larequille, Arpheuilles-Saint-Priest, Condat-lès-Montboissiers, Brousse, Auzelles, Domaize, Néris les Bains				Claude GAUDIAT
BasNorm	Saint-Clément-Rancoudray, Sainte Croix sur Orne, Caen, St-Germain de la Coudre, Gémages, Ceton	48.8788	0.5158	28	Aurélien CABARET, J.P. RIOULT, Samuel LAGRANGE, Michel LEMOINE, D. J. CHEVALIER
Bourg	Nicey, Raveau, La Marche, Chaulgnes, Chalaux, La Celle-sur-Loire, Malvaux, Michaugues, Limon, Bona, Chevagny-sur-Guye, Massangis, Nature Reserve of Bois du Parc, Coulanges-la-Vineuse, Saint-Germain-des-Champs	47.0607	4.3523	38	Cyril METIER, Edith VALETTE, Jerome BIDAUT, Claude CHAPALAIN, Arnauld OHL, Benoit FRITSCH, Christine COULOMB, Jacques MARCHAND, Eric FEDEROFF
Bretagne	Penmarc'h, Huelgoat, Ile de Sein, forest of d'Argentré du Plessis, Brielles, Paimpol	48.2022	2.9323	11	Dominique BEAUVAIS, Michel LEMOINE, Aurélia LACHAUD, Michelle SAUVE
Centre	Mery-ès-bois, forest of Saint Palais, La Chapelle Saint-Ursin, Luzillé, Souvigné, Forest of Montargis, Triguères, Orléans	47.7516	1.6751	20	Thierry GAUTROT, Jean-Claude BOURDIN, Martine CHOISNE, Françoise LE BERRE, Hugues SAVAY-GUERRAZ
ChampArd	Pont-Sainte-Maxence, Bernon	48.7916	4.4719	4	Christophe GALET, PETITPREZ
FranCom	Gilley, Morteau, Mouthe, Arbecey, Les Crozets, Moirans-en-Montagne, Charchilla, Forest of Revigny, Forest of Briod, Meussia, Grande-Rivière, Etival, Clairvaux-les-lacs, Forest of the Château des Près, La Rixouse, Les Rousses, Marnoz, Blois sur Seille, Menétru le Vignoble	47.1342	6.0218	53	Pascal GARDAVAUD, Jean GREMION, Florent LE MENE, Eric WOLF, Roberto FERRETO, Jean BRAUD, Odile GIROD, Alexis GENET
HteNorm	Sain-Aubin-Epinay, Montmain, St-Pierre-du-Vouvray, Belbeuf, forest of Bords	49.5065	0.9302	35	T. BERTONCINI
IleFrance	Beauvais, Loutteville, Dourdan, Angevilliers, Forest of Rambouillet, Forest of Carnelle, Roissy en Brie, Boissy aux Cailles, Bréviandes, Forest of Fontainebleau, Chanfroy, Viroflay, St Arnoult en Yvelines	48.8494	2.6381	195	Jean-Marie GRIMLER, Alodie SNIRC, Amandine CORNILLE, Didier DESSEAUX, Annick LARBOUILLAT, Françoise LE BERRE, Denis CHEVALLIER, Henri FOUREY, Jacqui GUIEYSSE, Christian CASSONNET
LangRous	Celle, Bassurels, Clermont sur Lauquet, Lairière, Greffeil, Boule d'Amont	43.5913	3.2583	6	Françoise DREGER, Marc VODOVAR, Udo POLHMANN
Limousin	Saint-Pardoux-Morterolles, Saint-Pierre-Bellevue, Royère-de-Vassivière, Le Donzeil	45.8974	1.5796	5	Jerome VANY, D. LE STRAT
Lorraine	Bionville, Reville-aux-Bois, Villiers-sur-Meuse, Burey-en-Vaux, Ambly-sur-Meuse, Bouchepern	48.8744	6.2081	12	Jean BRAUD, Jules WIZNIAK, Romain CHARLE, Yvon DALBORGO, François VERNIER
MidiPy	Foissac, Anson, Crostes, Montferran-Plavès, Saillet, Cauterets, Gèdre, Sousceyrac, Toulouse, Marcillat-en-Combraille	44.0859	1.5209	33	Dominique BEAUVAIS, Monique FROBERT ADAMO, Benoit VIGREUX, Jean-Claude GAUDIAT, Denise COMERA, Michel JAULIN, Emilie ESTIVAL, Nicolas FEAU, Tatiana GIRAUD
PaysLoi	Faye d'Anjou, Angers, Saint Barthélémy d'Anjou	47.7631	0.3300	15	Michèle RENOU, Yves LESPINASSE, Charles-Eric

						DUREL, Colette ROUTKOVSKY, Mireille CASSIN
	Picardie	Samoussy, Forest of Coucy Basse, Forest of Champs, Forest of Saint-Gobain, Verneuil-en-Halatte, Givet	49.6631	2.5285	100	Maurice ELARD, Christophe GALET, Philippe JACQUET, Franck LECOMPTE
	PoitChar	Tonnay-Boutonne, Terce, Vienne, Maviaux, Bonneuil-Matours, Vouneuil-sur-Vienne	45.9036	0.3092	13	Jean Renaud LEMAITRE, Philippe ROY, Maryvonne PACAUD, Bruno MERCIER, Hugues SAVAY-GUERRAZ
	ProACA	Saint-Jurs, Gorges de Daluis, Saint Martin d'Entraunes	43.9352	6.0679	6	Iliane VIAL, Françoise DUMAS
	RhonAlpes	La Balme, La Faurie, Lus-la-Croix-Haute, Les Communaux, Lake of Moras, Forest of Chambaran, Saint-Ismier, La Colombe, Valley of Garon, Chaponost, Benonce, Villard Reymond, Colliard, Montoux Brenod, Darbres, Lus-la-Croix-haute, Araches la Frasse	45.0937	5.4886	63	Philippe SURDON, Roberto FERRETO,Pierre GLADIEUX, Hugues SAVAY-GUERRAZ, Philippe JACQUES, Jean-Luc RICAUD, Christophe BONNET
	Deu_pop1	Reinhausen	56.4653	9.9831	21	Jorg KLEINSCHMIT, Wilfried STEINER
Germany	Deu_pop2	Kirchwehren	52.3583	9.7510	14	Jorg KLEINSCHMIT, Wilfried STEINER
	Deu_pop3	Hagen	51.3640	7.4599	10	Jorg KLEINSCHMIT, Wilfried STEINER
	Hun_p1	Erdőbénye	48.2682	21.2161	10	Lazlo NYARI
Hungary	Hun_p2	Visegrád	47.7798	18.5961	19	Lazlo NYARI
	Hun_p3	Pórszombat	46.7288	16.3459	11	Lazlo NYARI
	It_Cerva	Cervara di Roma, Prataglia	41.8948	13.3159	14	Alberto DOMINICCI
	It_Fil	Filettino, Colle Cardiglioso, Val Granara	41.8905	13.1978	4	Alberto DOMINICCI
	It_Pag	Camerata Nuova, Le Paghette	41.9972	13.1050	2	Alberto DOMINICCI
	It_parme	Salsomaggiore Terme	44.8112	9.9165	3	Stefano PORTA, Alfredo MARTIGNONI
Italy	It_rome	Fara Sabina (Rieti), Castel di Guido	41.9730	12.8101	5	François SALOMONE
	It_Subia	Subiaco, Valle delle mele	41.8718	13.2452	9	Alberto DOMINICCI
	It_Tosc	Manciano	42.4990	11.3208	1	François SALOMONE
	It_Trevi	Trevi nel Lazio, Valle Competa	41.8629	13.1516	11	Alberto DOMINICCI
	It_valle	Vallepietra, Fosso Fioio	41.9918	13.0949	13	Alberto DOMINICCI
	Italie	NA	38.7123	16.2792	5	Annick SCHNITZLER
Norway	No	NA	60.0672	10.2283	29	Per AVID
Poland	Pol_Bli	Blizyn	51.2414	20.7463	21	Jan KOWALSKY, Dzmitry KAHAN
	Pol_Zwi	Zwierzyniec	50.6549	22.5575	13	Jan KOWALSKY
Romania	Ro_Bra	Prejmer Forest, Brasov County	45.6500	25.6000	8	Lucian CURTUS

	Ro_Lavar	La Vârfuri	46.5196	24.5033	19	Lucian CURTUS
Spain	Es_norte	NA	38.0771	-2.7501	10	Carlos HERRERA
	Es_sur	Grenade	37.1414	-3.2045	7	Francisco DONAIRE SANCHEZ
Great Britain	Gbr	Scotland	55.5485	-2.9058	22	Stephen CAVERS
Ukraine	Ukn	Kharkiv	49.9481	36.1323	13	Roman VOLANSYANCHUK
***Malus domestica* (N=339)**						
Australia	NA	NA	NA	NA	3	INRA[3]
Belgium	NA	NA	NA	NA	4	INRA[3], CRA - W[1]
Canada	NA	NA	NA	NA	1	INRA[3]
France	NA	NA	NA	NA	227	INRA[3], CRA - W[1], Abbaye de Beauport[5]
Germany	NA	NA	NA	NA	4	INRA[3], CRA - W[1]
Great Britain	NA	NA	NA	NA	10	INRA[3], CRA - W[1]
Ireland	NA	NA	NA	NA	1	USDA-ARS[4]
Israel	NA	NA	NA	NA	1	INRA[3]
Japan	NA	NA	NA	NA	2	INRA[3]
Netherland	NA	NA	NA	NA	5	INRA[3], USDA-ARS[4]
New Zeland	NA	NA	NA	NA	2	INRA[3]
Russia	NA	NA	NA	NA	7	INRA[3], USDA-ARS[4]
Spain	NA	NA	NA	NA	2	INRA[3]
Switzerland	NA	NA	NA	NA	2	INRA[3]
Tunisia	NA	NA	NA	NA	2	INRA[3]
Ukraine	NA	NA	NA	NA	3	INRA[3], CRA - W[1]
USA	NA	NA	NA	NA	11	INRA[3]
Denmark	NA	NA	NA	NA	24	Einar FLENSTED-JENSEN
Unknown	NA	NA	NA	NA	28	INRA[3], CRA-W[1], USDA-ARS[4]

*For the total dataset for individuals identified a priori as *Malus sylvestris* (N=1,889), 77 genotypes were found to correspond to misidentified *Malus domestica* trees. These individuals were not presented in this table. *N*: Number of trees sampled; *NA*: non assigned, no information; [1] CRA –W: Centre wallons de Recherches Agronomiques, Belgium; [2]IVLO – PLANT: Institute for agricultural and Fisheries Research Plant Unit, Melle, Belgium; [3]INRA: Institut National de la Recherche Agronomique, France; [4]USDA – ARS: United States Departement of Agriculture - Agricultural Rsearch Service, Cornell University, USA; [5]Abbaye de Beauport, Paimpol, France - Conservatory Orchards of ancient apple varities; [6]EMR: East Malling Researh, Kent, UK

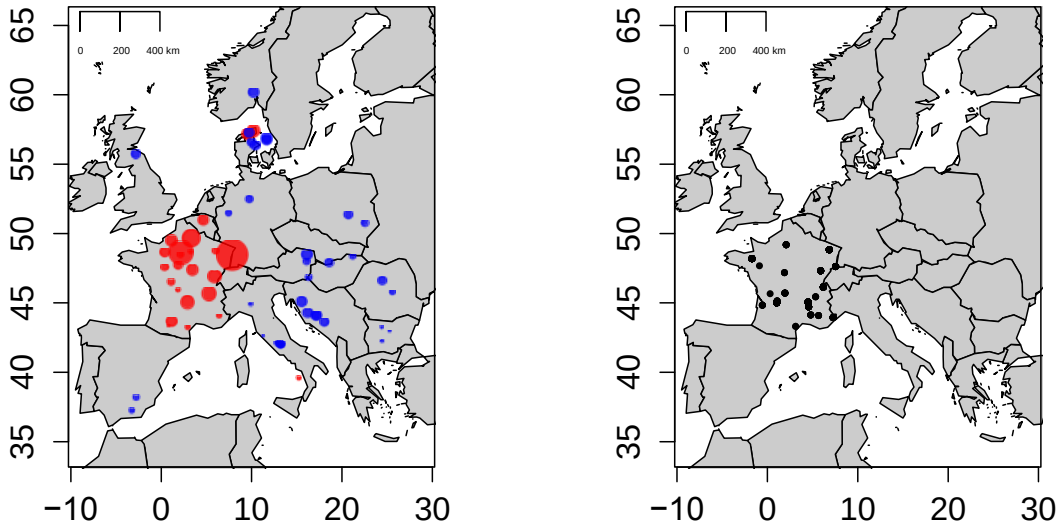


Figure S1 Geographical locations of the 66 sites across Europe (*N*=1,889 *Malus sylvestris* trees) used for genetic analyses (left). Blue: sites from previous studies (A. Cornille, Gladieux, & Giraud, 2013; A Cornille et al., 2013) and red: new sites included in this study. The circles are proportional to the size of the sample. Geographical locations of the 20 sites across France used for testing the factors affecting crop-to-wild gene flow are shown (right) (Deguines, Julliard, de Flores, & Fontaine, 2012).

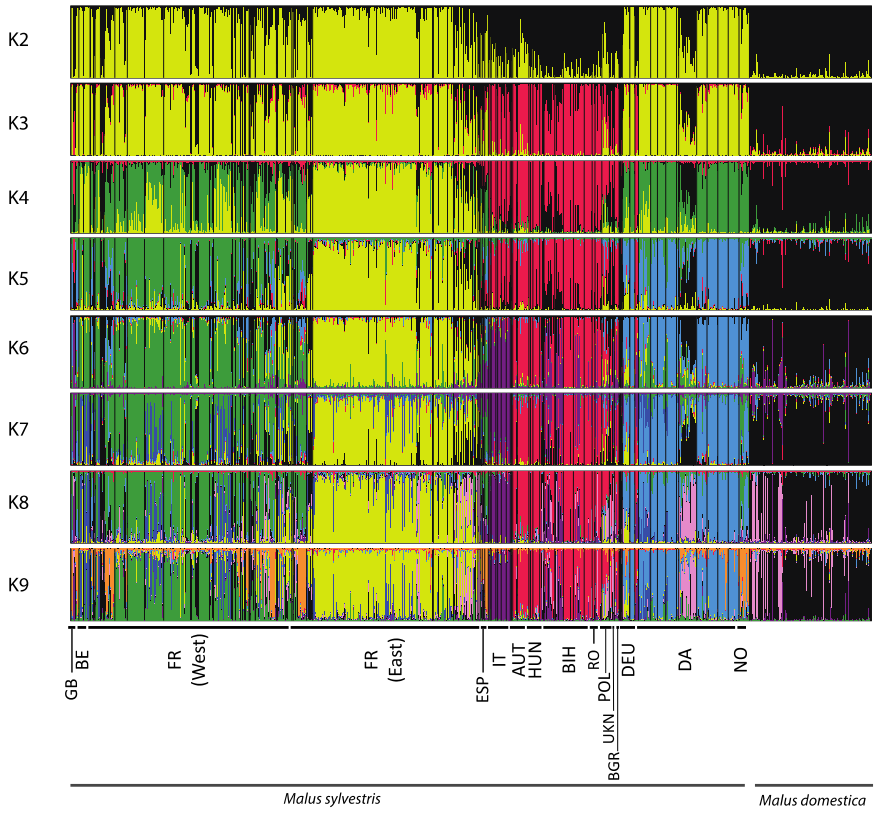


Figure S2 Bayesian clustering results obtained with STRUCTURE, for *K*=2 to *K*=9, based on the dataset including *Malus domestica* (*N*=339) and *Malus sylvestris* (*N*=1,889). Each individual is represented by a vertical bar, partitioned into *K* segments representing the proportions of ancestry of its genome in *K* clusters. Major clustering solutions ("modes") are shown. GB: Great Britain, BE: Belgium, FR (West): Western France, FR (East): Eastern France, ESP: Spain, IT: Italy, AUT: Austria, HUN: Hungary, BIH: Bosnia Herzegovina, RO: Romania, POL: Poland, BGR: Bulgaria, UKN: Ukraine, DEU: Germany, DA: Denmark, NO: Norway.

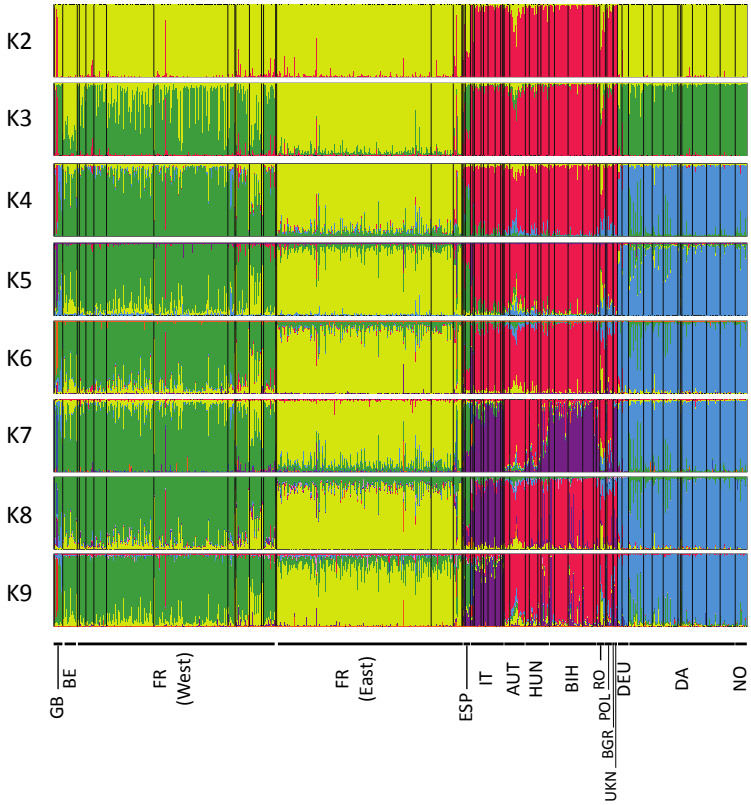


Figure S3 Bayesian clustering results for *Malus sylvestris* in Europe (*N*=1,376), obtained with TESS, for K_{max}=2 to K_{max}=9. Each individual is represented by a vertical bar, partitioned into *K* segments representing the proportions of ancestry of its genome in *K* clusters. Major clustering solutions ("modes") are shown. For clarity, the 37 sites are grouped by country. GB: Great Britain, BE: Belgium, FR (West): Western France, FR (East): Eastern France, ESP: Spain, IT: Italy, AUT: Austria, HUN: Hungary, BIH: Bosnia Herzegovina, RO: Romania, POL: Poland, BGR: Bulgaria, UKN: Ukraine, DEU: Germany, DA: Denmark, NO: Norway.

Figure S4 Estimated number of populations (K) in *Malus sylvestris,* from TESS analyses based on the DIC.

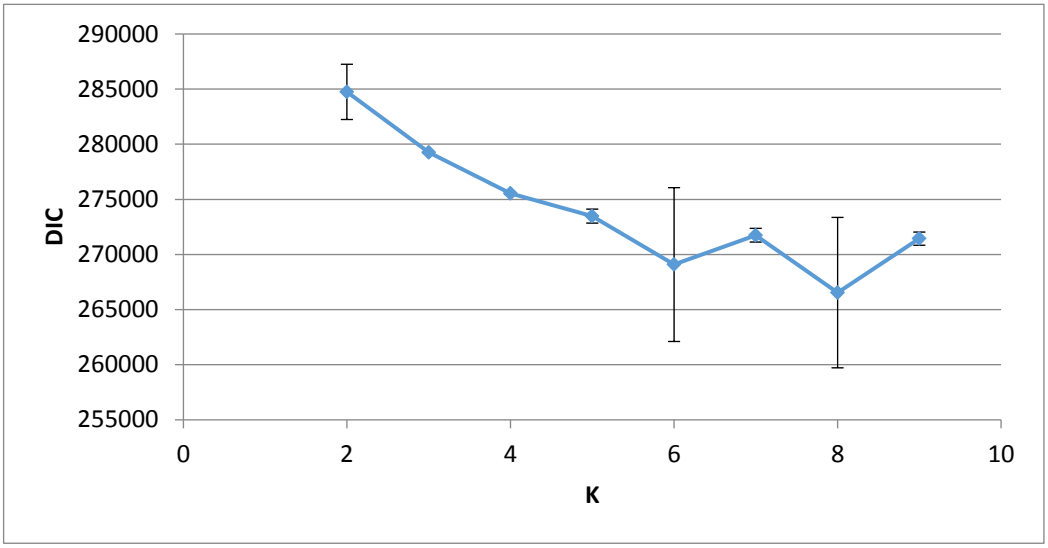


Table S2 Mean rates of introgression into *Malus sylvestris* from *Malus domestica* and number of pure *Malus sylvestris* trees and hybrids in the European sample (66 sites).

Site	N_{IND}	$P_{DOM->SYL}$	N_{PURE}	N_{HYB}
Alsace	333	0.035	307	26
Aquitaine	1	0.215	0	1
Aut_hern	12	0.058	11	1
Aut_stock	32	0.013	31	1
Auvergne	56	0.207	27	29
BasNorm	28	0.156	16	12
Bgr_Lov	2	0.590	0	2
Bgr_Plo	2	0.767	0	2
Bgr_tria	1	0.044	1	0
Bih_jab	30	0.149	16	14
Bih_kuk	22	0.192	11	11
Bih_Liv	28	0.021	27	1
Bih_Olo	30	0.019	29	1
Bih_Pre	22	0.016	21	1
Bourg	38	0.196	24	14
Bretagne	11	0.335	4	7
Centre	20	0.102	14	6
ChampArd	4	0.016	4	0
Da_Aaby	30	0.012	29	1
Da_he	29	0.065	21	8
Da_Hvolg	34	0.034	30	4
Da_k	30	0.019	28	2
Da_ka	29	0.015	27	2
Da_no	30	0.012	30	0
Da_Norha	6	0.088	5	1
Da_Nymo	50	0.497	3	47
Da_Oxholm	19	0.027	17	2
Da_Skeel	23	0.022	22	1

Deu_pop1	21	0.176	9	12
Deu_pop2	14	0.066	13	1
Deu_pop3	10	0.164	8	2
Es_norte	10	0.022	10	0
Es_sur	7	0.177	3	4
fland	30	0.012	29	1
FranCom	53	0.095	44	9
Gbr	22	0.118	17	5
HteNorm	35	0.135	25	10
Hun_p1	10	0.141	8	2
Hun_p2	19	0.037	17	2
Hun_p3	11	0.275	6	5
IleFrance	195	0.128	147	48
It_Cerva	14	0.014	14	0
It_Fil	4	0.031	4	0
It_Pag	2	0.008	2	0
It_parme	3	0.008	3	0
It_rome	5	0.017	5	0
It_Subia	9	0.007	9	0
It_Tosc	1	0.012	1	0
It_Trevi	11	0.012	11	0
It_valle	13	0.011	13	0
Italie	5	0.007	5	0
LangRous	6	0.217	3	3
Limousin	5	0.553	0	5
Lorraine	12	0.454	2	10
MidiPy	33	0.157	24	9
No	29	0.050	23	6
PaysLoi	15	0.084	13	2
Picardie	100	0.029	94	6
PoitChar	13	0.426	2	11

Pol_Bli	21	0.184	11	10
Pol_Zwi	13	0.445	3	10
ProACA	6	0.473	1	5
RhonAlpes	63	0.355	17	46
Ro_Bra	8	0.053	6	2
Ro_Lavar	19	0.179	7	12
Ukn	13	0.035	12	1
Total	**1,812**	**0.112**	**1,376**	**436**

N_{IND}: sample size for the site; $P_{DOM->SYL}$: proportion of admixture between the *M. domestica* and *M. sylvestris* gene pools; N_{PURE}: *M. sylvestris* individuals with a cumulative membership coefficient for the *M. sylvestris* gene pool >0.9; N_{HYB}: *M. sylvestris* individuals with a cumulative membership coefficient for *M. sylvestris* <0.9.

Table S3 Mean wild-to-crop introgression rates by country

Country	*Malus sylvestris* gene pool	*Malus domestica* gene pool	*N*
Germany	0.11	0.89	4
Great Britain	0.13	0.87	10
Australia	0.19	0.81	3
Belgium	0.14	0.86	4
Canada	0.02	0.98	1
Scotland	0.04	0.96	1
Spain	0.16	0.84	2
France	0.11	0.89	221
Ireland	0.09	0.91	1
Israel	0.41	0.59	1
Japan	0.04	0.96	2
NA	0.13	0.87	33
New_Zealand	0.02	0.98	2
The Netherlands	0.08	0.92	4
Russia	0.58	0.42	7
Switzerland	0.01	0.99	1
Switzerland/Germany	0.02	0.98	1
Sweden	0.22	0.78	24
Tunisia	0.35	0.65	2
Ukraine	0.31	0.69	3
USA	0.03	0.97	11
Total	0.15	0.85	338

N: sample size for each country. *NA:* cultivars of unknown origin.

Table S4 Genetic variation at each site, for *Malus sylvestris* (*N*=1,376 individuals, 62 sites)

Site	*N*	H_O	H_E	F_{IS}	A_R
ProACA	1	-	-	-	-
RhonAlpes	17	0.72	0.85	0.15	1.81
Picardie	94	0.77	0.86	0.11	1.77
Auvergne	27	0.76	0.86	0.12	1.78
ChampArd	4	0.82	0.84	0.027 NS	1.77
MidiPy	24	0.77	0.86	0.11	1.80
Alsace	307	0.76	0.86	0.11	1.79
BasNorm	16	0.78	0.85	0.09	1.78
PoitChar	2	-	-	-	-
Centre	14	0.76	0.82	0.08	1.78
Bourg	24	0.74	0.85	0.13	1.79
FranCom	44	0.73	0.83	0.12	1.76
IleFrance	147	0.74	0.84	0.12	1.77
HteNorm	25	0.73	0.80	0.08	1.73
Bretagne	4	0.66	0.80	0.17	1.77
LangRous	3	-	-	-	-
PaysLoi	13	0.76	0.89	0.15	1.79
Lorraine	2	-	-	-	-
fland	29	0.71	0.83	0.14	1.74
Da_Aaby	29	0.73	0.83	0.12	1.69
Da_Hvolg	30	0.74	0.85	0.13	1.69
Da_Norha	5	0.78	0.81	0.0331 NS	1.69
Da_Nymo	3	-	-	-	-
Da_Oxholm	17	0.71	0.78	0.08	1.67
Da_Skeel	22	0.70	0.76	0.08	1.67
Italie	5	0.69	0.81	0.15	-
Aut_hern	11	0.78	0.89	0.13	1.83
Aut_stock	31	0.74	0.86	0.14	-
Bih_jab	16	0.78	0.85	0.09	1.88

Bih_kuk	11	0.68	0.86	0.21	1.85
Bih_liv	27	0.70	0.80	0.12	1.86
Bih_olo	29	0.71	0.80	0.10	1.86
Bih_pre	21	0.71	0.78	0.09	1.86
Bgr_tria	1	-	-	-	-
Da_he	21	0.72	0.78	0.07	1.72
Da_k	28	0.67	0.74	0.09	1.69
Da_ka	27	0.75	0.86	0.12	1.69
Da_no	30	0.70	0.85	0.17	1.69
Deu_pop1	9	0.74	0.77	0.0481 NS	1.66
Deu_pop2	13	0.75	0.84	0.11	1.60
Deu_pop3	8	0.72	0.78	0.07	-
Gbr	17	0.71	0.83	0.15	1.81
Hun_p1	8	0.74	0.83	0.10	-
Hun_p2	17	0.68	0.79	0.13	1.84
Hun_p3	6	0.58	0.68	0.15	-
No	23	0.71	0.80	0.10	1.74
Pol_Bli	11	0.70	0.76	0.07	-
Pol_Zwi	3	-	-	-	-
Ro_Bra	6	0.65	0.66	0.0139 NS	-
Ro_Lavar	7	0.67	0.69	0.0205 NS	1.88
Es_sur	3	-	-	-	-
Es_norte	10	0.69	0.75	0.08	1.79
Ukn	12	0.75	0.88	0.14	1.84
It_parme	3	-	-	-	-
It_valle	13	0.64	0.68	0.06	1.83
It_rome	5	0.58	0.68	0.14	-
It_Subia	9	0.68	0.68	**-0.01 NS**	1.81
It_Cerva	14	0.68	0.77	0.12	1.83
It_Trevi	11	0.71	0.82	0.14	1.85
It_Fil	4	0.68	0.74	0.0877 NS	1.85

It_Tosc	1	-	-	-	-
It_Pag	2	-	-	-	-
Overall	1376	0.73	0.83	0.12	-

N: sample size for each site. H_O and H_E: observed and expected heterozygosity, respectively. F_{IS}: inbreeding coefficient. A_R: mean allelic richness across loci corrected by the rarefaction method. Overall. *11* sites with *N*<4 (ProACA. PoitChar. LangRous. Lorraine. Da_Nymo. Bgr_tria. Pol_Zwi. Es_sur. It_parme. It_Tosc and It_Pag) were excluded from these analyses. All F_{IS} values were significant (P<0.001) unless specifically annotated as NS (non-significant).

Table S5 Taxa diversity of insects visiting flowers of the apples in France

Order	Taxa diversity	Intensively	Less intensively
Araneida	8	0	8
Coleoptera	40	4	36
Diptera	109	25	84
Hemiptera	8	1	7
Hymenoptera	132	29	103
Lepidoptera	9	0	9
ALL	306	59	247

Intensively: intensively cultivated and managed areas; *Less intensively*: less intensively cultivated and managed areas; *ALL*: number of observations for all orders considered together.

References

Cornille A, Giraud T, Bellard C, Tellier A, Le Cam B, Smulders MJM, Kleinschmit J, Roldan-Ruiz I, Gladieux P. 2013a. Post-glacial recolonization history of the European crabapple (*Malus sylvestris* Mill.), a wild contributor to the domesticated apple. *Molecular Ecology* **22**: 2249–63.

Cornille A, Gladieux P, Giraud T. 2013b. Crop-to-wild gene flow and spatial genetic structure in the wild closest relatives of the cultivated apple. *Evolutionary Applications* **6**: 737–748.

Deguines N, Julliard R, de Flores M, Fontaine C. 2012. The Whereabouts of Flower Visitors: Contrasting Land-Use Preferences Revealed by a Country-Wide Survey Based on Citizen Science. *PLoS ONE* **7**: e45822.

Article 2. Flux de gènes des espèces cultivées aux espèces sauvages et ses conséquences sur la valeur sélective d'un fruitier sauvage : vers une stratégie globale de conservation du pommier sauvage en Europe

Après avoir estimé les niveaux d'hybridations historiques à partir des arbres adultes dans l'article précédent, un des buts pour ce second article était d'étudier les taux d'hybridations contemporains dans un environnement non agricole et donc relativement protégé des flux de gènes avec le pommier cultivé via l'étude génétique des graines dans une forêt. Afin de comprendre l'influence des hybridations sur le statut conservatoire du pommier sauvage, il était aussi indispensable de s'intéresser à la valeur sélective des hybrides, ce que nous avons fait en comparant les premiers stades de développement de plants de pommiers hybrides et non hybrides.

J'ai également cherché à étudier des aspects plus appliqués de la conservation du pommier sauvage. Pour cela, nous avons d'abord analysé les réseaux de ventes de graines de pommiers sauvages en France (et en Europe de l'Ouest via les distributeurs privés qui vendent souvent dans plusieurs pays). En effet, les stratégies de conservation *in situ* impliquent fréquemment des réintroductions d'individus, comme celles réalisées par l'association AFAF (association française d'agroforesterie). Il s'agit alors de s'assurer que les individus en question soient choisis correctement, notamment qu'ils soient des représentants génétiquement « purs » de l'espèce et non des hybrides, mais aussi que la diversité génétique réintroduite ne soit pas trop faible. Dans l'optique de créer un futur verger conservatoire pour *M. sylvestris* et de formuler des recommandations concernant son emplacement et la densité en arbres, nous avons également étudié la dispersion du pollen, qui au début de ma thèse n'avait été étudiée qu'à une petite échelle chez le pommier sauvage européen (Larsen & Kjaer, 2009).

Ma contribution à cet article comprend une participation à l'échantillonnage, aux étapes en laboratoire et à la vérification manuelle des génotypes (étapes pour lesquelles j'ai eu l'aide de deux stagiaires que j'ai co-encadrés avec Tatiana Giraud). Avec l'aide d'Amandine Cornille et de Tatiana Giraux pour la création du projet initial concernant la partie dispersion, j'ai développé le reste des questions scientifiques et mis en place les analyses et expériences nécessaires au projet. J'ai également réalisé l'ensemble des analyses bio-informatiques et statistiques et rédigé l'article. Une version révisée de cet article a été soumise à *Evolutionary Applications* après une recommendation de révisions mineures.

Crop-to-wild gene flow and its fitness consequences for a wild fruit tree: towards a comprehensive conservation strategy of the wild apple in Europe.

Feurtey Alice, Cornille Amandine, Shykoff Jacqui A., Snirc Alodie, Giraud Tatiana

Introduction

Interspecific hybridization can have important evolutionary consequences, especially in plants, by fostering adaptation or on the contrary introducing ill-suited alleles into populations. Gene flow between cultivated species and their wild relatives is of particular interest because it can be frequent (Ellstrand et al. 2013) and can impact the adaptive potential of both wild and domesticated species. Introgression from wild species into their cultivated relatives can contribute to variety improvement. For instance, hybridization between maize and the wild teosinte *Zea mays* spp. *mexicana* is likely to have contributed to crop resistance to the environmental conditions of the Mexican highlands (Hufford *et al.* 2013). On the other hand, hybridization between cultivated species and their wild relatives may have negative economic and ecological consequences. For example, hybrids growing in cultivated fields with their domesticated parents can reduce crop yield, e.g. in beets and carrots (Ellstrand *et al.* 2013). Hybridization from cultivated to wild populations can threaten the persistence of wild taxa, through either i) demographic swamping, in which hybrids with reduced fitness decrease population growth rates, or ii) genetic swamping, in which hybrids with moderate to high fitness replace pure wild genotypes over time (Wolf, Takebayashi & Rieseberg 2001; Todesco *et al.* 2016). It has long been assumed that crop-wild hybrids would have a lower fitness than their wild parent in natural habitats as has been shown, for instance, in sunflowers (Gutierrez, Cantamutto & Poverene 2011). However, in many other cases hybrids showed no decrease in fitness, for instance in Sorghum (Arriola & Ellstrand 1997). Some hybrids even display increased fitness compared to their parent species, e.g. in lettuce (Hooftman *et al.* 2007) and radishes (Hovick *et al.* 2012). Assessing the threat posed by hybridization to the conservation status and future of wild species facing global changes thus requires not only estimating the degree of gene flow, but also evaluating the relative fitness of hybrids.

While crop-to-wild hybridization has been intensively studied in annual crops such as wheat (Arrigo *et al.* 2011) and rice (Pusadee *et al.* 2013), its evolutionary consequence is still

poorly known in perennial plant species such as fruit trees. Yet, tree species in general have high dispersal capacities and weak interspecific barriers (Petit & Hampe 2006), two life-history traits promoting hybridization, and wild tree populations can be threatened by introgression from cultivated trees. In this context, various conservation management methods are used, with advantages and drawbacks for each method. *In situ* approaches (i.e. conservation of natural populations in their native areas) are generally the preferred biodiversity management method for threatened forest trees, as they allow the continued evolution of populations and help maintain genetic diversity (Koskela *et al.* 2013). *In situ* conservation can also involve reintroducing individuals into areas from which the species has disappeared to create bridges among isolated populations. These individuals can be planted in hedges or in agroforestry programs such as the ones carried out in France by the French Agroforestry Association (called hereafter *AFAF* for "*Association Française d'Agroforesterie*"). These conservation practices require the use of reliable pure and well-identified wild seed source populations. The main issue is then to find reliable production networks for seeds, free from inter-specific hybridization. One can harvest seeds from protected natural populations without introgressions. However, great effort is needed to find and harvest seed from enough wild trees to avoid a loss of genetic diversity in the replanted populations. Alternatively, seed orchards, i.e. a dynamic *ex situ* conservation approach (Balsemin & Collin 2004), allow harvesting large numbers of non-hybrid seeds easily. However, knowledge of pollen dispersal capacities and the factors affecting them is crucial for the creation of a seed orchard that would be isolated from crop gene flow in order to obtain seeds without introgressions.

One of the most widely cultivated fruit trees in the world is the apple tree *Malus domestica*. The genus *Malus* also includes ca. 30 wild species (Robinson, Harris & Juniper 2001), but its taxonomy is still unclear, in part due to weak interspecific barriers and also because outdated Latin names are still being used. For instance, *M. communis* is still used though it refers both to *M. domestica* and crabapples. Several crabapple species occur in areas with widespread cultivation of apples, including *M. sylvestris* in Europe, an emblematic tree of European forests. One major threat for *M. sylvestris* is the disappearance of its preferred habitats, wood margins and hedges, in intensively cultivated areas. Another major threat is frequent introgression from the cultivated apple tree *M. domestica* (Coart *et al.* 2006; Larsen *et al.* 2006; Gross *et al.* 2012; Cornille, Gladieux & Giraud 2013). Despite these threats, the conservation status of *M. sylvestris* is still unclear: while it is considered endangered in some European countries (e.g. Germany), the IUCN red list of threatened species indicates it as "data deficient" because the consequences of its hybridizations with *M. domestica* are still unknown (http://www.iucnredlist.org/). Such

introgression is facilitated by a lack of interspecific reproductive barriers, by the self-incompatible reproductive system that favours outcrossing, and by the widespread cultivation of apples (Cornille *et al.* 2015). A previous study revealed high frequency of hybridization and crop-to-wild introgression in *M. sylvestris* populations (Cornille *et al.* 2015). An earlier study (Larsen *et al.* 2006), based on few F1 genotypes, suggested that hybrids may germinate earlier than pure *M. sylvestris* seeds. Studying the frequencies and consequences of *M. domestica*-to-*M. sylvestris* introgressions is particularly timely and critical in light of the recent introduction of genetically modified apples, the "Arctic Apples" (Okanagan Specialty Fruits, Inc., www.fda.gov).

Our aim here was therefore to address the question of the conservation of the wild European crabapple in the face of introgression from *M. domestica*. More specifically, we wanted to 1) assess current hybridization rates, hybrid fitness and the suitability of current seed networks and 2) provide recommendations for creating seed orchards protected from introgression, by investigating pollen dispersal ability. Two previous studies (Larsen & Kjaer 2009; Reim *et al.* 2015) have reported pollen dispersal curves in European crabapples, but without studying dispersal in the light of hybridization or fitness differences between *M. sylvestris* and *M. domestica*. To investigate these issues, we chose the Dourdan forest situated 40 km southwest of Paris (France) as our study area because it contains the single seed orchard of putative *M. sylvestris* trees from which the French National Forest Agency (*ONF* hereafter, "*Office National des Forêts*") collects all *M. sylvestris* seeds that they use and supply to agroforestery programs. We also analysed crabapple seeds sold by seed companies and seeds used by the *AFAF*, a French non-governmental organization planting *M. sylvestris* trees, among other species, to reintroduce native trees into agrosystems.

We used microsatellite markers to genotype seeds from companies, adult trees and seeds from the Dourdan forest, and we grew seeds in a greenhouse, for addressing the following specific questions: 1) Are seeds sold and grown as crabapple seeds (from private companies, from the *ONF* seed orchard in the Dourdan forest and from the AFAF) free from introgression? 2) What is the frequency of introgression of the crabapples from the Dourdan forest, both in adult trees and seeds? 3) What is the impact of the degree of *M. domestica* ancestry level on crabapple seed fitness? In particular, what is the impact of the degree of *M. domestica* ancestry level of the mother tree on the germination of its seeds? Further, what is the impact of the degree of *M. domestica* ancestry level of each seed on its germination ability, i.e. its germination speed and the growth rate of the resulting seedling? 4) What are the pollen dispersal capacities of apple trees in a forest? Is this dispersal impacted by tree density and by the degree of *M. domestica*

ancestry level? What are the pollen dispersal capacities of apple trees in the Dourdan seed orchard?

Materials and methods

Sampling

Our sampling site, the Dourdan forest and seed orchard, is situated in France, covering 14 km^2. We collected leaves from 163 adult apple trees (mostly putative *M. sylvestris* but also a few *M. domestica* and putative hybrids) in the Dourdan forest and recorded their geographic coordinates (except for nine of them that were provided at a later stage by the Dourdan forest ranger). This sampling comprises all the apple trees known by the forest ranger in the area and therefore well represents the wild European crabapples population of this forest. We also collected leaves from the seed orchard in which *M. domestica* stocks were planted and grafted with branches of *M. sylvestris*. Since some of these grafted trees had been insufficiently pruned, we were able to collect leaves from the stock (*N*=6) and grafted (*N*=14) branches. We also collected leaves from three stock trees on which the grafts did not take. Among the 186 sampled apple trees, 112 bore fruits in the fall 2013 and 2014 when we sampled. We collected the apples on the tree when possible or around the trunk where no confusion was possible regarding the mother tree. We collected all the fruits we could find and reach, up to 73 apples per tree. We measured the diameter of the apples and the mean number of seeds they contained.

We also collected leaf samples from trees planted by the AFAF between 2010 and 2015 (*N*=12) and from seedlings obtained from seeds sold by commercial seed companies (*N*=13). The French seed company *Vilmorin* sells seeds of *"M. communis"*, the Belgian *Sluyza* company sells seeds of *"M. communis (sylvestris)"* and *Forestart* in the United Kingdom sells seeds of *"M. sylvestris"*. The Dutch *Kuiper* company sells seeds whose names oddly associate the wild species name with that of *M. domestica* cultivars (*M. sylvestris* "A2", *M. sylvestris* "Antonovka" and *M. sylvestris* "Bittenfelder"). We obtained a few samples from seedlings originating from three of these seed companies: *Vilmorin "M. communis"* (*N*=7), Sluyza *"M. communis (sylvestris)"* (*N*=2) and *Forestar*t, *"M. sylvestris"* (*N*=4). We could not obtain samples from the Kuiper seed stocks.

Seed emergence and seedling growth rates

In a first experiment, for fitness analyses as a function of mother genotype, seeds were stratified in the dark for three months at 5°C in a damp mix of sand and vermiculite (ratio 2:1).

Then 30-60 seeds per mother (or all the seeds collected when fewer than 30 seeds were available, giving *N*=3,927 in total) were planted in soil in a greenhouse with controlled temperature (between 20 and 25°C) and constant day length (16h), and kept moist. Germination was checked two months after planting. For analysis of pollen dispersal distance, leaves from 311 seedlings were sampled to be genotyped (from 28 mother-trees, 3 to 13 per mother, with a mean of 11 seedlings per mother).

In a second experiment, for fitness analyses as a function of both seed and mother genotypes, 344 seeds (7-14 seeds for 24 mothers) were stratified for two months at 5°C and were then moved in a greenhouse with controlled temperature (between 20 and 25°C) and constant day length (16h). Emergence of the seedlings was checked every three or four days and the height of each seedling and number of leaves were recorded. When seedlings were more than 1.5 cm tall on the first day observed, we considered they had emerged the previous day. When the seedlings had more than seven leaves, we collected leaves for genotyping (*N*=138). After two months, we also retrieved as many non-germinated seeds from the soil as possible (*N*=72). The genotypes obtained through this second experiment were also used for the pollen dispersal distance analysis.

Genotyping

We extracted DNA from leaves from 186 adult trees, 12 samples from the *AFAF*, 13 samples from commercial seeds, 449 seedlings issued from seeds collected from the Dourdan forest and orchard and from 172 seeds (including 100 seeds coming from the orchard and 72 non-germinated seeds retrieved after the second experiment described above). Seeds were dissected in order to keep only the embryo. Both leaves and embryos were frozen at -80°C before grinding with metal beads. DNA was then extracted with the Nucleo Spin plant DNA extraction kit II® (Macherey & Nagel). Multiplex microsatellite PCR amplifications were performed with a Multiplex PCR Kit® (QIAGEN, Inc.). We used 33 microsatellite markers spread across the 17 chromosomes using 10 different multiplex reactions as previously described (Cornille *et al.* 2012). The genotyping was done at the *Gentyane* platform (INRA). Only the microsatellite markers and individuals with less than 50% missing data were used for the rest of the analyses. The dataset was then composed of the genotypes for 30 microsatellite markers from 186 adult samples from Dourdan, 12 samples from the *AFAF*, 13 samples from commercial seeds, 442 from the Dourdan forest and orchard, and 167 seeds (including 99 seeds from the orchard and 68 non-germinated seeds) retrieved after the second experiment described above. We used the software Cervus (Marshall *et al.* 1998) to check for identical genotypes in the adult tree dataset.

Paternity and dispersal analyses

Paternity analyses were performed on the seedlings of the greenhouse experiments (*N*=442), from the seeds from the orchard (*N*=99) and from the non-germinated seeds retrieved after the second experiment described above (*N*=68). We used Cervus (Marshall *et al.* 1998) with standard settings to match the genotype of offspring and mother-tree in order to check both the identification of mother-trees and the validity of markers. We only kept the microsatellite markers with less than 3% of incongruences between the genotypes of mother-trees and their offspring. Any seed showing more than two mismatches from the genotypes expected given the mother's genotype were removed from the dispersal analyses. We thus retained 25 markers and 583 offspring genotypes. More than 80% of mismatches were found for homozygous loci and were thus probably due to null alleles (i.e. failed amplification of one allele). We then used Cervus to identify the father for each offspring using a paternity analysis. We used a two-mismatch threshold between tree and seed genotypes and a three-mismatch threshold when considering both parent trees and seed genotypes.

Hybridization analyses

In order to identify the hybrid status of the samples, we used reference genotypes from our collections previously identified as pure *M. domestica* (*N*=42) and pure *M. sylvestris* (*N*=31) (i.e. varieties assigned at a minimum of 90% to the appropriate genepool using STRUCTURE analyses, Cornille *et al.* 2012). In order to prevent any artefact in assignments due to population structure in *M. sylvestris* at the European scale, we used as references only *M. sylvestris* genotypes from regions belonging to the same genetic cluster as Dourdan (i.e. France and Belgium, Cornille *et al.* 2015).

For each genotype, we estimated a hybrid index with the add-on R-package *introgress* (Gompert & Buerkle 2010). This hybrid index is an estimate of the proportion of alleles that were inherited from one of two hybridizing parental populations, i.e. the level of ancestry from one of two populations. In our case, pure *M. domestica* genotypes had a "hybrid index" of one while pure *M. sylvestris* genotypes had a hybrid index of zero. We therefore call this "hybrid index" P_{dom} as it represents the proportion of assignment to the *M. domestica* genepool.

Genetic diversity and statistical analyses

The allelic richness A_R was estimated with ADZE (Szpiech, Jakobsson & Rosenberg 2008), a software implementing a rarefaction method in order to analyse allelic diversity across different populations while correcting for sample size differences.

To compare genetic diversity (A_R) between forest and orchard trees across markers, we used a signed-rank test on the differences in this A_R. To assess the impact of *M. domestica* ancestry level of the mother tree on its fitness and that of its progeny, several phenotypic traits related to fitness (apple diameter, number of seeds per fruit and germination rate) were regressed (linear regression) against the P_{dom} value of the mother. We averaged the measures for the duplicated genotypes of mother trees in the orchard (i.e. multiple grafting of the same clone), in order to avoid pseudo-replication. We tested whether trees that did or did not bear fruit differed for their P_{dom} with a t-test. To assess the impact of the P_{dom} of seeds on their germination ability (in the second experiment), we used a paired t-test to compare the mean P_{dom} of germinated *versus* non-germinated seeds from each mother. To evaluate the impact of *M. domestica* ancestry level of the seeds on their fitness, we calculated the growth rate and the rate of increase in leaf number over the time period where these rates were linear (*i.e.* when they were not significantly non-linear, which covered the period between days 2 and 29 after moving the seeds to the greenhouse for growth and between days 3 and 40 for leaf number). In order to avoid pseudo-replication due to the presence of multiple half-siblings (sharing their mother), we analysed the means of these traits per mother while including the sampling year as a blocking factor (which had no significant effect), retaining only the means done on at least three germinated seeds. We then regressed germination timing, growth rate and rate of increase in leaf number against the P_{dom} of the seedlings. A mixed-model multiple linear regression analysis, with the mother tree's identity as a random blocking factor, was run to investigate effects of tree density and P_{dom} of the parent trees on pollination distances. All of the statistical analyses were performed using the software JMP v7 (SAS Institute), using arcsin-transformed values for P_{dom} and rates to improve normality.

To check for a correlation between geographic distances between trees and similarity in their introgression levels, we calculated the difference in P_{dom} for each pair of trees. We then used a Mantel test to compare the introgression difference matrix with that of the pairwise geographic distance using the R package *ncf* (Bjornstad 2013).

Results

Hybridization levels

All samples from commercial seed companies that were sold as wild crabapples had substantial levels of *M. domestica* ancestry: $0.65 \leq P_{dom} \leq 0.98$ for *Vilmorin*, $0.35 \leq P_{dom} \leq 0.36$ for *Sluyza* and $0.76 \leq P_{dom} \leq 1$ for *Forestart* samples. Of the 12 trees planted by the *AFAF* that we analysed, five could be considered as pure *M. sylvestris* ($P_{dom} < 0.2$) and two were assigned to *M. domestica* ($P_{dom} > 0.8$); the others had a mean P_{dom} of 0.3.

In the seed orchard, only 7% of the genotyped seeds had a $P_{dom} > 0.2$. The orchard was composed of stock trees on which *M. sylvestris* branches had been grafted. However, the trees had been insufficiently pruned, bearing ungrafted *M. domestica* branches with flowers. Of the 17 trees in the orchard, three no longer bore grafted *M. sylvestris* branches at all. Among the 14 orchard trees with grafted branches, only three different *M. sylvestris* genotypes were found. These three genotypes had low *M. domestica* ancestry (two with $P_{dom} < 0.1$ and the third with $P_{dom} = 0.16$). Two trees with genotypes identical to those of the grafts were identified elsewhere in the forest, suggesting that these may have been two of the three graft-donor *M. sylvestris*

Figure 1: Spatial distribution of apple trees in the study area. Trees are represented by dots and their colour indicates their degree of Malus domestica ancestry (Pdom): green, Pdom < 0.2; red, Pdom > 0.8; and yellow, intermediate. The seed orchard is circled in black.

trees. The trees without grafted branches and the stock branches collected (N=6) had all the same, triploid *M. domestica* genotype.

The *M. sylvestris* genetic diversity was significantly lower in the orchard than in the forest, both for the adult trees (mean allelic richness across markers A_R=7.02 in the forest and A_R=4.06 in the orchards; t=13.09, df=24, p<0.0001) and for the seeds (A_R=7.06 in the forest and A_R=3.96 in the orchards; t=9.96, df=24, p<0.0001).

Estimates of the levels of *M. domestica* ancestry of the adult trees growing in the Dourdan forest (N=163) showed that 70% belonged to *M. sylvestris* (P_{dom}<0.2), 7% to *M. domestica* (P_{dom}>0.8) and 23% were hybrids (Figure 1). The *M. domestica* ancestry level of the seeds coming from *M. sylvestris* mothers (i.e. mothers with P_{dom}<0.2) was significantly higher than that of their mothers in the forest, but not in the orchard (paired t-test with the means per mother; in the forest: t=2.57, df=16, p=0.0032; in the orchard: t=0.30, df=11, p=0.6772; Figure 2).

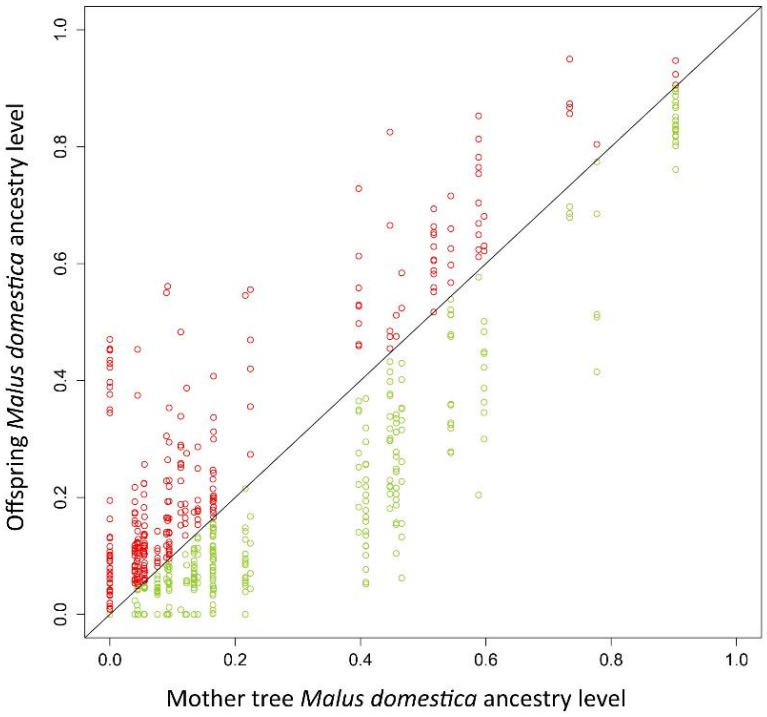


Figure 2: Level of Malus domestica ancestry in mother trees and in their progeny (602 progenies from 43 mothers sampled in the Dourdan forest and orchard). The line represents equal M. domestica ancestry between mother and seeds. Circles are coloured in red above the line and in green underneath.

Fitness of seeds as a function of hybridization level

In the first fitness experiment, we examined variation in fruit abundance and size, seed number and seed germination as a function of variation in hybridization level of mother trees. We detected no significant difference in mean *M. domestica* ancestry between trees with and without fruits (P_{dom}=0.24 and P_{dom}=0.23, respectively; t=0.19, df=158, p=0.8456). Trees with a higher level of *M. domestica* ancestry bore larger fruits (N=1,679 apples from 90 mothers, using mean fruit diameter from mothers with at least five measurable fruits, averaged over identical genotypes, N=77; $F_{(1,75)}$=62.43, slope ± SE=1.63±0.21; r^2=0.45; p<0.0001). However, these larger fruits contained fewer seeds, the mean number of seeds per fruit per mother decreasing with the level of *M. domestica* ancestry (using mean seed number per fruit per genotype from the mothers with at least five collected apples; N=2,160 apples; 11,481 seeds from 99 mothers; 87 genotypes; $F_{(1,85)}$=11.79; slope ± SE=-2.03±0.59; r^2=0.12; p=0.0009). Overall, 62% of the seeds germinated and the success of seed germination increased slightly but significantly with increasing mother *M. domestica* ancestry level (N=3,836 seeds, from 101 mothers with more than five planted seeds, germination rates being averaged across identical genotypes, N=87; $F_{(1,85)}$=8.99; slope ± SE=0.22±0.07; r^2=0.10; p=0.0036). On average, the germination success was 69% for seeds from *M. domestica* trees (0.8 <P_{dom}< 1), 57% for seeds from *M. sylvestris* trees (0 <P_{dom}<0.2), and 72% for seeds from hybrids (0.2<P_{dom}<0.8). Fitting a polynomial (degree 2) instead of a line did not improve the fit.

The second experiment was designed to check if germination ability and growth were influenced by the hybridization levels of the seeds themselves, in addition to that of their mothers, by individually genotyping and monitoring seedling emergence, number of leaves and growth. We successfully genotyped 138 germinated seeds and 68 non-germinated seeds from 24 different mother trees. For the 22 mothers, for which we had both germinated and non-germinated seeds, a paired t-test revealed no difference in their average hybridization levels (t=0.74, df=21, p=0.47). For the seedlings, we also recorded germination timing (i.e. how rapidly the seeds germinated), their rates of growth and of leaf number increase. Of the germinated seeds, maternal siblings with a higher *M. domestica* ancestry (P_{dom}) emerged significantly earlier than those with a lower P_{dom} ($F_{(1,18)}$=5.28, slope ± SE=-4.88±2.12, r^2=0.31, p=0.0338, Figure 3) and gave rise to seedlings that grew more rapidly ($F_{(1,18)}$=5.42, slope ± SE=0.11±0.05, r^2=0.25, p=0.0318, Figure 4). In contrast, the rate of leaf appearance was not significantly correlated with P_{dom} ($F_{(1,18)}$=2.18, slope ± SE=-0.04±0.02, r^2=0.11, p=0.1619). We also checked for a correlation between heterozygosity and growth rate, rate of leaf number increase and germination timing

(*i.e.* checking for heterosis or hybrid vigour), but we found no significant correlations (data not shown).

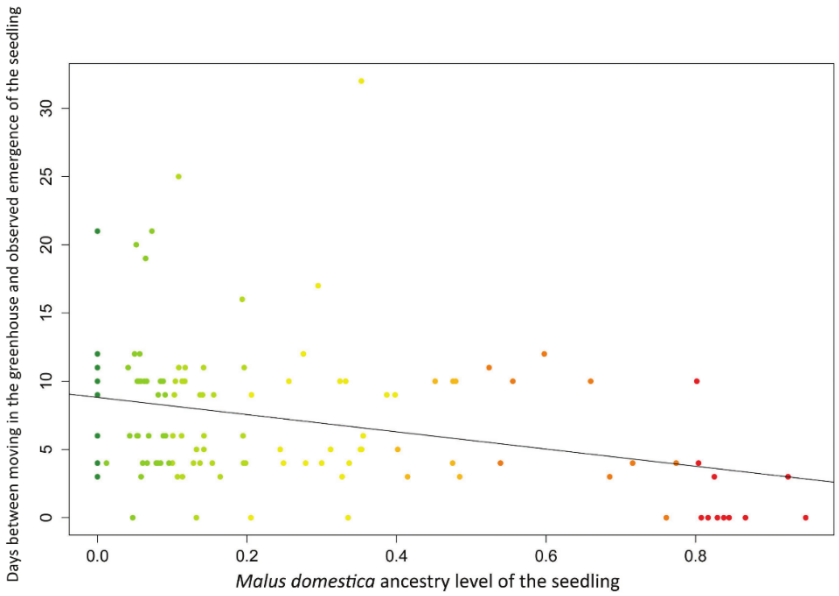


Figure 3: Days between moving the seeds to the greenhouse and observed emergence plotted against the level of Malus domestica ancestry of the seedlings (128 seedlings from 24 mothers sampled in the Dourdan forest and orchard). Colours indicate the level of Malus domestica *ancestry (red:* Malus domestica*, green:* Malus sylvestris *and shades of yellow/orange: hybrids). The line represents the linear regression.*

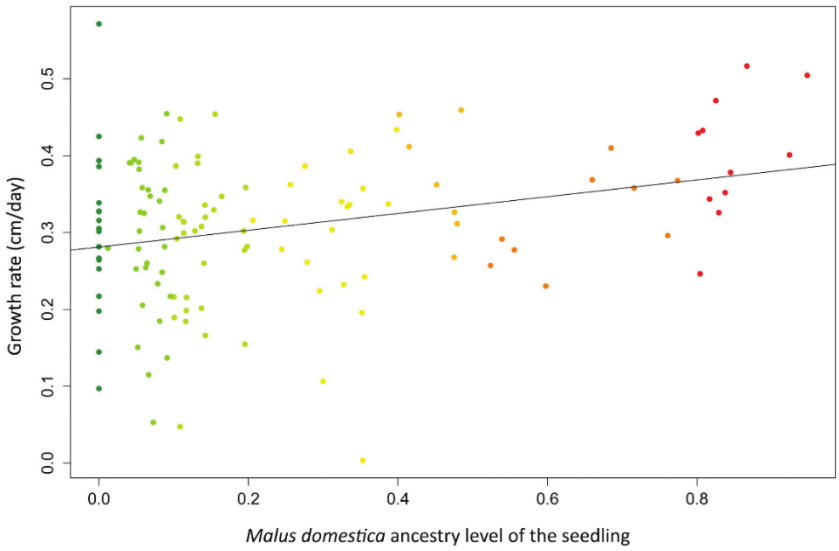


Figure 4: Growth rate of seedlings plotted against their level of Malus domestica ancestry (127 seedlings from 24 mothers sampled in the Dourdan forest and orchard). The colours indicate the level of Malus domestica ancestry (red: M. domestica, green: M. sylvestris *and shades of yellow/orange: hybrids). The line represents the linear regression.*

Pollen dispersal

A paternity analysis was performed on 583 offspring genotypes to identify adult tree genotypes compatible with the offspring genotype as possible pollen donors. We studied pollen dispersal and paternity separately for the forest and the seed orchard, because of its grafting history.

Among the 119 analysed offspring from the orchard, in no case was the putative male parent the *M. domestica* stock genotype. More than 90% of the seeds were sired by another of the three grafted *M. sylvestris* genotypes and only nine resulted from pollination events external to the orchard. For the four of these for which we could identify an unambiguous father in the forest, identified fathers stood from 132 to 167 m away from the mother tree.

For the forest seed samples, a male parent tree could be identified without ambiguity for 287 seeds from 30 mothers. For 26 seeds, several putative fathers were identified. We could identify no compatible father for 151 seeds. Half of these seeds came from only five mothers that had more than 80% of their offspring without identified fathers (three of them were located at the margins of our study area). Overall, we could only detect 10 seeds with genotypes consistent with self-pollination events (i.e. 1.7% of the total number of seed genotypes analysed).

The distance between the mother trees and the unambiguously detected male parent trees could be as large as 4 km (Figure 5). However, half of the pollination events for which we could identify the father occurred within less than 42 m, 25% at less than 15 m and 75% at less

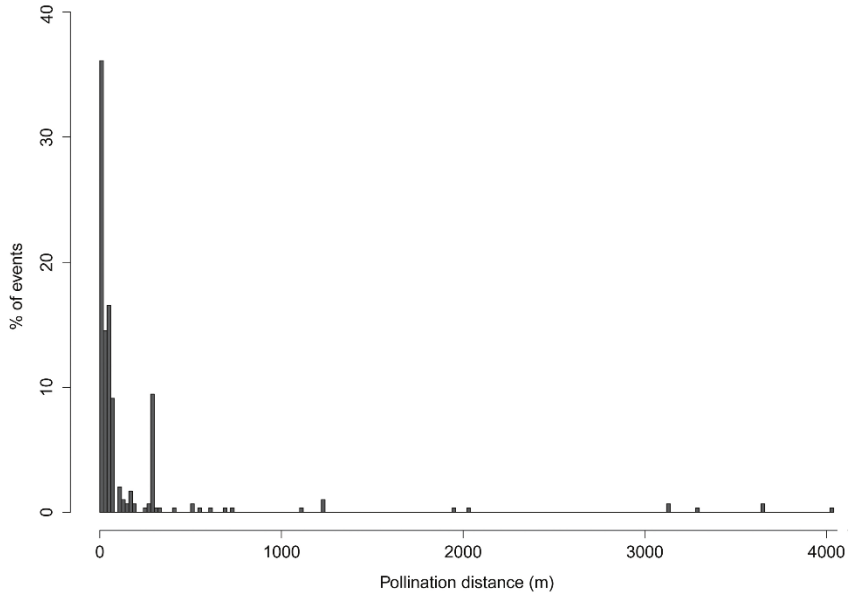


Figure 5: Distribution of the distance of pollination events (286 pollination events with a single putative father with known coordinates and from the 30 mothers in the forest) in the Malus trees of the Dourdan forest.

than 101 m. Altogether, although the vast majority of pollination events occurred at short distances, nearly 5% occurred between trees more than 1 km apart. A multiple regression analysis showed that pollination distances were significantly affected by apple tree density around parent trees (density being estimated as the number of apple trees within a radius of 100 m) and by the introgression level of the putative father (N=286, r^2=0.40, Table 1). Mothers in denser patches, as expected, had pollen donors that were nearer. The pollination distances increased with decreasing *M. domestica* ancestry in fathers.

Table 1 : Mixed-model multiple linear regression analysis investigating effects of tree density and degree of ancestry in Malus domestica (Pdom) of the parent trees on pollination distances (286 pollination events with a single putative father with known coordinates and from the 30 mothers in the forest; ndf: numerator degrees of freedom; ddf: denominator degrees of freedom).

Effect	ndf, ddf	F-ratio	p-value
Tree density around mothers	1, 24.3	15.35	0.0015
Tree density around fathers	1, 264.8	10.36	0.0023
Malus domestica ancestry of mothers	1, 15.4	2.21	0.1571
Malus domestica ancestry of fathers	1, 278.7	5.96	0.0153

We also tested for assortative mating by level of *M. domestica* ancestry. The mean P_{dom} value of seed sires was significantly and positively correlated with that of the mother (test on the means per mother, for the 27 trees from the forest with more than three pollination events with an identified putative father: $F_{(1,25)}$=10.20, *slope* ± SE=0.51±0.16, r^2=0.29; p=0.0038). We checked if this could result from spatial variation in introgression levels, i.e. if trees with similar levels of *M. domestica* ancestry were spatially clustered. We thus first calculated pairwise differences between levels of *M. domestica* ancestry among trees. This difference equals zero when two trees have similar levels of *M. domestica* ancestry and one when one is a pure *M. domestica* and the other a pure *M. sylvestris*. A Mantel test detected no significant correlation between pairwise geographic distance and this pairwise introgression difference index (r=-0.02, p=0.1309), indicating a lack of spatial structure of introgression levels.

Discussion

Crop-to-wild hybridization and introgression are major threats to endangered wild populations. Historical crop-to-wild gene flow has been previously investigated in the European crabapple (Cornille *et al.* 2015). Here, we aimed at assessing contemporary inter-specific gene flow as well as its fitness consequences for the wild crabapple. We also tested whether seeds sold or sampled as wild crabapples were free from introgression in order to provide recommendations for building orchards for growing *M. sylvestris* seeds, in particular by assessing pollen dispersal curves in forests with *M. sylvestris*, *M. domestica* and hybrid pollen.

Dispersal capacities are influenced by tree densities and introgression rates: consequences for future conservation programs

The pollen dispersal curve was leptokurtic and fat-tailed, i.e. with most pollination events occurring at short distances but with a substantial number of long-distance pollination events. Here, almost 5% occurred farther than 1 km. The occurrence of relatively frequent long-distance pollination events means that even *M. sylvestris* trees distant by several kilometres from any cultivated apple tree may still produce hybrid seeds. The pollen dispersal curve shape found here was consistent with the findings of previous studies on *M. sylvestris* (Larsen & Kjaer 2009; Reim *et al.* 2015), but the dispersal distances varied: the maximum dispersal distance was around 300 m and 11 km in the two previous studies, respectively, and 4 km in our study. This variation may be explained by our finding that the pollen dispersal curve depends on tree density, as well as on the spatial scale of the study area.

Our results also showed that pollination distances decreased with the density of trees from the same species around the mother tree, which is in agreement with previous findings (Reim *et al.* 2015). Although this result is intuitive (pollinators fly shorter distances when there are more trees to forage from), it is of crucial importance from a conservation perspective. First, it highlights a further detrimental effect of habitat fragmentation on a species like *M. sylvestris*: more long-distance pollination events potentially mean more pollination by *M. domestica* growing in gardens or orchards surrounding a forest or hedgerow. Second, this result helps in formulating guidelines for planting and maintaining seed orchards. High density of orchards will guard them against external pollen flow, effectively safeguarding the genetic pool of the seeds. In fact, in the orchard in the Dourdan forest fewer than 10% of pollinations were external to the orchard itself and, of the cases for which the pollen donor could be identified, involved neighbouring trees.

We also found that the level of *M. domestica* ancestry influenced pollination in two ways. First, pollination distances decreased with the level of *M. domestica* ancestry of the father tree: the "wilder" trees had a higher pollen dispersal distance than more "domesticated" apple trees. Second, there was significant assortative mating regarding the level of *M. domestica* ancestry. Assortative mating is unlikely to be due to physical clumping of mother trees since we found no correlation between similarity in introgression level and physical distance. Assortative mating is most likely explained by temporal barriers to gene flow between cultivated and wild apples. Indeed, a previous study reported only partial overlap in flowering time between cultivated and wild apple trees, with peak flowering time differing by 1 to 2 weeks (Schnitzler *et al.* 2014).

A positive effect of introgression on fitness?

The question of the fitness of hybrids in crop-wild complexes is relevant from both basic and applied perspectives. The consequences of crop-to-wild introgression have been studied in different annual crops such as maize, rice or lettuce (Ellstrand *et al.* 2013); however, to our knowledge, this is the first time that such a study has been conducted on a perennial fruit tree. From a conservation perspective, it is important to know whether hybrids display lower or higher fitness than pure wild apple trees. In our first experiment, germination was higher for seeds from introgressed and *M. domestica* mothers though the effect was not significant in our second experiment, likely because the number of seeds was smaller and the overall germination rate lower (60% *vs.* 40%). Seeds with a larger *M. domestica* ancestry germinated earlier than wild seeds, as had been suggested in a previous study (Larsen *et al.* 2006), using only five *M. domestica* cultivars and five *M. sylvestris* genotypes. We also found that seedlings with a higher *M. domestica* ancestry grew faster than wild seeds. Overall, fitness at early life stages thus increased with the level of *M. domestica* ancestry. Although our experiments focused on the early developmental stages only and were done in a greenhouse, these results suggest that hybrids probably suffer no reduced fitness in a natural environment either, at these stages. Therefore, hybrid seedlings can compete favourably with *M. sylvestris* seedlings where they occur naturally and may, over time, invade natural populations, if they display no disadvantages at later stages. This would explain the high levels of hybrids found here in the Dourdan forest and in a previous study all across Europe (Cornille *et al.* 2015).

Current crabapple seed network suffers from introgression, low genetic diversity and misidentified stocks

Our investigation of the French network providing *M. sylvestris* trees and seeds revealed several worrying aspects. First there is an ambiguity about what species is sold, due to the use of an obsolete species name. Indeed, the name "*M. communis*" seems to be used either as a synonym for *M. sylvestris*, or as a way to name apple trees coming from uncertain ancestry or even for known cultivated variety names. *Vilmorin* does not explicitly sell their "*M. communis*" as a "wild" species but nor does it call them "*M. domestica*". *Sluyza* also uses an ambiguous name, "*M. communis* (*sylvestris)*". *Kuiper* even uses "*M. sylvestris*" associated with the names "Antonovka" and "Bittenfelder" which are well known cultivated varieties. Every seed we could genotype coming from these companies had a *M. domestica* ancestry level of more than 0.3 (and very often more than 0.9), regardless of the name used. This means that buying genuine *M. sylvestris* seeds is extremely difficult if not impossible in the private sector. Indeed, identification of pure *M. sylvestris* individuals free from *M. domestica* pollination is quite difficult without genetic analyses, as attested by the high proportion of hybrids planted in agroforestry. In the French governmental *ONF* seed orchard, the trees from which the grafts were sourced belonged mostly to *M. sylvestris*. However, since we could only detect three distinct genotypes, the genetic pool of the orchard is very restricted, representing a very low sampling of the genetic diversity in *M. sylvestris.* This may lead to problems of inbreeding and poor adaptive potential if only seeds from this orchard are used to reintroduce the species into forests, as is the case currently.

Conclusion

Here, we have assessed the European crabapple pollen flow in the forest of Dourdan. Pollination distance decreased with increasing tree density and varied with *M. domestica* ancestry, with assortative mating probably due to some difference in flowering time between *M. sylvestris* and *M. domestica*. In addition, we have shown that substantial current crop-to-wild gene flow occurs in apple tree populations in natural habitats and that *M. domestica* trees and hybrids showed higher fitness than the wild apple tree at the early growth stages investigated. Overall, our results suggest increasing threats for wild apple integrity. Therefore, populations of European wild apples need to be protected by reducing interspecific gene flow and replanting genotypes free from introgressions. However, we have shown that the seed network feeding the programs aiming at reintegrating the wild apple into forests and agrosystems suffers from poor

genetic diversity and/or introgressions, and even misidentification of the seed species. These results altogether set a baseline for concrete conservation programs of the wild apple. First, more *M. sylvestris* seed orchards should be set up, with high genetic diversity and protection from cultivated apple pollen, the latter being facilitated by planting stands of high local density far from cultivated apples. The trees used for producing source seeds for reintroduction schemes should be free of *M. domestica* introgressions, which can only be verified based on genotyping at several loci. A best first approach would be to take them from several forests previously identified as free from introgression (Cornille *et al.* 2015). However, care should also be taken to not mix too different genotypes, possibly locally adapted. In fact, a previous study has shown a strong population subdivision of *M. sylvestris* at the European scale (Cornille *et al.* 2015). These lineages should be kept separate in different orchards and replanted in their region of origin.

Acknowledgements

We declare no conflict of interest. We thank the Région Ile de France (PICRI and DIM R2DS), IDEEV, and laboratoire d'Excellence BASC for funding, and the Plateforme de Génotypage GENTYANE INRA UMR 1095, Douae BenHDech and William Fontanaud, for assistance with genotyping and fitness measures. We thank all those who helped with sampling (Table S1), in particular Didier Desseaux (ONF). We also thank Maud Tenaillon, Pierre Gladieux, Eric Collin, Séverin Lavoyer and Charly Chauveau for help during the project.

References

Arrigo, N., Guadagnuolo, R., Lappe, S., Pasche, S., Parisod, C. & Felber, F. (2011) Gene flow between wheat and wild relatives: empirical evidence from Aegilops geniculata, Ae. neglecta and Ae. triuncialis. *Evolutionary Applications,* 4, 685-695.

Arriola, P.E. & Ellstrand, N.C. (1997) Fitness of interspecific hybrids in the genus Sorghum: Persistence of crop genes in wild populations. *Ecological Applications,* 7, 512-518.

Balsemin, E. & Collin, E. (2004) Conservation in situ des ressources génétiques des arbres forestiers en France métropolitaine. *Ingénieries,* 40, 51-60.

Bjornstad, O.N. (2013) ncf: spatial nonparametric covariance functions. R package.

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.

Cornille, A., Feurtey, A., Gelin, U., Ropars, J., Misvanderbrugge, K., Gladieux, P. & Giraud, T. (2015) Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: a basis for conservation and breeding programs in apples. *Evolutionary Applications,* 8, 373-384.

Cornille, A., Gladieux, P. & Giraud, T. (2013) Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *Evolutionary Applications,* 6, 737-748.

Cornille, A., Gladieux, P., Smulders, M.J.M., Roldan-Ruiz, I., Laurens, F., Le Cam, B., Nersesyan, A., Clavel, J., Olonova, M., Feugey, L., Gabrielyan, I., Zhang, X.-G., Tenaillon, 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 Genetics,* 8.

Ellstrand, N.C., Meirmans, P., Rong, J., Bartsch, D., Ghosh, A., de Jong, T.J., Haccou, P., Lu, B.-R., Snow, A.A., Stewart, C.N., Jr., Strasburg, J.L., van Tienderen, P.H., Vrieling, K. & Hooftman, D. (2013) Introgression of Crop Alleles into Wild or Weedy Populations. *Annual Review of Ecology, Evolution, and Systematics, Vol 44* (ed. D.J. Futuyma), pp. 325-345.

Gompert, Z. & Buerkle, C.A. (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources,* 10, 378-384.

Gross, B.L., Henk, A.D., Forsline, P.L., Richards, C.M. & Volk, G.M. (2012) Identification of interspecific hybrids among domesticated apple and its wild relatives. *Tree Genetics & Genomes,* 8, 1223-1235.

Gutierrez, A., Cantamutto, M. & Poverene, M. (2011) Persistence of sunflower crop traits and fitness in Helianthus petiolaris populations. *Plant Biology,* 13, 821-830.

Hooftman, D.A.P., Jong, M.J.D., Oostermeijer, J.G.B. & Den Nijs, H.C.M. (2007) Modelling the long-term consequences of crop-wild relative hybridization: a case study using four generations of hybrids. *Journal of Applied Ecology,* 44, 1035-1045.

Hovick, S.M., Campbell, L.G., Snow, A.A. & Whitney, K.D. (2012) Hybridization Alters Early Life-History Traits and Increases Plant Colonization Success in a Novel Region. *American Naturalist,* 179, 192-203.

Hufford, M.B., Lubinksy, P., Pyhaejaervi, T., Devengenzo, M.T., Ellstrand, N.C. & Ross-Ibarra, J. (2013) The Genomic Signature of Crop-Wild Introgression in Maize. *Plos Genetics,* 9.

Koskela, J., Lefevre, F., Schueler, S., Kraigher, H., Olrik, D.C., Hubert, J., Longauer, R., Bozzano, M., Yrjana, L., Alizoti, P., Rotach, P., Vietto, L., Bordacs, S., Myking, T., Eysteinsson, T., Souvannavong, O., Fady, B., De Cuyper, B., Heinze, B., von Wuehlisch, G., Ducousso, A. &

Ditlevsen, B. (2013) Translating conservation genetics into management: Pan-European minimum requirements for dynamic conservation units of forest tree genetic diversity. *Biological Conservation,* 157, 39-49.

Larsen, A.S., Asmussen, C.B., Coart, E., Olrik, D.C. & Kjaer, E.D. (2006) Hybridization and genetic variation in Danish populations of European crab apple (Malus sylvestris). *Tree Genetics & Genomes,* 2, 86-97.

Larsen, A.S. & Kjaer, E.D. (2009) Pollen mediated gene flow in a native population of Malus sylvestris and its implications for contemporary gene conservation management. *Conservation Genetics,* 10, 1637-1646.

Marshall, T.C., Slate, J., Kruuk, L.E.B. & Pemberton, J.M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology,* 7, 639-655.

Petit, R.J. & Hampe, A. (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics*, pp. 187-214.

Pusadee, T., Schaal, B.A., Rerkasem, B. & Jamjod, S. (2013) Population structure of the primary gene pool of Oryza sativa in Thailand. *Genetic Resources and Crop Evolution,* 60, 335-353.

Reim, S., Proft, A., Heinz, S., Lochschmidt, F., Höfer, M., Tröber, U. & Wold, H. (2015) Pollen movement in a Malus sylvestris population and conclusions for conservation measures. *Plant Genetic Resources,* FirstView, 9.

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 Systematics and Evolution,* 226, 35-58.

Schnitzler, A., Arnold, C., Cornille, A., Bachmann, O. & Schnitzler, C. (2014) Wild European apple *(Malus sylvestris* (L.) Mill.) Population Dynamics: Insight from Genetics and Ecology in the Rhine Valley. Priorities for a Future Conservation Programme. *Plos One,* 9.

Szpiech, Z.A., Jakobsson, M. & Rosenberg, N.A. (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics,* 24, 2498-2504.

Todesco, M., Pascual, M.A., Owens, G.L., Ostevik, K.L., Moyers, B.T., Hübner, S., Heredia, S.M., Hahn, M.A., Caseys, C., Bock, D.G. & Rieseberg, L.H. (2016) Hybridization and extinction. *Evolutionary Applications*.

Wolf, D.E., Takebayashi, N. & Rieseberg, L.H. (2001) Predicting the risk of extinction through hybridization. *Conservation Biology,* 15, 1039-1053.

Article 3. Une histoire génomique de la domestication du pommier

Comme expliqué en introduction, l'importance des introgressions de *M. sylvestris* vers *M. domestica* dans l'histoire du pommier cultivé était encore en débat au début de ma thèse. Ainsi, des données nucléaires de microsatellites et des données chloroplastiques montraient une importance cruciale de ces flux de gènes (Cornille et al 2012, Nikiforova et al 2013) alors que des analyses à partir de 23 gènes nucléaires (Velasco et al 2010, mais dont l'interprétation a été débattue dans Nicola Harrison and Harrison 2011 et Micheletti et al 2011) indiquait le contraire. Dans l'étude de Velasco et al. (2010), aucune différenciation n'a été détectée entre les individus *M. domestica* et *M. sieversii*.

Ce débat ouvert nous a conduits à vouloir étudier plus en détail l'histoire du pommier cultivé, en ce qui concerne la structure génétique des pommiers cultivés et des pommiers sauvages européens et asiatiques et donc potentiellement l'histoire de domestication du pommier et l'importance respective des deux espèces sauvages dans la composition du génome de *M. domestica*.

Pour cela, nous avons obtenu les données de séquençage pour 32 variétés cultivées du consortium Fruitbreedomics. Nous avons également reséquencé les génomes complets de cinq individus *M. sylvestris* et de trois individus *M. sieversii*.

Le génome de la pomme a été séquencé et assemblé pour la première fois en 2010 à partir d'un individu diploïde de la variété Golden Delicious. Les pommiers cultivés sont extrêmement hétérozygotes ce qui pose de nombreux problèmes techniques pour l'assemblage du génome. Ceci explique probablement le fait que ce premier assemblage divisait les 750 Mb du génome en plus de 120 000 contigs (la troisième version de cet assemblage en comptant environ 90 000). Un second projet d'assemblage a été monté au cours de la dernière année de ma thèse par Etienne Bucher à l'INRA d'Angers. Ce projet se base sur le séquençage d'un individu double haploïde par la technologie Pacific Biosciences SMRT. L'utilisation d'un double haploïde permet de s'affranchir des problèmes posés par l'hétérozygotie et la technologie Pacific Biosciences SMRT permet d'obtenir des fragments d'ADN séquencé de grande longueur. La combinaison des deux devrait donc mener à un assemblage de bien meilleure qualité.

L'ébauche d'article présentée ici a été rédigée sur la base d'alignements des données de séquençage sur la version du génome de *M. domestica* assemblée à partir de données de séquençage par Illumina de la variété Golden Delicious (Velasco et al 2010). Or, travailler à partir d'un meilleur assemblage apporterait beaucoup dans l'étude présentée ici puisqu'un génome trop

fractionné limite le travail en fenêtres glissantes sur tout le génome afin d'identifier des régions précises d'introgressions de *M. domestica* par *M. sylvestris* et empêche de détecter d'éventelles grandes régions génomiques plus ou moins perméables aux flux de gènes. Nous avons donc décidé de refaire cette étude à l'aide du nouvel assemblage de référence réalisé par Etienne Bucher. C'est pourquoi les résultats présentés dans cet article ne sont que des résultats préliminaires ajoutés à cette thèse pour discussion.

En plus d'utiliser un nouvel assemblage de référence, nous avons également décidé d'inclure ce projet dans une collaboration avec l'équipe de Riccardo Velasco. Nous prévoyons de mener à bien deux projets jumeaux avec près de 70 génomes de variétés cultivées en plus des huit génomes de pommiers sauvages que nous avons reséquencés. Notre projet répondra aux questions que j'ai présentées ci-dessus alors que le projet dirigé par nos collaborateurs s'intéressera à la domestication du pommier sous un angle différent, en recherchant des régions du génome sous sélection chez le pommier cultivé.

Ma participation à ce projet a commencé avec la mise en place des questions scientifiques. J'ai participé à l'échantillonnage des individus sauvages et ai ensuite réalisé les extractions d'ADN préalables au séquençage. J'ai ensuite réalisé l'ensemble des analyses de bio-informatique détaillées ci-après. J'ai rédigé l'ébauche d'article présentée ici (comprenant également une discussion qui est uniquement en français dans ce manuscrit).

A genomic history of apple domestication

Feurtey Alice, Micheletti Diego, Cornille Amandine, Kerschbamer Emanuela, Bucher Etienne, Troggio Michela, Branca Antoine, Cestaro Alessandro, Tenaillon Maud, Velasco Ricardo & Giraud Tatiana

Introduction

Domestication is a slow and continuous process, it begins when organisms are free-living and humans begin investing in their survival or reproduction and it can reach a point where organisms cannot reproduce or survive without a large human investment (Gross and Olsen 2010). Crop domestication has been extensively studied in annual plants and especially in the "big three" (maize, rice and wheat). Only few studies have focused on perennials (i.e. plants living more than two years, such as fruit trees or grapes) (Goldschmidt 2013), despite the fact that perennials produce important resources as food, drink and oil staples and account for one-eigth of the world food-producing surface, and that the life history traits of perennials species can be very different from annuals (Gaut et al 2015). To name a few specific life history traits, perennial species often have long juvenile phases and are frequently obligate outcrossers, with high rates of intra and interspecific gene flow (Petit and Hampe 2006, Savolainen et al 2007), which can impact the process of domestication (Gaut et al 2015). For instance, while bottlenecks have been found frequently in domesticated annuals (with an average of 60% of neutral genetic diversity retained in domesticated species compared to their wild relatives), this does not seem to be the case in perennials, that retained an average of 95% of the wild populations' diversity (Miller and Gross 2011). Another life history trait impacting domestication is the clonal propagation of many perennial crops, especially woody plants (McKey et al 2010). Clonal propagation ensures that desired phenotypes are reproduced intact and limits wild-to-crop gene flow that could lower the economic value of the next crop generation. Clonal propagation is however quite recent in terms of evolutionary times and perennial plants often have outcrossing mating systems favoring gene flow, in contrast to mostly selfing domesticated annuals (Gaut et al 2015).

The outcrossing mating systems and weak inter-specific barriers found in perennial crops often promoted gene flow between domesticated forms and their wild relatives, which had non-negligible importance in the history of domestication of a number of species. Indeed, hybridizations have been at the origin of many domesticated species. It has been the case even

in annual species such as in the textbook wheat example (reviewed in Charmet 2011), with several successive hybridization events leading to the durum wheat, *Triticum turgidum* subsp. *durum*, and then the bread wheat, *Triticum aestivum*. Inter-specific ybridizations have been of even greater importance in the domestication of cultivated perennials. For instance, the majority of banana varieties, the most cultivated fruit in the world (according to the statistics of the Food and Agriculture Organization of the United Nations, FAO), are derived from crosses between the wild species *Musa acuminate* and *M. balbisiana* (Heslop-Harrison and Schwarzacher 2007). Another striking example is the *Citrus* genus in which different hybridization events gave rise to lemon, oranges and mandarins among others (Velasco and Licciardello 2014, Wu et al 2014).

Beyond the cultivation of hybrid forms, hybridizations can also impact the evolution of domesticated species through introgressions from wild relatives (Ellstrand et al 1999). This possibility of introgressions of wild alleles in cultivated crops has been used deliberately in breeding programs, in annuals such as rice, wheat and tomatoes, and in perennials such as bananas, peaches and grapes (Hajjar and Hodgkin 2007, Miller and Gross 2011). Adaptive introgressions have also occurred outside of deliberate breeding programs earlier in domestication history. For instance, Hufford et al (2013) showed that resistance of maize to Mexican highland harsh conditions may have originated from introgressions from wild teosintes.

Among important perennial crops, apples are the most cultivated fruit in temperate areas in the world and as such are economically crucial in these areas (FAO). Apples are also important culturally in many countries, as shown by their presence in religions, myths and fairy-tales. Surprisingly, considering their importance in temperate areas culture and economy, it is only recently that their history has been thoroughly studied using genetic markers. One of the first hints on the origin of cultivated apple trees, *Malus domestica*, came from Vavilov in the 20's who observed forests areas of wild apple trees in the Tian Shan Mountains. He suggested that this area was the center of origin of the cultivated apple tree and that the phenotypically diverse wild species growing there, *Malus sieversii*, was its progenitor (Harris et al 2002). This hypothesis has been recently supported by the sequencing of 23 nuclear genes, to the point that no genetic differentiation was detected between *M. sieversii* and cultivated individuals (Velasco et al 2010). Therefore, the authors even questioned the occurrence of an evolutionary process of domestication in the apple tree and suggested the use of a unique species name (*M. pumila*). However, microsatellite markers in contrast suggested that *M. domestica* and *M. sieversii* constituted two differentiated and well-identified genepools (Cornille et al 2012).

The story of the cultivated apple tree did not end at its supposed domestication in Central Asia a few thousand years ago (Zohary and Hopf 2000). The cultivated trees were then imported in Europe via the Silk Roads, thereby crossing the distribution range of several wild apple tree species, including *M. orientalis* in the Near and Middle East, and *M. sylvestris* in Europe (Cornille et al 2014). Microsatellite markers and chloroplast sequences have shown recently that gene flow occurred between these two wild species and *M. domestica* and that *M. sylvestris*, in particular, is an important secondary contributor to the chloroplast (Nikiforova et al 2013) and nuclear cultivated apple genomes (as detected with microsatellite markers in Cornille et al. 2012). However, the location in the nuclear genome and the extent of the introgressions is still subject to debate. Indeed, the results concerning nuclear data were obtained with only two dozens of microsatellite markers and stand in contrast to the analysis of 23 nuclear genes that did not reveal any introgressions between *M. domestica* and *M. sylvestris* (Harrison and Harrison 2011, Micheletti et al 2011, Velasco et al 2010). This discrepancy in the results obtained with different markers could be explained by the fact that microsatellite markers are often located in non-coding regions that might be affected by different evolutionary processes. Whole genome analysis is therefore required to elucidate the domestication history, by providing more power and access to different genomic regions under various types of selection.

In continuity with these studies of the apple tree domestication, a recent study has shed light on the genetic structure of *M. domestica* cultivated varieties across Europe using 16 microsatellite markers (Urrestarazu et al 2016). It has revealed the existence of three genetic groups corresponding to different geographic areas, grouping North-Western, South-Western and North-Eastern varieties, respectively.

To better understand the evolution of the cultivated apple tree, we used whole genome sequences from *M. domestica, M. sylvestris* and *M. sieversii* to address two main questions:

1. Is there a genome-wide genetic differentiation between *M. sieversii* and *M. domestica* indicative of a domestication process?
2. Can we detect introgressions from *M. sylvestris* into the genome of *M. domestica*? If we can, where are they localized in the genomes (i.e. in coding/non-coding regions, in similar loci in different varieties)?
3. Can we detect a genetic subdivision within *M. domestica* using whole genomes? If so, do these groups result from a difference in domestication history?

Material and methods

Sampling and sequencing

Of the five *M. sylvestris*, two came from Stockerau (Austria), one from the Fontainebleau forest (France), one from the Jura region (France) and the last one from the collection of the Angers INRA was collected from the Rambouillet forest (France). The three *M. sieversii* came from the Angers INRA collection, and they were sampled in the Tarbagataï moutains, in Kazakhstan and in the Djungarsky mountains.

The DNA from the Djungarsky mountains *M. sieversii* and from the Rambouillet forest *M. sylvestris* was extracted from fresh leaves using the Macherey-Nagel mini plant DNA extraction kit following their protocol after finely grinding the leaves with liquid nitrogen. The sequencing was done in paired-end with a read length of 100 bp at the Centre for Genomic Regulation in Barcelona (Spain) with a HiSeq Illumina sequencer. The sequencing depth for these samples was around 60X.

For the other wild samples, the leaves were dried in silica gel and stored at -80°C. DNA was extracted with the Macherey-Nagel midi plant DNA extraction kit following their protocol after finely grinding the leaves with liquid nitrogen. The sequencing was done in paired-end with a read length of 125 bp at the INRA Genotoul platform with a HiSeq Illumina sequencer. The sequencing depth for these samples was between 15 and 20X.

The sequencing reads data for 32 cultivated varieties was obtained through the FruitBreedomics consortium and our collaboration with Dr. Velasco's team (FEM, Trento, Italy). These were chosen randomly among the 70 apple tree sequencing reads that they currently have.

Reads were trimmed with Trimmomatic (Bolger et al 2014) to a length of 80 and a quality of 34 both at the beginning and end of each reads for the *M. domestica* genomes, to a length of 80 and a quality of 32 both at the beginning and end of each reads for the two wild genome with a deeper sequencing and to a length of 100 bp and a quality of 36 both at the beginning and end of each reads for the rest of the wild genomes. We only kept the reads that were still paired after trimming.

SNP calling and filtering

All genomes were mapped on the *M. domestica* genome assembly v3.0.a1 (available on rosaceae.org, Velasco et al 2010) with the algorithm mem from the software Bwa (Li and Durbin 2009). We then used to Picard v1.134 (http://broadinstitute.github.io/picard) and GATK v3.4-46

(McKenna et al 2010) to prepare the reads for the SNP calling, i.e. marking duplicate reads and performing local realignment for indels following GATK Best Practices (Van der Auwera et al 2013, DePristo et al 2011). The SNP calling was then done using a two-step process: first HaplotypeCaller was used independently for every sample and then a multi-sample callset was created by a joint genotyping step.

The variants obtained through these steps (N=38 861 598) were then filtered according to several criteria. First, we retained only the SNPs with a quality above the thresholds recommended for this type of SNP calling (namely FS < 60.0, MQ > 40.0 and QD < 2.0; N= 26,087,592). We filtered on read depth per site (no more than 1,700; N= 24,441,074) and per genotype (no less than 8; these genotypes were marked as missing data) and then on the amount of missing data (no more than 30%; N= 19,624,105). Finally, we only kept biallelic SNP (N= 17,986,962), the only ones that many analysis tools can take.

After preliminary analyses, it appeared that a tree identified as *M. sieversii* actually belonged to *M. domestica* and that a tree identified as *M. domestica* individual appeared as an outlier and may not belong to any of the three species studied here. For the analyses presented here, these two individuals were therefore removed. The second case will be studied more thoroughly after alignments on the new assembled genome are done.

Analyses

The population structure was investigated in two different ways. First, we used fastStructure (Raj et al 2014), an algorithm based on a Bayesian framework for posterior inference. We used a subsample of 500,000 SNPs randomly drawn SNPs after a filtering step eliminating all SNPs with a minor allele count smaller than four. For each value of K (indicating the maximum number of genetic clusters to be found), we did ten runs and used the software CLUMPP (Jakobsson and Rosenberg 2007) to identify multimodality.

We also used a non-model based approach through the use of principal component analysis (PCA) with the package R SNPRelate (Zheng et al 2012), using all the biallelic SNPs. F_{ST} were calculated per SNP with vcftools (Danecek et al 2011) and then averaged. In this analysis, the individuals kept for each species were the ones with a corresponding ancestry proportion of more than 0.99 (as calculated by fastStructure at K=3).

We used the R package introgress (Gompert and Buerkle 2010) in order to calculate a "hybrid index", corresponding to an estimate of the proportion of the genome inherited from *M. sylvestris*, that we therefore hereafter name P_{syl}. This P_{syl} index was estimated for each individual on sliding windows using the contigs larger than 10,000 bp (corresponding to ca. 60% of the

genome). The windows encompassed 5000 bp with a sliding step of 2500 bp. The individuals used as reference for the two parental genepools were those with an ancestry proportion of more than 0.99 for one of the two wild species as calculated by fastStructure at K=3. To prevent problems of ancestry detection due to shared ancestral polymorphisms, only the SNPs with a F_{ST} of more than 0.8 between the two wild species were kept (N=861,913). Introgress also gives the lower and upper limits of the 95% confidence interval for each P_{syl} estimate and we only kept the values for which this interval was below 0.3. The results in this manuscript are only preliminary since this analysis was performed for only 69,428 out of the 182,819 possible windows, due to time constraints. The statistical tests checking for correlations of differences in means between groups were done with JMP (SAS Institute).

Results

A model-based approach to elucidate population structure revealed both differentiation between the three *Malus* species and the existence of genetically differentiated clusters within *M. domestica*. In total, four genetic clusters were identified (Figure 1). At K=2, a first cluster separated *M. sylvestris* from all other *Malus* trees. A second cluster consisted of the individuals sampled as *M. sieversii* but also several cultivated varieties (with four varieties assigned entirely to this group). A third cluster encompassed the remaining of the *M. domestica* varieties. A fourth cluster appeared in minor modes and included only *M. domestica* varieties coming from North-Eastern Europe (data not shown).

We calculated F_{ST} values per position and between all pairs of species, keeping only the samples assigned to the three genetic groups with a posterior probability of at least 0.99 in the fastStructure analysis (K=3). The highest differentiation was found between *M. sieversii* and *M. sylvestris* genetic groups (i.e. including cultivated varieties assigned purely to these groups; mean F_{ST} = 0.22). The lowest differentiation was between *M. domestica* and *M. sieversii* (mean F_{ST} = 0.07). The differentiation between *M. domestica* and *M. sylvestris* had an intermediate value (mean F_{ST} = 0.14).

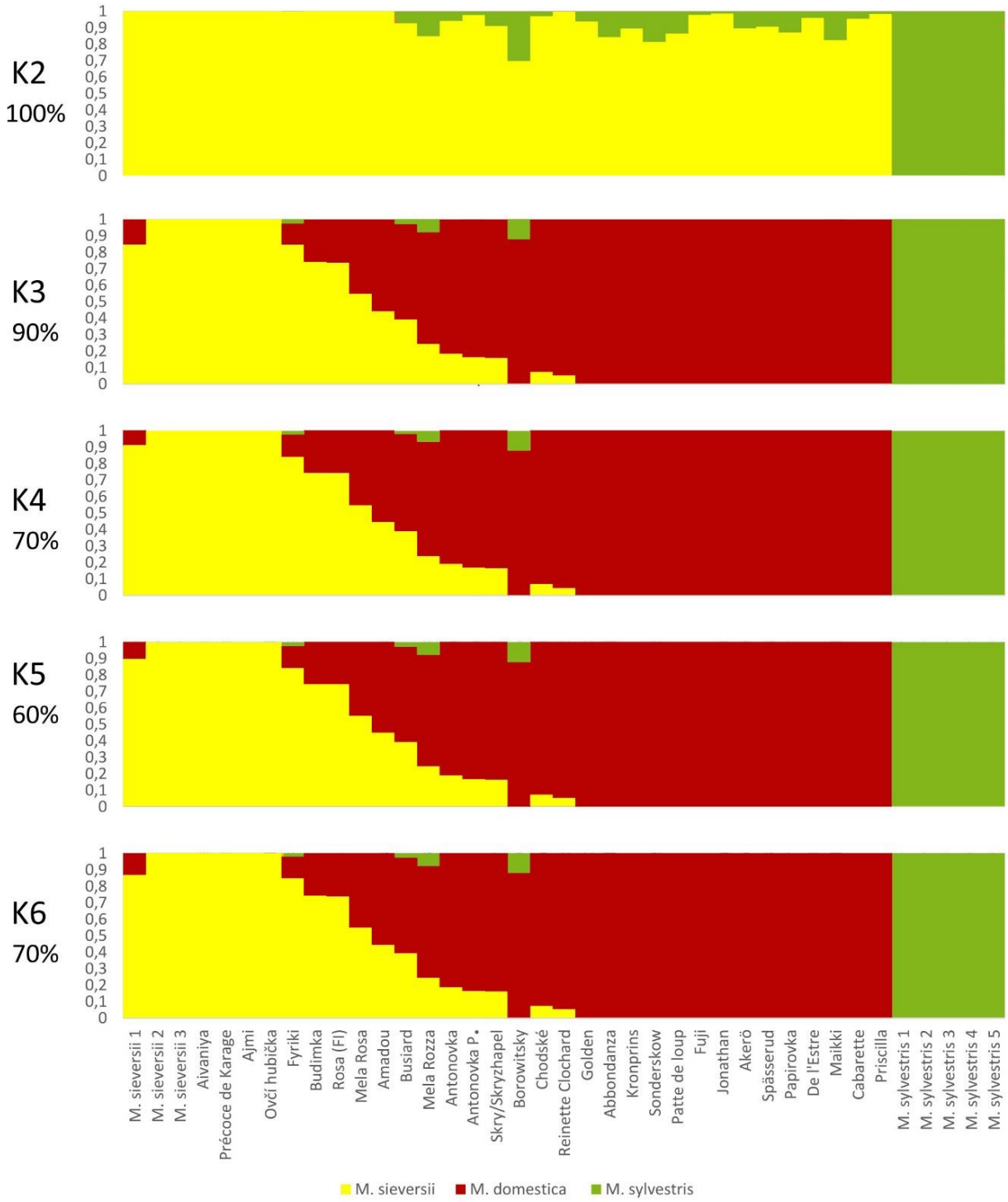


Figure 1: Ancestry proportion of genomes for the three Malus *species (*M. sylvestris*: N=5,* M. sieversii*: N=3 and* M. domes
*cultivated varieties: N=31) by fastStructure in K=2 to K=6 genetic clusters for each sample. Only major modes as found v
CLUMPP are presented here with the percentage of corresponding runs indicated at the left of each barplot. Each c
represents a differentiated genetic group and each vertical bar represents a genome, divided into K colored segme
representing its amount of ancestry proportion in K genetic groups.*

When removing the five *M. sylvestris* samples, the sub-structure in *M. domestica* was even clearer (Figure 2) and it was the major mode in the fastStructure analyses from K=3 upward. We also calculated F_{ST} values between pairs of genetic groups, keeping only the samples assigned to the genetic groups with a posterior probability of at least 0.9. The differentiation between the two *M. domestica* genetic groups was slightly lower (mean F_{ST} = 0.06), than between them and *M. sieversii* (mean F_{ST} = 0.8 for the two groups). The mean F_{ST} between the two *M. domestica* groups and *M. sylvestris* was 0.15.

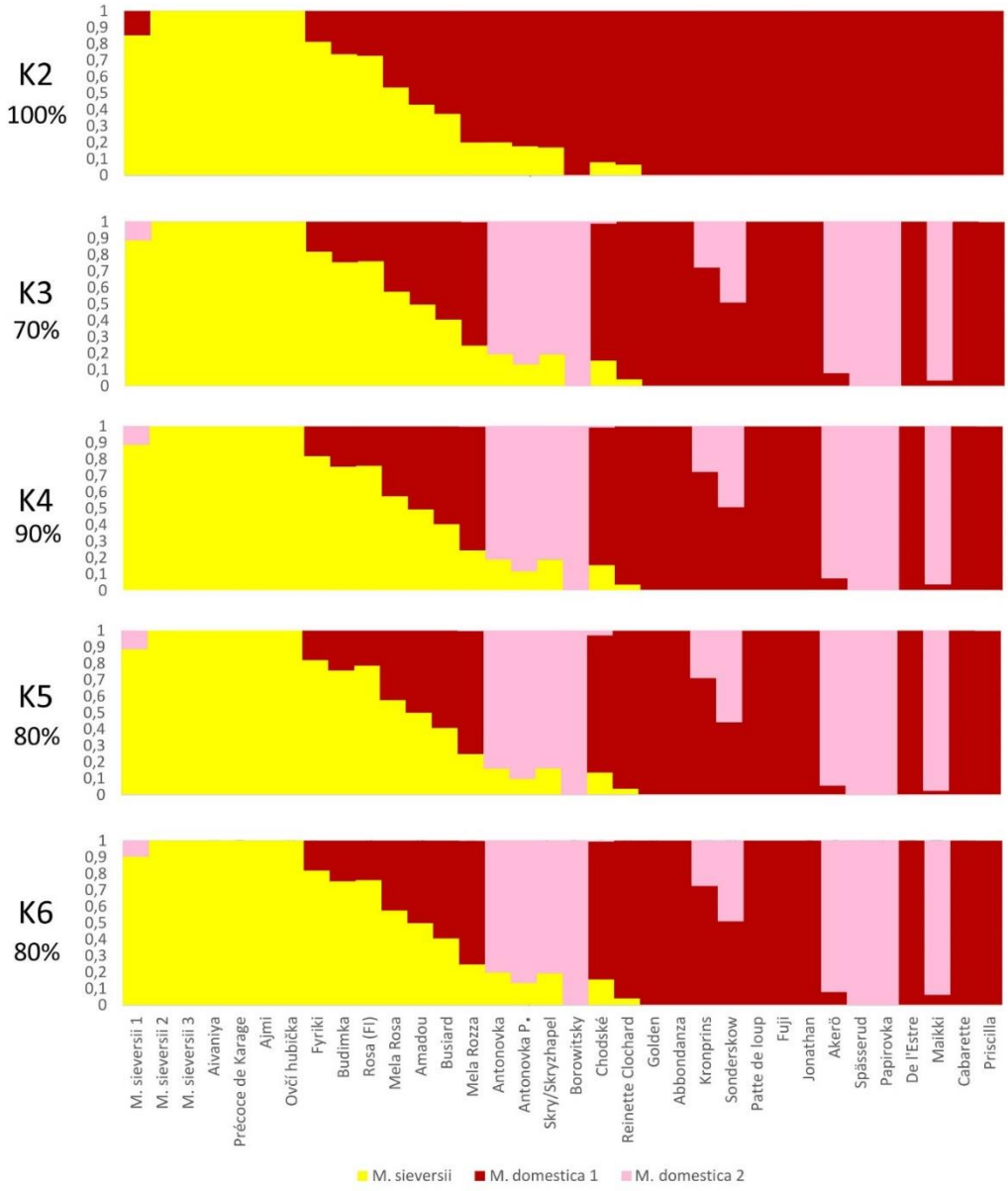


Figure 2: Ancestry proportion of genomes for trees sampled as M. sieversii *(N=3) and cultivated varieties (N=31) by fastStructure in K=2 to K=6 genetic clusters for each sample. Only major modes as found with CLUMPP are presented here with the percentage of corresponding runs indicated at the left of each barplot. Each color represents a differentiated genetic group and each vertical bar represents a genome, divided into K colored segments representing its amount of ancestry proportion in K genetic groups.*

The same genetic structure was found with a principal component analysis (PCA), which is not model-based, i.e., without any assumption on panmixia or Hardy-Weinberg. The first axis indeed separated *M. sylvestris* from the other trees and the second axis separated *M. sieversii* and its closely related cultivated varieties on the one hand from the rest of the cultivated varieties on the other hand. The genetic cluster composed of the North-Eastern European varieties identified above was isolated by the fourth axis (Fig. 3).

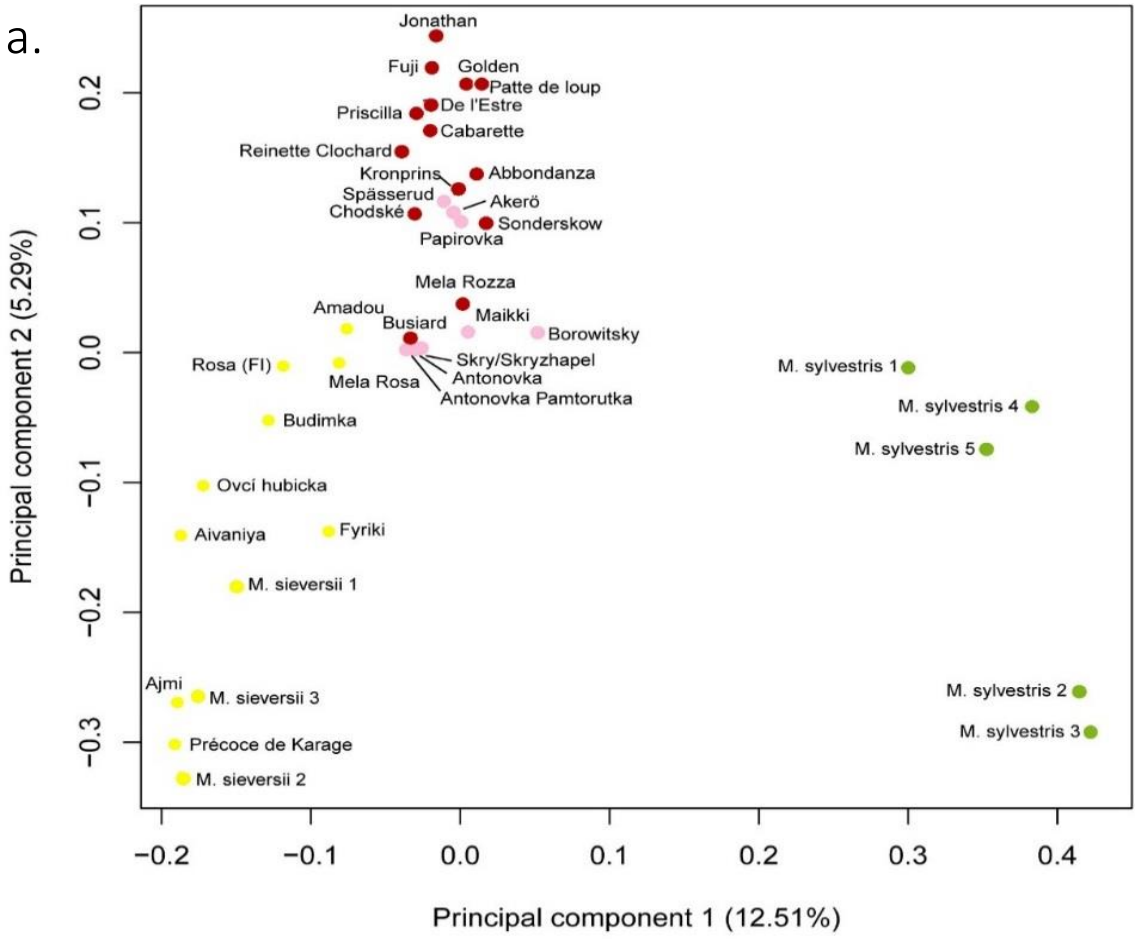


Figure 3: Principal component analysis (PCA) on genomic SNPs of 39 individuals from three Malus *species. a. Principal components 1 and 2 and b. Principal components 3 and 4. The dots are colored according to the genetic group to which the individual was predominantly assigned in a fastStructure analysis. Green:* M. sylvestris*, yellow:* M. sieversii *and associated varieties, red: main group of* M. domestica*, pink: North-Eastern* M. domestica *varieties*

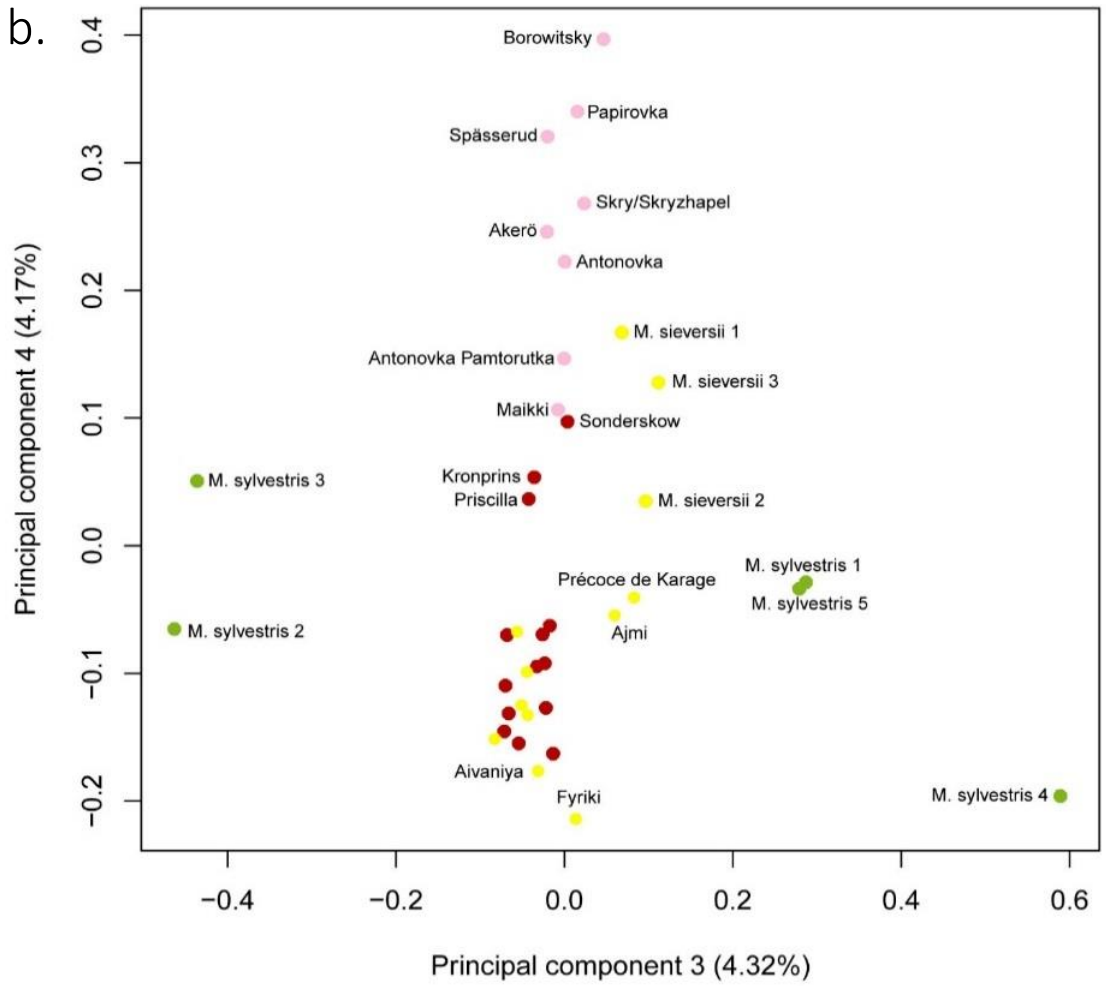


Figure 3: Principal component analysis (PCA) on genomic SNPs of 39 individuals from three Malus *species. a. Principal components 1 and 2 and b. Principal components 3 and 4. The dots are colored according to the genetic group to which the individual was predominantly assigned in a fastStructure analysis. Green:* M. sylvestris*, yellow:* M. sieversii *and associated varieties, red: main group of* M. domestica*, pink: North-Eastern* M. domestica *varieties*

The nucleotide diversity was also calculated per position and then averaged. The mean π was the highest in *M. domestica* (0.18), the lowest in *M. sylvestris* (0.16) and intermediate in *M. sieversii* (0.17).

The *M. sylvestris* ancestry proportion (P_{syl}) was also estimated per 5kb windows. Because we kept only the SNP differentiating *M. sylvestris* from *M. sieversii*, we could only estimate P_{syl} for 27,262 windows out of the 69,428 initial windows. The P_{syl} per window per cultivated varieties belonging to the *M. domestica* genetic cluster (with a threshold of 0.99) seemed to be trimodal, with peaks around 0 (*i.e.* pure *M. sieversii* ancestry), 1 (*i.e.* pure *M. sylvestris* ancestry) and 0.5 (*i.e.* mixed ancestry) (Fig. 4).

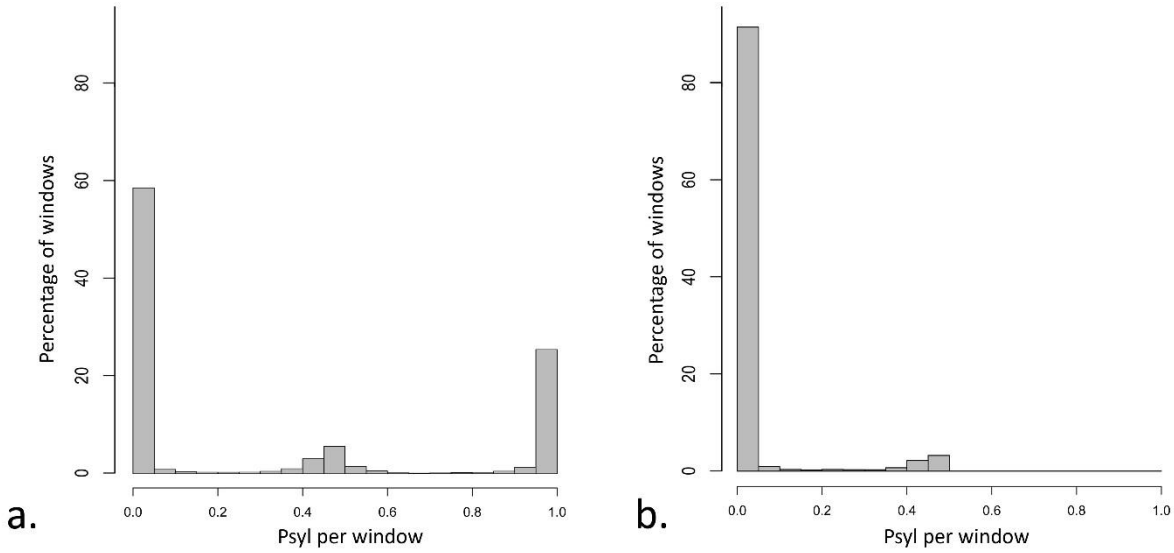


Figure 4: Percentages of sliding windows with a given proportion of M. sylvestris *ancestry (a) in the genomes of the 14 trees assigned to the* M. domestica *genetic cluster with a proportion higher than 0.99 by fastStructure at K=3; and (b) in the genome of the sample named* M. sieversii *1 in the figures above, added here for comparison*

We could detect 516 windows (ca. 2%) where more than 90% of pure M. domestica samples (N=14) had a mixed or pure M. sylvestris ancestry (i.e. Psyl higher than 0.4 with less than 50% of samples with missing data or poor quality measures). There were 4,568 windows (ca. 17%) in which 90% of pure M. domestica samples had a M. sieversii ancestry. There was no significant difference in the mean P_{syl} per window between the samples from the two *M. domestica* genetic groups (t-test: *t-ratio*=0.30, *df*=11.03, *p*=0.77).

At K=2 in fastStructure analyses, the mean P_{syl} per genome over all windows was correlated with both *M. sieversii* and *M. sylvestris* ancestry proportions (Figure 5; *M. sieversii*: r^2=0.45; $F_{(1,26)}$=21.19, slope ± SE=-0.88±0.19; *p*<0.0001; *M. sylvestris*: r^2=0.45; $F_{(1,26)}$= 21.19, slope ± SE=+0.88±0.19; *p*<0.0001). At K=3, the mean P_{syl} per genome over all windows was not significantly correlated anymore to the *M. sylvestris* ancestry proportion as estimated by fastStructure, although the P-value was close to the 0.05 treshold (r^2=0.11; $F_{(1,26)}$=3.34, slope ± SE=1.23±0.67; *p*=0.0791). It was, however, still significantly correlated with the *M. sieversii*

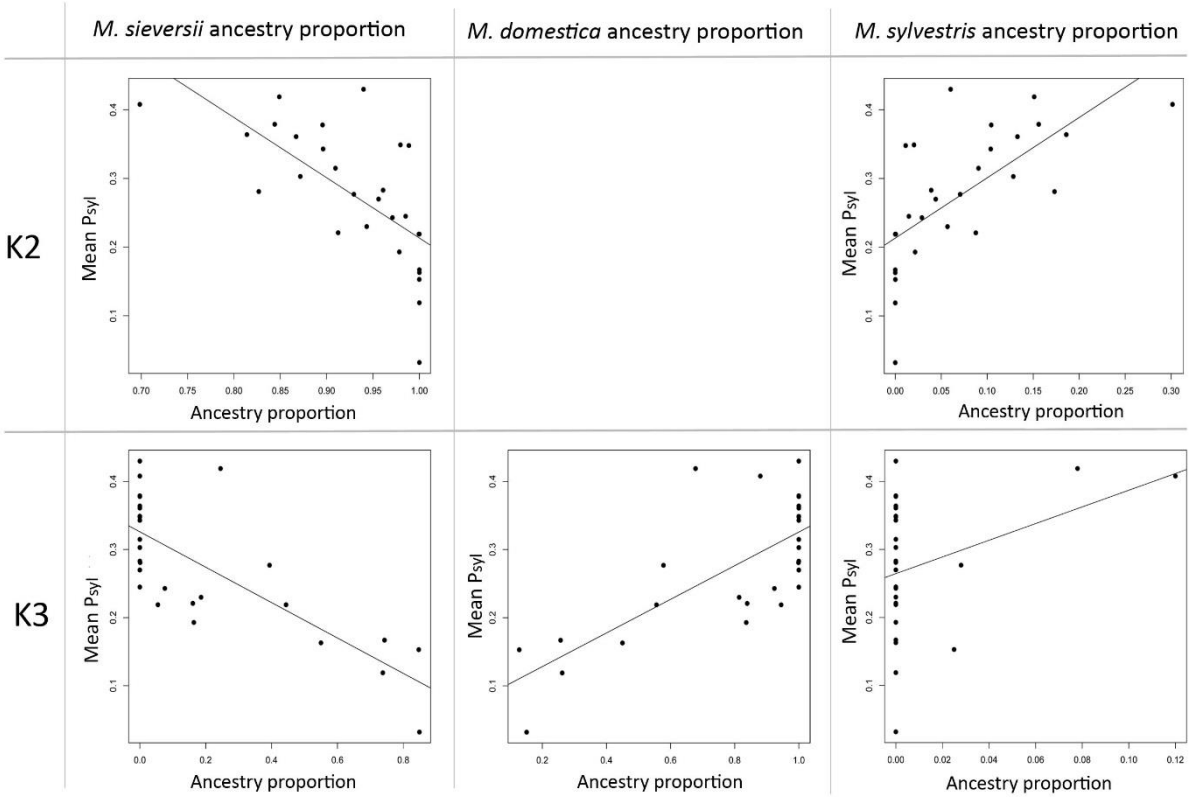


Figure 5: Mean Psyl *over all windows per sample against the ancestry proportion in the* M. sieversii, M. domestica *and* M. sylvestris *genetic groups inferred by fastStructure at K=2 and K=3.*

ancestry proportion (r^2=0.59; $F_{(1,26)}$=38.18, slope ± SE=-0.26±0.04; p<0.0001) and with the *M. domestica* ancestry proportion (r^2=0.54; $F_{(1,26)}$=30.78, slope ± SE=0.25±0.04; p<0.0001). This suggests the possibility that the *M. domestica* cluster identified group by fastStructure is differentiated from the other cultivated varieties mainly by introgressions from *M. sylvestris*.

N.B.: These results are discussed in the "Discussion de la section 1" in French hereunder.

References

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. http://doi.org/10.1093/bioinformatics/btu170

Charmet, G. (2011). Wheat domestication: Lessons for the future. *Comptes Rendus - Biologies*, *334*(3), 212–220. http://doi.org/10.1016/j.crvi.2010.12.013

Cornille, A., Giraud, T., Smulders, M. J. M., Roldan-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, *30*(2), 57–65. Journal Article. http://doi.org/10.1016/j.tig.2013.10.002

Cornille, A., Gladieux, P., Smulders, M. J. M., Roldan-Ruiz, I., Laurens, F., Le Cam, B., … 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 Genetics*, *8*(5). Journal Article. http://doi.org/10.1371/journal.pgen.1002703

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. http://doi.org/10.1093/bioinformatics/btr330

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. http://doi.org/10.1038/ng.806

Ellstrand, N. C., Prentice, H. C., & James, F. (1999). Gene Flow and Introgression from Domesticated Plants into their Wild Relatives. *Annual Review of Ecology and Systematics*, *30*, 539–563.

Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. *Trends in Genetics*, *31*(12), 709–719. http://doi.org/10.1016/j.tig.2015.10.002

Goldschmidt, E. E. (2013). The Evolution of Fruit Tree Productivity: A Review. *Economic Botany*, *67*(1), 51–62. http://doi.org/10.1007/s12231-012-9219-y

Gompert, Z., & Buerkle, C. A. (2010). INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, *10*(2), 378–384. Journal Article. http://doi.org/10.1111/j.1755-0998.2009.02733.x

Gross, B. L., & Olsen, K. M. (2010). Genetic perspectives on crop domestication. *Trends in Plant Science*, *15*(9), 529–537. http://doi.org/10.1016/j.tplants.2010.05.008

Hajjar, R., & Hodgkin, T. (2007). The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica*, *156*(1–2), 1–13. http://doi.org/10.1007/s10681-007-9363-0

Harris, S. A., Robinson, J. P., & Juniper, B. E. (2002). Genetic clues to the origin of the apple. *Trends in Genetics*, *18*(8), 426–430. Journal Article. http://doi.org/10.1016/s0168-9525(02)02689-6

Harrison, N., & Harrison, R. J. (2011). On the evolutionary history of the domesticated apple. *Nature Genetics*, *43*(11), 1043–1044. http://doi.org/10.1038/ng.935

Heslop-Harrison, J. S., & Schwarzacher, T. (2007). Domestication, genomics and the future for banana. *Annals of Botany*, *100*(5), 1073–1084. http://doi.org/10.1093/aob/mcm191

Hufford, M. B., Lubinksy, P., Pyhaejaervi, T., Devengenzo, M. T., Ellstrand, N. C., & Ross-Ibarra, J. (2013). The Genomic Signature of Crop-Wild Introgression in Maize. *Plos Genetics*, *9*(5). Journal Article. http://doi.org/10.1371/journal.pgen.1003477

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*(14), 1801–1806. http://doi.org/10.1093/bioinformatics/btm233

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. http://doi.org/10.1093/bioinformatics/btp324

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. http://doi.org/10.1101/gr.107524.110

McKey, D., Elias, M., Pujol, B., & Duputié, A. (2010). The evolutionary ecology of clonally propagated domesticated plants. *New Phytologist*, *186*(2), 318–332. http://doi.org/10.1111/j.1469-8137.2010.03210.x

Micheletti, D., Troggio, M., Salamini, F., Viola, R., Velasco, R., & Salvi, S. (2011). On the evolutionary history of the domesticated apple. *Nature Genetics*, *43*(11), 1044–1045. http://doi.org/10.1038/ng.983

Miller, A. J., & Gross, B. L. (2011). From forest to field: Perennial fruit crop domestication. *American Journal of Botany*, *98*(9), 1389–1414. http://doi.org/10.3732/ajb.1000522

Nikiforova, S. V., Cavalieri, D., Velasco, R., & Goremykin, V. (2013). Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. *Molecular Biology and Evolution*, *30*(8), 1751–1760. http://doi.org/10.1093/molbev/mst092

Petit, R. J., & Hampe, A. (2006). Some evolutionary consequences of being a tree. In *Annual Review of Ecology Evolution and Systematics* (Vol. 37, pp. 187–214). Book Section. http://doi.org/10.1146/annurev.ecolsys.37.091305.110215

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, *197*(2), 573–589. http://doi.org/10.1534/genetics.114.164350

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics*, *38*(1), 595–619. http://doi.org/10.1146/annurev.ecolsys.38.091206.095646

Urrestarazu, J., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., … Durel, C.-E. (2016). Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. *BMC Plant Biology*, *16*(1), 130. http://doi.org/10.1186/s12870-016-0818-0

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. In *Current Protocols in Bioinformatics* (p. 11.10.1-11.10.33). Hoboken, NJ, USA: John Wiley & Sons, Inc. http://doi.org/10.1002/0471250953.bi1110s43

Velasco, R., & Licciardello, C. (2014). A genealogy of the citrus family. *Nature Biotechnology*, *32*(7), 640–2. http://doi.org/10.1038/nbt.2954

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., … Viola, R. (2010). The genome of the domesticated apple (Malus × domestica Borkh.). *Nat Genet*, *42*(10), 833–839. http://doi.org/10.1038/ng.654

Wu, G. A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., … Rokhsar, D. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology*, *32*(7), 656–662. http://doi.org/10.1038/nbt.2906

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. http://doi.org/10.1093/bioinformatics/bts606

Zohary, D., & Hopf, M. (2000). *Domestication of Plants in the Old World* (3rd ed.). Oxford University Press.

IV. Discussion de la section 1

Flux de gènes des espèces sauvages vers les espèces cultivées

Comme précisé en introduction, les résultats présentés dans ce dernier article sont des résultats préliminaires obtenus à partir de seulement 31 variétés cultivées et d'un génome de référence très fragmenté. Ils nous permettent cependant de tirer des premières conclusions quant aux questions posées dans cette étude.

Structure génétique et domestication

Les résultats obtenus à partir du logiciel fastStructure et de l'analyse de cinq génomes de *M. sylvestris*, trois génomes de M. *sieversii* et 31 génomes de *M. domestica* montrent l'existence de trois groupes génétiques principaux. Le premier groupe est constitué uniquement des échantillons *M. sylvestris*. Les autres individus sont séparés en deux groupes. L'un d'eux inclut les échantillons *M. sieversii* ainsi que quelques variétés cultivées comme la Précoce de Karaj (supposée iranienne) ou encore l'Aivaniya, populaire en Bulgarie dans les années 1960. L'autre groupe est constitué uniquement de variétés cultivées, certaines très populaires commercialement aujourd'hui comme la Jonathan ou la Fuji. Cette structure séparant un groupe de pommiers *M. domestica* des pommiers sauvages d'Asie Centrale et des *M. sylvestris* confirme les résultats obtenus avec l'analyse de marqueurs microsatellites par Cornille et al. (2012) indiquant l'existence de trois groupes génétiques distincts. Elle contredit en revanche la théorie proposée par Velasco et al. (2010) selon laquelle *M. domestica* dans son ensemble ne serait pas différencié génétiquement de son progéniteur *M. sieversii* et aurait seulement subit ce qu'ils nomment une « domestication instantanée » par greffe. Il semblerait cependant que certaines variétés soient réellement directement issues d'arbres sauvages au phénotype intéressant pour

la culture et dont le génotype est propagé par greffe. En effet, le groupe génétique contenant les pommiers sauvages d'Asie Centrale comprend également des variétés cultivées.

Contrairement aux résultats obtenus avec les marqueurs microsatellites (Cornille et al. 2012), les valeurs de F_{ST} calculées sur l'ensemble des SNPs biallélicques montrent que *M. domestica* est plus proche de *M. sieversii* que de *M. sylvestris*. Ce résultat est en accord avec l'histoire présumée du pommier cultivé, c'est-à-dire avec la domestication à partir de *M. sieversii* en Asie Centrale.

La diversité (π) était plus élevée dans les *M. domestica* que dans les deux espèces sauvages. Ceci avait déjà été observé avec d'autres types de marqueurs chez le pommier (AFLP, Coart et al 2003). L'absence de goulet d'étranglement accompagnant la domestication, voire même une diversité plus élevée comme observé chez le pommier cultivé, semble être fréquente dans le cas des espèces pérennes (Miller & Gross, 2011), à cause notamment de modes de reproduction allofécondants et de l'existence d'introgressions depuis d'autres espèces sauvages.

Détection des introgressions

Les analyses obtenues avec fastStructure ne révélent pas d'introgressions massives de *M. sylvestris* dans le génome des variétés cultivées. En effet, la proportion d'ancestralité moyenne dans le groupe *M. sylvestris* obtenue à K=2 pour l'ensemble des variétés cultivées est d'environ 6% et seule une variété a une proportion d'ancestralité *M. sylvestris* de plus de 0,2. Ce résultat est très différent de celui obtenu dans Cornille et al. (2012) qui donnait 26% des individus dépassant ce seuil. Plusieurs explications peuvent être avancées pour cette différence.

La première hypothèse pourrait être la différence dans l'échantillonnage des variétés cultivées. En effet, de nombreuses variétés étudiées dans l'article de Cornille et al. (2012) provenaient de conservatoires bretons, dont les variétés auraient pu être plus proches des *M. sylvestris* que la moyenne. Seules trois variétés étaient communes entre les analyses des génomes et des marqueurs microsatellites (Ajmi, Fuji et Jonathan) et les assignations aux groupes génétiques n'étaient en accord que pour les deux dernières. Une autre hypothèse est la nature des marqueurs, les microsatellites pouvant donner une vision biaisée par rapport aux génomes complets. Pour tester ces hypothèses, je compte extraire les marqueurs microsatellites du jeu de données obtenus avec les génomes alignés sur le nouvel assemblage et je ferais une analyse similaire avec celle de Cornille et al. (2012). Il est possible cependant que le faible nombre d'individus dans mon jeu de données ne permettent pas d'obtenir de résultat clair avec peu de marqueurs et que je ne puisse pas valider ou invalider cette hypothèse.

Une dernière hypothèse pour expliquer les divergences de résultats entre les études est en lien avec les outils utilisés. En effet, bien que fastStructure ait été créé pour remplacer STRUCTURE pour les jeux de données issus du séquençage massif de nouvelle génération, ces deux logiciels ne fonctionnent pas exactement de la même façon. Dans l'article décrivant fastStructure, une comparaison a été faite des performances des deux logiciels sur deux jeux de données avec une forte structuration de la population et avec une faible structuration (Raj, Stephens, & Pritchard, 2014). Le logiciel STRUCTURE pouvait inférer à tort une hybridation lorsque la structuration est faible (Fig. 1.5). Cependant, le texte de l'article précise que détecter de faibles taux de flux de gènes en utilisant fastStructure n'est peut-être pas faisable en raison de sa tendance « aggressive » à n'utiliser que le nombre de groupes expliquant au mieux la structuration au lieu d'utiliser tous les groupes disponibles pour expliquer les données (c'est-à-dire qu'il n'utiliserait que trois groupes même à K=5 comme dans la Fig. 1.5). Une comparaison sur notre jeu de données à K=4 semble confirmer cette idée, avec huit échantillons assignés à trois groupes différents avec une proportion supérieure à 0,1 par STRUCTURE contre seulement un échantillon par fastStructure (données non présentées). Je n'ai pas pu faire la comparaison inverse puisque fastStructure ne peut gérer que des locus bi-alléliques et ne peut donc pas estimer les proportions d'admixture à partir des microsatellites.

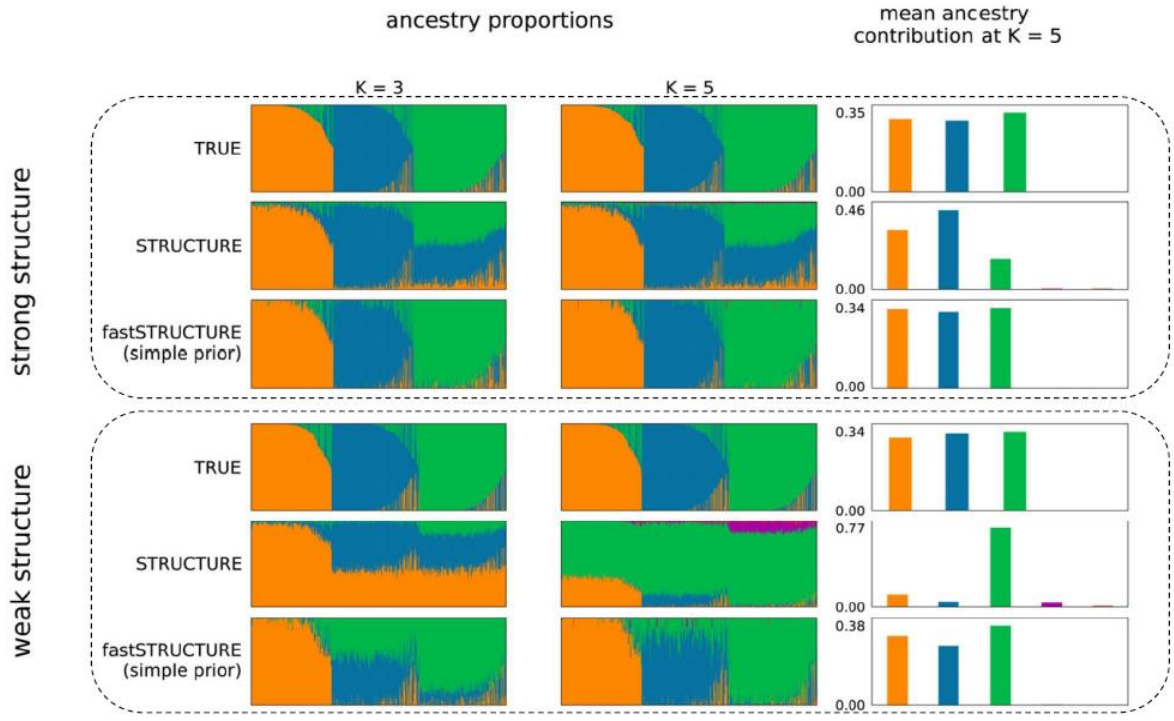


Figure 1.5: (Gauche) Proportions d'ancestralité estimées pour K=3 et K=5 par différents algorithmes sur deux jeux de données simulées, le premier avec une structuration forte (en haut) et le second avec une structuration faible (en bas). Chaque couleur représente un groupe génétique et chaque barre verticale représente un individu. (Droite) Contributions moyennes de chaque groupe génétique possible lorsque K=5. Figure adaptée de Raj et al, 2014

Après avoir fait la constatation de l'absence d'introgressions massives apparentes dans les analyses fastStructure sur les génomes complets, j'ai cherché à voir si des introgressions plus localisées pouvaient être observées en analysant les génomes non plus dans leur ensemble, mais à l'échelle locale. En effet, d'autres signaux, comme la place intermédiaire des variétés cultivées sur l'axe de l'ACP séparant les *M. sylvestris* des *M. sieversii,* ou les valeurs de F_{ST}, suggéraient une possibilité d'introgressions. Le calcul des F_{ST} entre *M. sieversii* and *M. sylvestris* par position a montré un obstacle à la détection des introgressions : la différenciation entre ces deux espèces est faible pour de très nombreux SNPs en raison de polymorphisme ancestral partagé. Pour contourner cette difficulté, j'ai réalisé l'analyse en ne conservant que les SNPs ayant un F_{ST} de plus de 0,8 (N=861 913).

L'analyse par fenêtre glissante avec seulement les SNP différenciant les deux espèces sauvages a révélé l'existence de régions avec un P_{syl} élevé dans les variétés cultivées. Ainsi, parmi les 14 échantillons assignés au groupe *M. domestica*, 26% des fenêtres analysées par *Introgress* avaient un P_{syl} supérieur à 0,9. Ces fenêtres correspondent probablement à des régions introgressées par *M. sylvestris* dans le génome de ces variétés. Dans l'ensemble, la moyenne de P_{syl} pour les individus analysés ici est corrélée à la proportion d'assignation au groupe génétique *M. domestica* ce qui suggère que ce groupe génétique distinct s'est formé au moins en partie sur la base d'introgressions avec *M. sylvestris*. J'ai d'autre part identifié 514 fenêtres introgressées dans toutes les variétés cultivées *M. domestica*, ce qui correspond à 209 régions chromosomiques différentes (certaines fenêtres étant adjacentes). J'ai également identifié 4568 fenêtres qui n'étaient jamais introgressées par *M. sylvestris*, correspondant à 1372 locus différents.

Lorsque les analyses seront refaites avec le nouvel assemblage moins fragmenté, il sera possible de travailler sur des fenêtres plus grandes permettant d'englober un plus grand nombre de SNP par fenêtre. Cela permettra certainement d'obtenir des estimations du P_{syl} avec un intervalle de confiance plus réduit et donc d'avoir moins de fenêtres traitées comme manquantes car peu fiables. Je chercherai également des régions chromosomiques qui pourraient avoir des valeurs de P_{syl} élevées dans toutes les variétés étudiées ou bien seulement dans certaines d'entre elles ou bien, à l'inverse, des régions qui n'apparaitraient jamais introgressées. Un meilleur assemblage permettra de visualiser ces introgressions le long des chromosomes.

Si de telles régions sont identifiées, il sera alors intéressant d'identifier le contenu, et éventuellement les gènes et QTL (pour « quantitative trait loci », c'est-à-dire une région chromosomique liée à un caractère phénotypique quantitatif), situés dans ces régions systématiquement ou jamais introgressées afin de déterminer si le pommier sauvage européen a contribué à des caractères phénotypiques d'intérêt agronomique. Le pommier sauvage européen

est connu pour sa résistance à certains pathogènes parasitant le pommier cultivé, comme *Venturia inaequalis*, responsable de la tavelure (Le Van et al., 2012), et il est possible qu'il ait contribué au génome du pommier cultivé pour des gènes de résistance.

Sous-structuration chez le pommier cultivé

Inférer la structure génétique sans les pommiers sauvages européens a permis de confirmer l'existence d'une structure génétique chez *M. domestica*, déjà identifiée précédemment sur la base de marqueurs microsatellites (Urrestarazu et al 2016, Fig.1.6). En plus des pommiers cultivés détectés comme appartenant au groupe génétique de *M. sieversii*, deux autres groupes de pommiers se distinguent avec l'analyse de fastStructure. Ces trois groupes sont également visibles sur le premier et le quatrième axe de l'ACP. La grande quantité de cultivars utilisés dans l'étude de Urrestarazu and Denancé et al. (2016, N=1869) leur a permis de relier ces trois groupes génétiques qu'ils ont également détectés à trois régions d'Europe : Nord-Est, Ouest et Sud. Dans notre cas, les variétés appartenant au groupe identifié comme Nord-Est grâce aux marqueurs microsatellites proviennent effectivement de pays correspondant à cette zone géographique comme la Spässerud (suédoise), la Maikki (finlandaise) ou encore la Papirovka (russe). Les variétés appartenant au groupe Sud sont également issues de pays comme la Bulgarie, l'Iran ou la Tunisie (avec l'exception d'une variété russe). Ce dernier groupe est celui qui inclut les pommiers sauvages *M. sieversii* dans notre analyse.

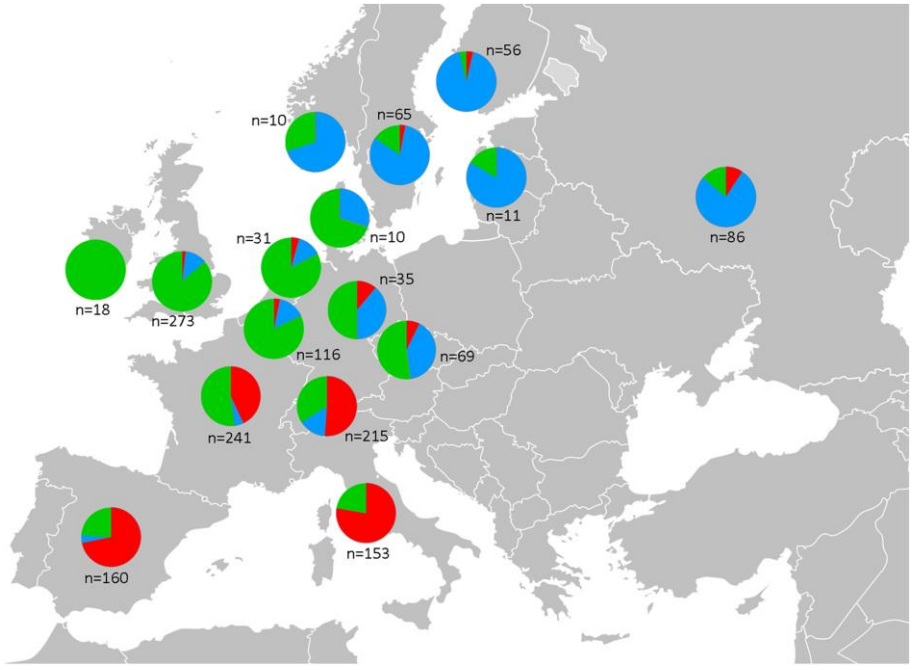


Figure 1.6 : Proportions des assignations aux différents groupes à K=3 par le logiciel STRUCTURE dans variétés cultivées originaires de différents pays d'Europe. Figure tirée de Urrestarazu and Denancé et al. 2016.

Deux hypothèses sont proposées par Urrestarazu and Denancé et al. (2016) pour expliquer l'existence de ces groupes génétiques. La première hypothèse est une différenciation sous l'action de la dérive génétique comme dans le cas de l'abricot (Bourguiba et al. 2012). Bien que cette hypothèse ne puisse être exclue complètement, la domestication du pommier ne semble pas s'être accompagnée, comme pour l'abricot, d'une baisse de diversité, ce qui peut limiter l'intensité de la dérive génétique. Il est possible cependant qu'il y ait eu un bottleneck à un moment de l'histoire de domestication mais que son effet soit masqué par l'existence d'introgressions venant du pommier sauvage européen. Une fois que les régions génomiques concernées par ces introgressions seront identifiées, il sera peut-être possible de recalculer la diversité génétique dans les seuls locus non introgressés afin de tester la validité de cette hypothèse.

La seconde hypothèse proposée pour expliquer l'existence des trois groupes génétiques est l'adaptation locale aux conditions environnementales des trois régions concernées, comme dans le cas du raisin, pour lequel une partie des groupes révélés par des analyses génétiques se différencient par l'usage des fruits, comme raisin de table ou raisin à vin (Bacilieri et al., 2013). N'ayant pas accès, à ce jour, à des données phénotypiques complètes sur les variétés analysées, cette hypothèse est difficile à tester. Pour contourner ce problème, j'utiliserai les données produites avec le nouvel assemblage pour chercher des régions du génome qui pourraient être plus différenciées entre ces trois groupes génétiques que le reste du génome ou introgressées à des locus différents par *M. sylvestris*. Nos collaborateurs italiens ayant prévu d'identifier des régions sous sélection dans le génome, je pourrais ensuite comparer de potentielles régions différenciées avec les régions sous sélection qu'ils auront détectées.

L'addition dans notre jeu de données de pommiers sauvages nous permet de proposer une troisième hypothèse expliquant la structure des populations de *M. domestica* (qui n'exclut pas les hypothèses ci-dessus). Puisque le groupe Sud détecté avec les marqueurs microsatellites (Urrestarazu et al., 2016) correspond dans mes analyses à celui incluant les *M. sieversii*, ce groupe de variétés pourrait être un groupe ancestral de variétés cultivées n'ayant que peu changé depuis l'importation des premiers pommiers cultivés par la route de la soie aboutissant à Antioche, aujourd'hui en Turquie. Les deux groupes situés plus au Nord auraient pu se différencier du groupe ancestral notamment par hybridations avec le pommier sauvage européen dont c'est l'aire de répartition. L'existence de deux groupes distincts pourrait alors être expliquée par deux routes de diffusion de la culture du pommier ou par des introgressions différentes.

Une quatrième hypothèse serait que ces groupes génétiques résultent de plusieurs évènements de domestication indépendants, suivis par des introgressions par le pommier sauvage européen dans le cas des deux groupes d'Europe du Nord. Ainsi, dans le cas de l'abricotier, des domestications indépendantes semblent avoir été à l'origine des différents groupes de variétés trouvées sur le continent eurasiatique (Decroocq et al., 2016).

Afin de départager ces hypothèses, j'envisage de réaliser des analyses ABC (de l'anglais « Approximate Bayesian Computation ») permettant de comparer l'adéquation de ces scénarios historiques et démographiques à notre jeu de données. Brièvement, des données génétiques seraient simulées sur la base de différents scénarios évolutifs puis les simulations seraient comparées aux données observées par le biais de statistiques résumées (comme le F_{ST}, Pi etc.). Le scénario favorisé sera celui dont les statistiques résumées sont les plus proches de celles des données observées.

En conclusion, les analyses présentées dans cette ébauche d'article, bien que préliminaires, permettent déjà de tirer quelques conclusions concernant l'histoire du pommier cultivé. Nous avons détecté l'existence d'un groupe génétique *M. domestica* qui semble inclure des introgressions de *M. sylvestris*. Ces introgressions seront localisées dans des analyses futures facilitées par le nouvel assemblage de meilleure qualité. A l'inverse, certaines variétés cultivées du sud de l'Europe et du Nord de l'Afrique appartiennent au même groupe génétique que *M. sieversii* et semblent dépourvues d'introgressions. Une sous-structure a été identifiée dans le groupe génétique *M. domestica* dont l'origine sera à étudier plus en détail dans les analyses que nous présenterons dans l'article final.

Flux de gènes des espèces cultivées vers les espèces sauvages

En plus d'avoir étudié l'influence des flux de gènes entre *M. sylvestris* et *M. domestica* sur la domestication du pommier cultivé, je me suis intéressée, dans deux études distinctes présentées ci-dessus, aux conséquences de ces flux de gènes sur la conservation du pommier sauvage européen. Au cours de ce projet, j'ai donc cherché à répondre pour notre cas d'étude aux trois questions posées par la Société pour la Biologie de la Conservation, c'est-à-dire : 1) Comment la diversité génétique de *M. sylvestris* est-elle distribuée à l'échelle de son aire de répartition ? 2) A quelles menaces fait face cette espèce (dans cette thèse, je me suis focalisée sur la menace apportée par les flux de gènes avec *M. domestica*) ? 3) Quelles actions peuvent être entreprises afin de réduire ou d'éliminer ces menaces et de restaurer la diversité biologique ?

Structure des populations : étude des climats passés et conservation

La première étude présentée ici nous a permis d'étudier la diversité génétique du pommier sauvage européen à travers l'Europe et sa structure de population. Ce nouvel échantillonnage couvrant quasiment l'ensemble de l'aire de répartition de *M. sylvestris* a permis d'obtenir une vision bien plus nette de la structure de la population de cette espèce que l'étude précédente qui révélait trois groupes génétiques (Cornille et al 2013a). Nous avons ainsi pu identifier 5 groupes génétiques cohérents géographiquement. Nous avons également pu observer une répartition particulière de la diversité génétique, avec une diversité plus importante dans les régions du sud de l'Europe que vers le nord. Nous avons interprété ces résultats dans le cadre des changements climatiques passés et notamment d'une contraction de l'aire de répartition du pommier sauvage en Europe au cours de la dernière grande glaciation. Cependant, puisque nous avons identifié davantage de groupes génétiques que les trois correspondant aux péninsules du sud de l'Europe, le pommier sauvage a certainement une histoire plus complexe qu'une simple recolonisation issue de ces trois grands refuges. L'existence d'un groupe génétique au Nord de l'Europe, de faible diversité génétique, pourrait provenir d'un phénomène d' « allele surfing » qui créé un goulot d'étranglement sur le front des vagues de colonisation (Excoffier & Ray, 2008). L'identification de deux groupes génétiques en France au lieu d'un groupe unique qui serait issu d'un refuge ibérique pourrait provenir de l'existence d'une fragmentation des populations dans ce refuge en raison d'un climat non-homogène (Gómez & Lunt, 2007). Il est également possible que ces groupes proviennent de l'existence de refuges alternatifs ou micro-refuges situés dans le sud de la France et dont l'existence a été montrée chez d'autres espèces d'arbres comme le hêtre (de Lafontaine, Amasifuen Guerra, Ducousso, & Petit, 2014; de Lafontaine, Amasifuen Guerra, Ducousso, Sánchez-Goñi, & Petit, 2014).

Du point de vue de la conservation du pommier sauvage en France, il est important de prendre en compte la structure génétique qui différencie deux groupes génétiques suivant un axe est-ouest. En effet, ces deux groupes forment des unités de conservation puisque ce sont des populations isolées historiquement, et avec des fréquences alléliques différentes (Newton et al., 1999). Cependant, selon de Guia and Saitoh (2007), les unités de conservation ne peuvent être que partiellement définies par des données sur des marqueurs neutres, et il serait idéal d'obtenir dans l'avenir des informations sur la valeur sélective et adaptative des populations à travers la France, et notamment sur l'existence ou non d'une adaptation locale.

D'importants flux de gènes influencés par des facteurs anthropiques

Dans une seconde partie de cette première étude, nous avons non seulement étudié l'intensité des flux de gènes d'une espèce d'arbre cultivé, *M. domestica*, vers son apparenté sauvage, *M. sylvestris*, à l'échelle de son aire de répartition, mais également les facteurs anthropiques influençant ces flux de gènes. Ainsi nous avons montré que les hybridations entre *M. domestica* et *M. sylvestris* sont très fréquentes puisque 23% des individus étudiés montraient des signes d'introgressions. Nous avons également montré que les flux de gènes entre ces espèces sont bien corrélés avec l'intensité des activités humaines dans les milieux agricoles de façon directe. La forte intensité des flux de gènes de *M. domestica* vers *M. sylvestris* est ainsi liée à la production locale de pommes, au nombre de vergers et à leur surface. Nous avons également montré que l'agriculture intensive pourrait avoir un effet sur les flux de gènes à la fois entre espèces et entre populations indirectement. Ainsi une réduction du nombre de pollinisateurs visitant les fleurs de pommiers a été observée dans les régions d'agriculture intensive. Ces résultats montrent que les activités agricoles humaines influencent bien les flux de gènes des espèces cultivées vers leurs apparentés sauvages. Ils confirment également l'idée qu'une culture plus intensive accroit le risque de « genetic swamping » pour les espèces sauvages apparentées aux espèces cultivées.

Valeur sélective aux premiers stades de développement

Les conséquences des flux de gènes des espèces cultivées vers les espèces sauvages ont déjà été étudiées chez différentes espèces végétales (Ellstrand et al. 2013). Cependant, à notre connaissance, notre étude est la première à porter sur une espèce cultivée fruitière pérenne. Nous avons également montré que, du moins dans les premiers stades de développement et dans des conditions environnementales contrôlées, le taux d'ancestralité dans les deux espèces étudiées à un impact sur la croissance que nous avons utilisée comme une composante de la valeur sélective. En effet, les graines germent plus rapidement et les jeunes plants poussent plus rapidement lorsque leur proportion d'ancestralité dans l'espèce *M. domestica* est plus haute. Ce résultat suggère que les hybrides entre *M. domestica* et *M. sylvestris* n'ont pas une valeur sélective plus faible, au moins dans de bonnes conditions de croissance en serre, ce qui contribue à expliquer les forts taux d'introgressions observés en Europe et à l'échelle locale et qui confirme la menace posée pour l'intégrité génétique de l'espèce *M. sylvestris*. Il serait intéressant d'étudier la croissance sous différents stress biotiques et abiotiques ainsi que dans différents sols. *Malus sylvestris* est connu pour sa croissance dans des conditions défavorables comme des milieux très

humides ou à l'inverse très peu humides (Schnitzler et al., 2014) ou encore sa résistance à différents pathogènes, comme la tavelure (Le Van et al., 2012). Confirmer expérimentalement les capacités de résistance de *M. sylvestris* à différents stress permettrait de donner plus de poids aux projets pour sa conservation puisque cette espèce pourrait alors être inclue dans les programmes d'amélioration variétale de *M. domestica*, en vue de créer notamment de nouveaux portes-greffes.

Capacités de dispersion

Les flux de gènes sont en grande partie liés aux capacités de dispersion des espèces impliquées. Nous avons utilisé deux approches différentes pour estimer les capacités de dispersion de *M. sylvestris* : dans la première étude, nous avons utilisé une approche indirecte, que nous avons complétée dans la seconde par une approche directe.

La première approche que nous avons utilisée est indirecte puisque c'est, en fait, une estimation de l'intensité de la structure génétique spatiale, qui elle-même découle des capacités de dispersion de la plante (dispersion par le pollen ou par les graines). Nous avons pour cela utilisé le paramètre *Sp* qui représente le taux de diminution de l'apparentement entre paires d'individus avec la distance et qui a l'avantage de pouvoir être comparé entre différentes espèces (Vekemans & Hardy, 2004). Dans notre étude, l'intensité de la structure génétique spatiale était faible, ce qui suggère de fortes capacités de dispersion (et/ou de fortes tailles efficaces). Les valeurs de *Sp* (*Sp* moyen = 0,012) sont cependant bien supérieures à celles obtenues dans l'étude précédente sur le pommier sauvage (*Sp* compris entre 0,005 et -0,003 ; Amandine Cornille, Gladieux, et al 2013), certainement grâce à l'échantillonnage plus complet. Ces valeurs peuvent être comparées à celles estimées par Vekemans et Hardy (2004) pour des espèces ayant des traits d'histoire de vie communs au pommier sauvage. Elles sont similaires à celles obtenues pour les espèces de plantes ayant des systèmes d'auto-incompatibilité (*Sp* moyen = 0,013). Cette valeur de *Sp* est plus faible que la moyenne obtenue pour les espèces dont le pollen est dispersé par des animaux (*Sp* moyen = 0,017) mais plus élevée que la moyenne chez les espèces dont les graines sont dispersées par des animaux (*Sp* moyen = 0,009). Dans l'ensemble nos estimations sont en accord avec celles qui ont été effectuées sur d'autres espèces de plantes partageant des traits d'histoire de vie avec *M. sylvestris*.

Dans la seconde étude présentée, nous avons cherché à approfondir la question de la dispersion et à obtenir, cette fois-ci, une estimation directe de la capacité de dispersion du pollen. Grâce à des analyses de parenté et à un échantillonnage quasi-exhaustif des pommiers du site d'étude, nous avons observé l'existence de flux de pollen à longue distance (environ 4 km) et

donc de flux de gènes entre arbres distants. Des évènements de pollinisation à longue distance ont également été observés par Reim et al (2015) dans une étude publiée au cours de nos analyses. Nous avons montré que la distance des flux de pollen est dépendante de la densité en arbres de cette espèce, avec des distances plus courtes lorsque la densité est plus élevée. Ce résultat est intuitif : les pollinisateurs volent d'autant moins loin que des arbres en fleurs sont disponibles à proche distance. Néanmoins, ce résultat est très important du point de vue de la conservation. Tout d'abord, il renforce les problèmes causés par la fragmentation de l'habitat naturel puisque dans un environnement fragmenté, et donc moins dense, les distances de pollinisation sont plus importantes, ce qui augmente le risque de flux entre espèces. De plus, savoir que les distances de pollinisation sont plus courtes dans un environnement dense en pommiers permet de souligner le fait qu'un potentiel verger à graines devra être aussi dense que possible sans compromettre la croissance des arbres permettant ainsi de limiter les flux de gènes vers les arbres extérieurs au verger.

Etude des filières existantes de vente de Malus sylvestris

Notre étude du système actuel de vente du pommier sauvage européen en France a révélé de nombreuses insuffisances. Nous avons étudié ce système sous différents angles : la filière publique, la filière privée et une filière locale et associative.

En tant que représentant de la filière locale, nous avons étudié le cas de l'association française d'agroforesterie (AFAF) avec laquelle nous avons établi une collaboration sur plusieurs années. L'association participe à de nombreux projets de recherche et de développement afin de permettre le transfert des connaissances des laboratoires de recherches vers les responsables institutionnels, les agriculteurs et d'autres associations dont les buts sont plus appliqués. Parmi celles-ci figure l'association « Arbre et Paysage du Gers » qui conçoit et accompagne la réalisation de nombreux de projets de plantations d'arbres ou de gestion de la végétation existante dans le cadre de l'agroforesterie. L'agroforesterie désigne les pratiques associant arbres, cultures et/ou animaux sur les mêmes parcelles agricoles, que ce soit en plein champs ou en bordure de champs. Dans ce cadre, le pommier sauvage est utilisé pour former des haies entre champs ou le long des chemins où il est apprécié pour son ancrage et sa vigueur y compris sur sol de qualité médiocre. Cependant, la difficulté pour les agriculteurs et les associations participant aux projets d'agroforesterie est de trouver des filières fiables pour se pourvoir en graines et plants de pommiers sauvages européens dont l'origine est connue et vérifiée. Le risque est d'introduire des individus provenant de populations exogènes mal adaptés aux conditions locales ou même étant hybrides avec d'autres espèces de pommiers. L'association avec laquelle

nous avons travaillé se procure elle-même les graines qu'elle utilise et, parmi les plants étudiés, seule la moitié a effectivement été confirmée comme étant du *M. sylvestris*. Ceci souligne tout d'abord la difficulté d'identifier des arbres sauvages sans analyses génétiques, mais aussi et surtout, le manque d'une filière de vente connue et fiable pour cette espèce.

Dans l'ensemble, les filières privées se sont révélées particulièrement problématiques. Le premier problème est un manque de clarté dans les appellations. Les pommiers « sauvages » vendus sont indifféremment nommés « *M. communis* », « *M. communis (sylvestris)* », « *M. sylvestris* », « Pommiers francs », « Pommiers sauvages » ou encore « Pommier commun ». Le nom *M. communis* est obsolète et a désigné à la fois le pommier sauvage européen, le pommier cultivé et même parfois le poirier, ce qui peut conduire naturellement à des confusions. Nous avons même trouvé une compagnie néerlandaise associant le nom d'espèce « *Malus sylvestris* » dans son catalogue avec des noms de variétés cultivées connues comme « Antonovka » et « Bittenfelder ». Charly Chauveau, qui a réalisé un stage en collaboration entre l'ESE et l'association AFAF, a également mis en évidence le fait qu'une fois les plants vendus par les pépinières, l'origine des graines devient floue, avec le pays de récolte comme seule référence et les noms sont parfois modifiés dans les catalogues, apparemment sans lien strict avec la dénomination employée par les fournisseurs de graines. En plus de cette confusion dans les appellations, tous les échantillons étudiés se sont révélés avoir de forts taux d'introgression par *M. domestica,* voire même être des pommiers cultivés sans traces de l'espèce sauvage. Ces problèmes proviennent probablement d'une part du manque de règlementation concernant le pommier sauvage européen en France et d'autre part du fait que le marché de cette espèce reste très marginal (moins de 100 000 plants commercialisés annuellement en France).

La filière de production de plants de pommiers sauvages publique fut plus simple à étudier. En effet, il n'existe qu'une seule source de graines gérée par l'Office National des Forêts qui est le verger à graines situé dans la forêt de Dourdan. Le verger lui-même n'est que marginalement entretenu et j'ai observé des branches du porte-greffe *M. domestica* qui fleurissent toujours sur les arbres avec les branches sauvages greffées. Par chance, cette variété cultivée semble incapable de fertiliser les fleurs sauvages (aucune des graines que nous avons analysées n'a été fertilisée par cette variété cultivée), probablement car elle est triploïde, alors que les individus sauvages greffés sont diploïdes. Les analyses génétiques que nous avons faites ont révélé que les arbres sauvages greffés dans le verger sont effectivement des *M. sylvestris*. Cependant, le verger montre une diversité génétique remarquablement faible : sur les 14 arbres greffés, nous n'avons retrouvé que trois génotypes sauvages différents. Pour ajouter à ce problème de diversité, les trois arbres sur lesquels ont été prélevés les greffons proviennent tous

de la forêt de Dourdan elle-même (un arbre aujourd'hui disparu et deux que nous avons très probablement identifiés ; communication personnelle de Didier Desseaux, le garde forestier de la forêt ayant mis en place le verger). A titre de comparaison, Stephan et al (2003) recommandent un nombre minimum de 50 génotypes différents pour créer un verger conservatoire pour *M. sylvestris*.

Mise en pratique des conclusions obtenues

Après avoir fait le constat que le système en place au début de ma thèse n'apportait pas de protection suffisante au pommier sauvage européen en France, nous avons établi une collaboration avec l'institut national de recherche en sciences et technologies pour l'environnement et l'agriculture (IRSTEA), et en particulier avec Eric Collin et Anne Pierangelo, afin de mettre en place une nouvelle règlementation et d'améliorer le processus de vente et de réimplantation du pommier sauvage européen en France. Cette réglementation a été mise en place durant ma dernière année de thèse avec la création de deux régions de provenance pour les matériels forestiers de reproduction ou MFR, correspondant aux deux groupes génétiques identifiés dans le premier article présenté dans cette section. La diffusion des conseils correspondant à la réglementation a été faite en partie grâce à une fiche technique rédigée par Eric Collin et Anne Pierangelo et mise à disposition par le ministère de l'agriculture (Annexe 2).

Une fois la réglementation en place, il devient urgent de trouver des sources alternatives fiables de semences. Dans un premier temps, j'ai accompagné une équipe constituée de gardes forestiers de l'ONF et de représentants de l'INRA et de l'IRSTEA dans la forêt de Dourdan afin d'identifier sur place des arbres pouvant potentiellement servir à la récolte des prochaines années. Pour cela, j'ai pris en compte différents critères comme la proportion d'ancestralité dans le pool génétique du pommier sauvage de l'arbre, mais aussi lorsque c'était possible, des graines récoltées sur l'arbre, l'accessibilité des arbres avec des véhicules et la densité en individus d'intérêt afin de faciliter la récolte.

Cependant, cette solution n'est destinée qu'à être temporaire en attendant la mise en place d'un nouveau verger conservatoire produisant des semences génétiquement diverses. Avec nos collaborateurs, nous travaillons actuellement à mettre en place un nouveau verger conservatoire, soit à l'emplacement du verger actuel, soit à un emplacement complètement nouveau. Les arbres greffés pourraient être les arbres que nous avons identifiés comme appartenant à l'espèce *M. sylvestris* sans introgressions dans l'étude de Cornille et al (2015) et devraient appartenir aux deux régions de provenance des MFR afin de représenter la diversité

génétique complète à l'échelle de la France. Il serait alors préférable de faire deux vergers différents afin d'éviter des hybridations inter-clusters.

SECTION 2

RELATIONS ENTRE HÔTES ET PATHOGÈNES

I. Introduction générale de la section 2

Interactions entre organismes et évolution

Différents types d'interactions entre organismes

Les interactions entre organismes peuvent avoir lieu entre individus d'une même espèce ou bien d'une même population, sous forme, par exemple, de compétition ou de coopération. Mais elles peuvent également impliquer des individus appartenant à différentes espèces. Alors que certaines interactions n'impliquent pas de changements dans la valeur sélective des partenaires (on parle alors de neutralisme), d'autres types d'interactions ont un effet sur la valeur sélective d'au moins un des deux partenaires. En modifiant l'action de la sélection naturelle, ces interactions entre individus conduisent à des changements évolutifs réciproques, un processus qu'on désigne par le terme de « coévolution ». Pour illustrer cette notion, une allégorie en référence au livre de Lewis Caroll, *Alice de l'Autre Côté du Miroir* (Fig. 2.1) est fréquemment utilisée sous le nom de théorie de la Reine Rouge depuis son invention par van Valen (1973). La course folle d'Alice et de la Reine Rouge nécessaire pour rester sur place dans le monde fantastique en mouvement du roman est ainsi utilisée pour symboliser l'évolution permanente indispensable aux espèces pour maintenir un *statu quo* face aux espèces avec lesquelles elles co-évoluent.

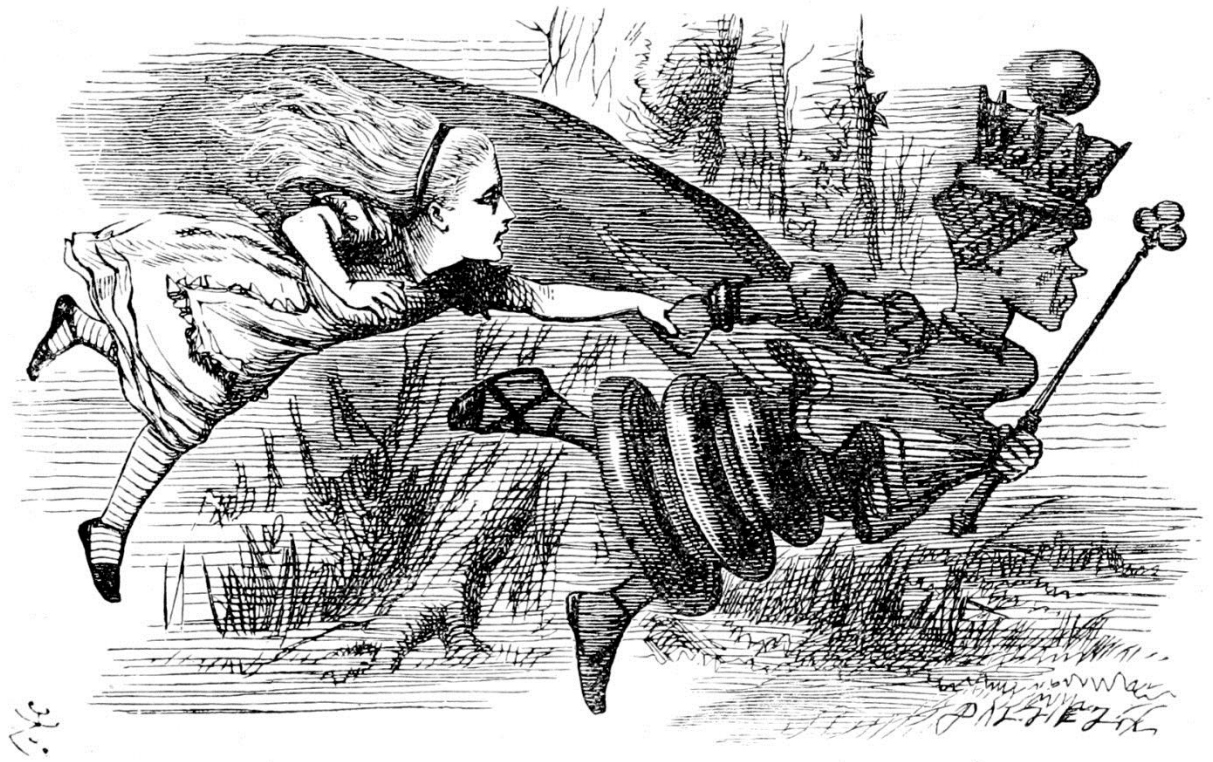


Figure 2.1: Gravure représentant la course d'Alice et de la Reine Rouge utilisée comme allégorie pour illustrer la coévolution entre organismes. "Alice looked round in great surprise. 'Why, I do believe we've been under this tree the whole time! Everything's just as it was!' 'Of course it is,' said the Queen, 'what would you have it?' 'Well in *our* country,' said Alice, still panting a little, 'you'd generally get to somewhere else if you ran very fast for a long time, as we've been doing.' 'A slow sort of country!' said the Queen. 'Now, *here*, you see, it takes all the running *you* can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!'" extrait du livre *Through the Looking Glass and What Alice Found There* (1874) par Lewis Caroll (1832-1898). Illustration par Sir John Tenniel (1820-1914).

Il existe différents types d'interactions qu'on distingue selon leurs conséquences sur la valeur sélective des individus impliqués et qui auront donc des effets différents sur l'évolution des partenaires. Certaines interactions sont neutres pour l'une des deux parties, mais impactent la seconde. C'est le cas du commensalisme, lorsque les conséquences sont positives sur la valeur sélective du second partenaire. Dans d'autres cas, les espèces impliquées tirent des bénéfices réciproques des interactions et voient leur valeur sélective positivement impactée. On parle alors de mutualisme ou encore, lorsque les relations entre les espèces sont obligatoires, de symbiose. Celles-ci peuvent avoir des conséquences très importantes sur l'évolution des partenaires. L'un des exemples les plus extrêmes et frappants est le cas des mitochondries, anciens organismes unicellulaires indépendants, ayant évolué en organites et ayant permis l'émergence des eucaryotes.

Mais toutes les interactions entre individus et espèces n'ont pas seulement des conséquences positives sur la valeur sélective des partenaires. L'amensalisme impacte négativement une espèce tout en n'ayant aucune conséquence sur l'autre. De même, la

compétition ou concurrence, interaction entre organismes partageant une même niche écologique et luttant donc pour un accès à des ressources limitées, abaisse la valeur sélective des deux espèces impliquées. Enfin, il existe aussi des types d'interactions dans lesquelles l'une des deux espèces subit une réduction de sa valeur sélective alors que celle de la seconde espèce augmente. C'est le cas de la prédation ou du parasitisme.

On distingue prédation et parasitisme notamment par le nombre d'individus touchés par un prédateur ou parasite. Ainsi, les prédateurs (comme les brouteurs) se nourrissent de nombreux individus, au contraire des parasites qui obtiennent leur nourriture d'un très petit nombre d'individus et souvent même d'un unique individu, dit hôte (Begon, 2009). D'autres critères sont parfois également utilisés comme la taille, puisque les parasites sont plus petits que l'hôte alors que les prédateurs sont fréquemment plus grands. On parle de pathogène, lorsque l'infection par un parasite cause des symptômes délétères à son hôte, c'est-à-dire lorsque l'infection donne une maladie à son hôte (Begon, 2009).

Les pathogènes et leur impact évolutif

Les pathogènes sont un groupe d'organismes ayant un impact très lourd, à la fois en termes de santé publique et en termes économiques, qu'il s'agisse de pathogènes humains ou infectant les animaux domestiqués et les plantes cultivées. Ils peuvent appartenir à de très nombreux et très divers groupes d'organismes : virus, bactéries, protozoaires ou encore champignons. La définition du parasite implique, comme mentionné précédemment, une baisse de la valeur sélective de son hôte. Celle-ci peut se manifester de différentes façons. Par exemple, l'infection peut modifier les capacités reproductrices de l'hôte, causant des stérilités ou des distorsions du ratio entre les sexes des descendants. Le pathogène peut également impacter la capacité de développement ou de survie de leur hôte, en le rendant plus susceptible à la compétition ou à la prédation, par exemple, ou tout simplement en ayant des conséquences létales.

En affectant la valeur sélective de leurs hôtes, les pathogènes modifient l'histoire évolutive de ces espèces. Ainsi, il a été montré que, du moins chez certaines espèces comme l'Homme, la pression de sélection induite par les pathogènes est l'une des forces évolutives les plus importantes (Fumagalli et al., 2011). Cependant, la relation de pathogénicité affecte non seulement l'évolution des hôtes mais aussi celle des pathogènes, en une boucle infinie de sélection réciproque s'exerçant sur les deux partenaires, la coévolution hôte-pathogène.

La coévolution entre hôte et pathogènes

De la coévolution à la différenciation chez les pathogènes

Le processus de coévolution implique une pression de sélection sur le pathogène pour obtenir des capacités d'infection plus importantes et sur l'hôte pour mieux reconnaitre et contrer son parasite. Il est important de noter que j'utilise ici ce terme de façon distincte de celui de co-spéciation et de co-divergence, bien que ces termes soient occasionnellement utilisés en tant que synonymes (De Vienne et al., 2013).

En cas de différenciation chez l'hôte, la coévolution peut faire que les spécialistes auront une meilleure valeur sélective que les généralistes ; dans ce cas, à condition qu'un isolement reproductif se mette en place, la coévolution peut mener à la spécialisation du pathogène, c'est-à-dire l'évolution de génotypes spécifiques de chaque lignée d'hôtes en divergence. Par ce phénomène de co-divergence, la spéciation chez les hôtes peut conduire à une spéciation chez les pathogènes associés, un phénomène nommé co-spéciation (De Vienne et al., 2013; Giraud, Refrégier, Le Gac, de Vienne, & Hood, 2008). En conséquence, il est souvent attendu que la phylogénie des pathogènes reflète la phylogénie de leur hôte (Giraud et al., 2008). Cependant, il a été montré que cet attendu n'est pas souvent vrai. Ainsi, il semblerait que la spéciation des pathogènes par saut d'hôte, c'est-à-dire l'apparition d'une espèce-fille de pathogène par « saut » et spécialisation sur un nouvel hôte, soit plus fréquente que la co-spéciation (De Vienne et al., 2013). La spécialisation par saut d'hôte reste un phénomène qui est lié à la coévolution entre hôtes et pathogènes (De Vienne et al., 2013).

La différenciation dans les couples hôtes-pathogènes a souvent été étudiée dans le contexte de la spéciation, cependant, comme je l'ai évoqué dans l'introduction de la première section de cette thèse, la spéciation n'est pas un phénomène tout ou rien, mais un processus progressif. La divergence entre lignées dans les couples hôtes-pathogènes peut donc également être étudié entre populations au sein même d'une espèce.

L'adaptation locale comme une mesure de la co-évolution

L'existence de populations différenciées chez les espèces d'hôtes cause des irrégularités spatiales dans l'environnement des pathogènes. En raison de grandes tailles de populations, de temps de génération généralement assez courts et de taux de mutations élevés, les pathogènes pourraient évoluer plus rapidement que leurs hôtes (Dybdahl & Storfer, 2003). Ceci conduit à la

possibilité de populations de pathogènes localement adaptées à leurs hôtes (Greischar & Koskella, 2007). L'adaptation locale est alors à la population ce que la spécialisation est à l'espèce.

L'adaptation locale est mise en évidence lorsque chaque population possède une meilleure valeur sélective dans son milieu d'origine que dans un autre milieu (Kawecki & Ebert, 2004). Cela peut être testé soit en comparant, dans un environnement donné, les performances d'individus issu de cet environnement (population locale) à celles d'individus provenant d'un autre environnement (population étrangère), soit en comparant les performances des mêmes individus dans son environnement local et dans un environnement étranger.

Dans le cas plus spécifique des organismes pathogènes, l'environnement correspond notamment à leurs hôtes. On peut alors mesurer leur adaptation locale à la population d'hôtes via, par exemple, la mesure de pouvoir infectieux. Le pouvoir infectieux est la capacité pour un pathogène à infecter des hôtes et est basé sur des stratégies d'évitement de la détection par ces hôtes (Dybdahl & Storfer, 2003). Le pouvoir infectieux est une mesure qualitative très utile pour les expérimentateurs car facilement mesurable par simple comptage du nombre d'hôtes infectés et il correspond à une forte différence de valeur sélective. On peut aussi mesurer la sévérité des symptômes, une mesure quantitative également importante en termes de valeur sélective.

Bases moléculaires de la coévolution hôte-pathogène

Pouvoir infectieux et sévérité des symptômes dépendent de la résistance de la plante et donc notamment de la capacité de son système immunitaire à reconnaitre le pathogène. La perception du pathogène par la plante déclenche des réactions locales et systémiques permettant à la plante de répondre à l'attaque et cette réponse se fait à plusieurs niveaux.

En plus de résistances constitutives telles que la présence de la paroi végétale, il existe deux niveaux de défenses induites chez les plantes en réponse aux maladies (Selin et al. 2016). Le premier niveau est la défense basale, qui se déclenche en réponse à l'infection causée par un large nombre de pathogènes. Ce niveau de défense peut être déclenché par les composés dérivés de la membrane des cellules après dégradation par un pathogène, par exemple, mais aussi par la reconnaissance de motifs moléculaires associés au pathogène (PAMP pour l'anglais « pathogen-association molecular patterns », Gururani et al 2012). Ces PAMP sont des molécules essentielles pour le pathogène qui ne peuvent ni être supprimées, ni mutées et sont fortement conservées à travers un grand nombre d'espèces de pathogènes. La reconnaissance des PAMP par des récepteurs chez les plantes initie une réponse de défense dite PTI pour l'anglais « PAMP-triggered immunity », c'est-à-dire une série de réactions bien souvent suffisantes à repousser l'infection (Boyd, Ridout, O'Sullivan, Leach, & Leung, 2013). Les micro-organismes pathogènes

expriment une série de molécules, dites « effecteurs », qui servent à contrer ces mécanismes de défense et à réussir à coloniser l'hôte. C'est la détection de ces effecteurs qui déclenche le second niveau de défense des plantes.

Ce second niveau de défense passe par une reconnaissance du pathogène par un récepteur de la plante hôte ; cette reconnaissance peut être directe, c'est-à-dire que le récepteur se lie à l'effecteur, mais aussi indirecte, quand le récepteur de la plante hôte reconnait, non pas l'effecteur lui-même, mais la version modifiée par l'effecteur d'une de ses propres protéines (par exemple, les protéines impliquées dans la réponse de défense basale, souvent prises pour cibles par les effecteurs des pathogènes ; Bomblies and Weigel 2007). Une réponse de défense dite ETI pour l'anglais « effector-triggered immunity » est alors mise en place. Ce type de défense, parfois nommée également « résistance aux pathogènes médiées par les gènes R », fonctionne généralement sur un modèle gène-pour-gène (Gururani et al., 2012), expliqué ci-dessous.

Le modèle gène-pour-gène

Le modèle gène-pour-gène semble être très répandu, notamment chez les plantes (Brown & Tellier, 2011). Dans les modèles gène-pour-gène les plus simples et les plus anciennement développés, la sélection s'applique à un unique locus avec deux allèles dans les deux partenaires (Flor, 1971). La plante a un allèle de résistance (RES), qui correspond généralement à une molécule reconnaissant un motif moléculaire particulier du pathogène, et un allèle de susceptibilité (S ou res). Le pathogène, quant à lui, a un allèle dit d'avirulence (A ou AVR) et un allèle de virulence (V ou avr) ; l'allèle de virulence correspond en général a une absence d'un motif moléculaire qui est reconnu par la plante comme une menace. La résistance est observée dans le seul le cas où la plante porte l'allèle R et le pathogène l'allèle A puisque la plante est alors capable de reconnaître le pathogène et se défendre. Dans tous les autres cas, soit la plante (S/res) est incapable de reconnaitre le pathogène, soit le pathogène échappe aux défenses de la plantes (V/avr). Des modèles plus complexes peuvent inclure plusieurs couples de gènes (Sasaki, 2000).

Selon le modèle gène-pour-gène, il existe deux issues possibles au processus de co-évolution. Dans la première, des cycles épidémiques successifs s'accompagnent du maintien à long terme du polymorphisme chez l'hôte comme chez le pathogène, avec des fréquences alléliques variant de façon cyclique, et le pouvoir infectieux avançant et se repliant dynamiquement à la façon d'une guerre des tranchées. A l'inverse, il peut arriver que les allèles de résistance et de virulence soient fixés et que la sélection s'exerce sur un nouveau couple de gènes. Cette issue conduit à une escalade du nombre de gènes de résistance et de virulence,

comparable à une course aux armements (Brown & Tellier, 2011; Summers et al., 2003). Cette course aux armements entre hôtes et pathogènes est, bien sûr, présente dans les systèmes naturels, mais elle n'est jamais aussi visible que dans les cas où les Hommes s'efforcent de les contrôler, comme lorsque l'usage abusif de multiples antibiotiques, conduit à l'apparition de nombreuses souches de bactéries pathogènes multi-résistantes (Williams, 2010) ou encore dans les agrosystèmes, avec l'accumulation de virulences dans les champignons pathogènes (Brown & Tellier, 2011).

Les pathogènes dans les agrosystèmes

Dans les populations naturelles, les gènes de résistance semblent avoir une longue durée de vie et la guerre des tranchées y est le scénario évolutif prédominant (Brown & Tellier, 2011). Cela n'est pas le cas dans les agrosystèmes où les gènes de résistance sont généralement contournés rapidement après leur dissémination (parfois seulement en quelques années) en raison de l'émergence de populations virulentes du pathogène visé (McDonald & Linde, 2002).

Scénarios évolutifs d'émergence des pathogènes dans les agrosystèmes

L'émergence des pathogènes dans les agrosystèmes a fait l'objet un article de synthèse par Stukenbrock and Mcdonald (2008). A travers leur compte-rendu de la bibliographie sur le sujet, ils identifient quatre mécanismes évolutifs majeurs mis en jeu dans l'émergence de nouveaux types de pathogènes. Les deux premiers mécanismes évolutifs sont également étudiés dans le cadre des populations naturelles de pathogènes et je les ai déjà évoqués plus haut. Il s'agit de la co-divergence (« host-tracking ») et des sauts d'hôtes. Ainsi, un saut d'hôte a été observé dans le cas de la pyriculariose du riz, par exemple. Dans ce saut d'hôte ancien, le pathogène, *Magnaporthe oryzae*, semble être passé d'un millet appartenant au genre *Setaria* vers le riz (Couch, 2005).

Le scénario de co-divergence implique des différenciations de lignées simultanées de l'hôte et du pathogène au cours du processus de domestication et du développement de l'agrosystème propre à l'espèce domestiquée. Au cours de la domestication volontaire de la plante hôte par l'Homme, le processus de sélection et de transformation de la plante comme de l'agrosystème a sélectionné les pathogènes les plus adaptés à ce nouvel environnement. Cela a conduit à ce que certains nomme une domestication involontaire ou encore « pestification » (De Gracia et al., 2015; Saleh, Milazzo, Adreit, Fournier, & Tharreau, 2014). Dans ce cas, on peut s'attendre à ce que l'étude de populations domestiques et sauvages chez l'hôte et le pathogène

révèle une histoire commune avec un même centre d'origine. Ce mécanisme de co-divergence ou host-tracking a été montré dans plusieurs cas de pathogènes de plantes domestiquées comme dans le cas du pathogène responsable de la tavelure du pommier, *Venturia inaequalis* (P Gladieux et al., 2010).

Deux autres des scénarios mentionnés par Stukenbrock et McDonald (2008) sur l'émergence des pathogènes dans les agrosystèmes sont liés aux flux de gènes. En premier lieu, les transferts horizontaux permettent à un organisme non ou peu pathogène de devenir virulent grâce à l'acquisition de gènes de virulence venant d'une autre espèce. Ce phénomène est bien connu chez les bactéries mais a également été identifié chez les champignons comme dans le cas de *Pyrenophora tritici-repenti*, un pathogène du blé, ayant obtenu un gène d'un autre pathogène, *Staganospora nodorum* (Friesen et al., 2006)). Le second mécanisme d'émergence de pathogènes lié aux flux de gènes, l'hybridation inter-spécifique, passe par la reproduction sexuée et n'implique pas toujours l'échange d'un petit nombre de gènes seulement, mais peut avoir des effets à l'échelle du génome complet. L'un des exemples les plus frappants et les plus récents est le champignon responsable de l'oïdium des céréales sur le triticale, *Blumeria graminis triticale*, émergeant depuis seulement 2001 et issu d'hybridations interspécifiques entre les pathogènes responsables de l'oïdium du blé, *B. g. tritici*, et du seigle, *B. g. secalis* (Menardo et al., 2015). De façon intéressante, le triticale est lui-même une céréale hybride entre le blé et le seigle, créé dans les années 1960.

Résistance chez les plantes cultivées

Malgré l'émergence possible de nouveaux pathogènes virulents sur les cultures anciennement résistantes, utiliser des gènes de résistance aux pathogènes dans les programmes d'amélioration variétale reste une alternative économique et écologique à l'usage de pesticides (Gururani et al., 2012). Bien que les résistances portées par les gènes R soient rapidement contournées par les pathogènes, les gènes R sont traditionnellement favorisés par les sélectionneurs dans les programmes d'amélioration variétale, probablement parce qu'ils s'expriment en un phénotype clair et dominant de résistance (Boyd et al., 2013). La recherche de gènes de résistance à différents pathogènes est donc très active pour un usage dans les programmes d'amélioration variétale et l'introgression de nouveaux gènes de résistance dans les variétés cultivées implique bien souvent des croisements avec des espèces apparentées, généralement sauvages (Boyd et al., 2013; Hajjar & Hodgkin, 2007). Ce fut le cas, par exemple, chez le pommier cultivé, pour lequel le gène de résistance à la tavelure Rvi6 (anciennement

appelé Vf), le plus largement introgressé dans les variétés issues de programme d'amélioration, fut initialement identifié chez un pommier non issu de l'agriculture (Lemaire et al., 2016).

II. Les pommiers et leur pathogène *Venturia inaequalis*

Introduction aux espèces étudiées : *Venturia inaequalis*

Le genre *Malus,* et en particulier le pommier domestique, ayant été introduit en détail dans la section précédente, je n'introduis ici que son pathogène. *Venturia inaequalis* est un champignon ascomycète responsable de la tavelure. Il infecte notamment de nombreuses espèces de pommiers, dont le pommier cultivé *M. domestica*, son ancêtre *M. sieversii*, et le pommier sauvage européen *M. sylvestris*. Sur les feuilles, la tavelure cause des taches brunes ou noirâtres provoquant des déformations de la feuille. Les fruits sont également affectés par les symptômes, induisant des lésions superficielles qui peuvent crevasser. C'est en raison de l'apparition de ces lésions que ce pathogène a une importance économique cruciale puisque les fruits atteints perdent une grande partie de leur valeur commerciale (Fig. 2.2). Or, de très nombreuses variétés sont sensibles à la tavelure incluant des variétés extrêmement populaires comme les variétés Golden Delicious et Gala. Une grande quantité de traitements anti-tavelure est donc nécessaire

Figure 2.2 : Symptômes de tavelure sur fruits (A) et feuilles (B). Photos Le Cam et photothèque INRA. Tirées de la thèse d'Amandine Le Van 2012

pour maintenir les récoltes intactes, ce qui en fait une maladie des pommes couteuse pour les arboriculteurs.

Historiquement, les premières mentions officielles de tavelure datent du début du XIXème siècle, cependant des traces de tavelure sont détectables bien avant cette époque. Ainsi, les pommes représentées sur le tableau du Caravage « Le souper à Emmaüs » peint vers 1601 ou 1602 sont tachées de tavelure (Fig. 2.3). MacHardy et al (2001) ont proposé une explication à l'absence de mention de la maladie précédant le début du XIXème siècle. Selon eux, il est possible que les modes de culture ne se soient pas prêtés à un développement extensif de la maladie. Durant le Moyen-Age en Europe, les vergers étaient essentiellement présents dans les monastères et des génotypes résistants et très différents étaient certainement utilisés, maintenant l'infection à des taux faibles. Ensuite, et jusqu'au XIXème siècle, les pommiers ont été cultivés souvent dans des prés (où les animaux d'élevage débarrassaient le sol des feuilles et fruits infectés), dans les haies ou bien dans des vergers de tailles modestes incluant d'autres espèces fruitières comme des poiriers ou des pruniers. Là encore, les variétés étaient certainement très nombreuses, possédant donc différentes résistances face au pathogène. Au XIXème siècle, des vergers à grande échelle ont commencé à se répandre. Pour répondre aux attentes plus spécifiques d'une clientèle plus nombreuse, les variétés présentes dans les vergers ont été homogénéisées et, pour maximiser la production, les arbres ont été plantés en rang. Ces

Figure 2.3 : « Le souper à Emmaüs » peint par Michelangelo Merisi da Caravaggio dit le Caravage (1571-1610). Les pommes représentées dans la coupe sont tachées de tavelure.

nouvelles habitudes de culture auraient alors permis un développement plus important de la tavelure qui aurait alors commencé à causer des dégâts économiques importants.

La tavelure est particulièrement problématique dans les régions où il pleut fréquemment au printemps. En effet, l'infection des pommiers par *V. inaequalis* se fait au printemps et au début de l'été par des ascospores (spores sexuées) qui sont relâchées lors d'évènements pluvieux. Ce moment coïncide parfaitement avec l'apparition sur l'hôte des jeunes bourgeons plus susceptibles à la maladie (MacHardy et al., 2001). Les ascospores sont expulsées sur une très petite distance, mais peuvent ensuite être dispersées sur quelques dizaines de mètres par le vent. Leur arrivée sur les tissus susceptibles d'un hôte marque un changement brusque d'une phase saprophytique à une phase parasitaire. Les ascospores germent alors et pénètrent dans l'organe hôte. Un mycélium sous-cuticulaire se forme et produit des conidiophores par plusieurs cycles de reproduction asexuée. Ce sont ces formations qui causent les symptômes apparents. Les conidies formées peuvent alors servir d'inoculum secondaire. A l'automne, les feuilles et pommes infectées tombent au sol où le parasite va se reproduire de façon sexuée et repasser dans

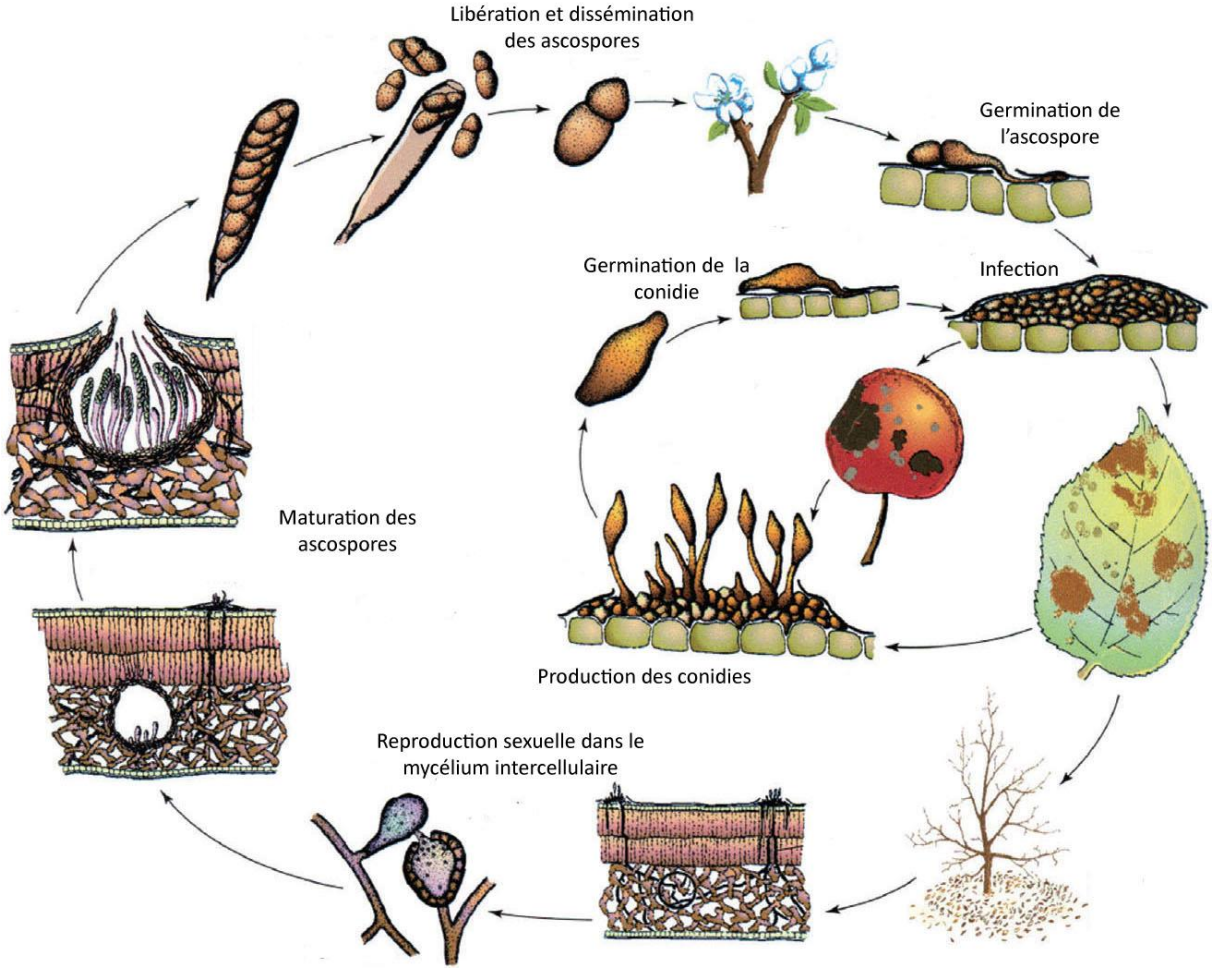


Figure 2.4 : Le cycle de vie de Venturia inaequalis. Figure adaptée de Bowen et al. 2011. Le diagramme fut initialement publié dans Agrios, Plant Pathology, p. 506. Copyright Elsevier 2005.

sa phase saprophytique. *Venturia inaequalis* passe alors l'hiver dans ces feuilles avant de commencer un nouveau cycle (Fig. 2.4) (Bowen et al., 2011; MacHardy et al., 2001).

Découverte de deux populations de V. inaequalis *en Asie Centrale*

Une étude de 2010 par Gladieux et al. a cherché à comprendre l'évolution de *V. inaequalis* en réponse à la domestication de son hôte. En utilisant un échantillonnage en Europe et en Asie centrale sur les trois espèces hôtes *M. domestica*, *M. sylvestris* et *M. sieversii*, trois groupes génétiques ont été identifiés. Un premier groupe en Europe infecte à la fois *M. domestica* et *M. sylvestris*. Deux autres groupes ont été trouvés en Asie centrale, l'un infectant les arbres *M. domestica* et *M. sieversii* poussant dans les agrosystèmes des plaines, et un autre sur des arbres sauvages *M. sieversii* isolés dans les montagnes. L'une des premières questions d'intérêt était l'origine des populations de *V. inaequalis* infectant les pommiers cultivés. La présence de la population de *V. inaequalis* dans les montagnes sauvages asiatiques sur le progéniteur initial du pommier domestiqué et dans son environnement originel suggérait que la population des plaines résultait d'une différenciation par « host-tracking » lors de la domestication du pommier. Ainsi l'évolution de *V. inaequalis* reflèterait-elle celle de son hôte : les populations ancestrales seraient originelles des montagnes d'Asie centrale, puis le pommier aurait subi un processus de différenciation due à la domestication, suivie par une différenciation chez le pathogène.

La migration en Europe du pommier cultivé a sans doute induit une différenciation en une lignée supplémentaire de *V. inaequalis* en Europe, qui a aussi infecté *M. sylvestris.* La faible différenciation entre la population européenne et la population des agrosystèmes d'Asie centrale rend en effet improbable une origine de la population européenne sur l'espèce hôte *M. sylvestris*. De même, la faible diversité dans cette population européenne n'est pas compatible avec une origine hybride impliquant une population ancestrale présente sur *M. sylvestris* et une population apportée avec le pommier cultivé. L'hypothèse la plus probable semble être que *M. sylvestris* est actuellement parasité par des souches de *V. inaequalis* importées en Europe avec le pommier cultivé, il s'agirait donc ici d'un saut d'hôte.

Le phénotype de V. inaequalis *a été affecté par le processus de « pestification »*

Des expériences d'inoculation (Le Van et al., 2012) utilisant des souches de *V. inaequalis* provenant des trois groupes génétiques identifiés précédemment a montré que : 1) L'espèce *M. domestica* est sensible à toutes les souches (entre 80 et 100% des hôtes infectés) à l'exception aux souches collectées sur *M. sieversii* dans le milieu sauvage (0% d'infection) ; 2) *M. sylvestris*,

lui, est résistant à toutes les souches (entre 0 et 10% d'infections), à l'exception des souches collectées sur lui-même (environ 20% d'infection) *; 3) M. sieversii* peut être infecté par l'ensemble des souches collectées, bien que les taux d'infection soient moyens (entre 30 et 80%).

La comparaison du pouvoir infectieux des souches collectées sur *M. sieversii* en milieu sauvage et de celles collectées en milieu agricole permet d'éclairer les changements évolutifs associés au processus de domestication et de culture de l'hôte à grande échelle. Le fait que les populations de pathogènes « sauvages » présentes sur le progéniteur du pommier cultivé soient incapables d'infecter *M. domestica* suggère que des gènes de résistances ont été acquis par le pommier au cours de sa domestication. Ces gènes pourraient provenir soit d'un processus de sélection artificiel de résistance à partir de certains individus *M. sieversii*, soit de flux de gènes par les hybridations avec différentes espèces sauvages le long de la route de la soie. Le pouvoir infectieux élevé des souches de types agricoles sur *M. domestica* et sur *M. sieversii* suggère que *V. inaequalis* a également acquis des gènes de virulence au cours de son évolution suivant la domestication du pommier, possiblement à la suite d'introgressions également.

Le pouvoir infectieux ne fut pas le seul caractère phénotypique affecté chez *V. inaequalis* par la « pestification ». En effet, Le Van et al. (2012) ont également montré que l'agressivité (variation quantitative du pouvoir pathogène) de la population de type agriculturale collectée en Asie Centrale était significativement supérieure à celle de la population de type sauvage collectée dans la même région. Une étude plus récente (De Gracia et al., 2015) a également montré que la taille des spores et les capacités de sporulation sont différentes entre ces deux populations.

Article 4. Conséquences de l'introduction des pommiers cultivés dans leur zone d'origine : une histoire sur l'hôte et sur son pathogène

Après avoir identifié l'existence de populations différenciées génétiquement et phénotypiquement en Asie Centrale selon le type d'environnement (agricole ou naturel), il était important d'étudier l'impact de l'introduction du pommier cultivé dans les zones montagneuses d'Asie Centrale sur les populations naturelles de *V. inaequalis* mais également de son hôte. En effet, les études exposées dans la première section de manuscrit ont montré que l'introduction du pommier cultivé en Europe a eu des conséquences évolutives sévères sur l'espèce de pommiers sauvages indigènes, puisque l'intégrité génétique de *M. sylvestris* est maintenant menacée par les flux de gènes avec les pommiers cultivés. L'étude de Le Van et al. (2012) détaillée ci-dessus a également montré que cette introduction a conduit à la différenciation du pathogène en une population européenne qui est capable d'infecter l'espèce *M. sylvestris* jusqu'alors très probablement exempte de tavelure, puisqu'on ne trouve pas de populations endémique à cette espèce.

Dans l'article que je présente ici et dont je suis co-premier auteur, nous nous sommes intéressés au contact secondaire entre populations de types agricoles et sauvages de pommiers et de *V. inaequalis* dans les populations naturelles des montagnes d'Asie centrale. Nous avons notamment cherché à comprendre si l'introduction du pommier cultivé menace l'intégrité génétique de *M. sieversii*, si le contact secondaire entre les deux types de pathogènes favorise les flux de gènes, si le pathogène de type agricole ou le pathogène de type hybride sont une menace pour les populations naturelles de *M. sieversii,* et enfin si nous pouvions détecter des incompatibilités génétiques entre les deux populations de *V. inaequalis*. Ces incompatibilités pourraient être des incompatibilités dites de Dobzhansky-Muller, conduisant à des baisses de viabilité ou de fertilité chez les hybrides, à cause d'interactions négatives entre allèles à des loci différents, ces allèles ayant été fixés dans chaque population lors d'une divergence sans flux de gènes (Johnson, 2008; C.-I. Wu & Ting, 2004).

Ma contribution à cet article a été de faire le génotypage et les analyses concernant les pommiers. J'ai également participé aux analyses de comparaison entre hôte et pathogènes ainsi qu'à la rédaction de l'article. Au moment d'écrire ce manuscrit, toutes les données prévues pour la version finale de l'article ne sont pas encore disponibles. En effet, une étude expérimentale est

en cours permettant de mesurer les taux d'infection des pathogènes hybrides sur pommiers cultivés.

Threat on the wild Asian apple trees by introduction and gene flow from domesticated apples and their pathogen

Feurtey A.*, Lemaire C.*, de Gracia M., Cascales M., Leroy T., Bellanger M-N., Expert P., Gouzy J., Bierne N., Giraud T.#, Le Cam B.#

*Co-first authors

#Co-supervisers

Introduction

With increasing globalization, worldwide human-mediated transport of domesticated organisms has increased the likelihood of secondary contacts between allopatric populations of both hosts and pathogens. In plants in particular, domesticated species are frequently able to hybridize with related wild species and especially with their progenitor species (Zeder 2006). Following potentially human-mediated secondary contact between wild and domesticated species, such hybridizations can thus lead to introgressions in both directions, which has important evolutionary outcomes, either beneficial or deleterious, on all the involved species (Ellstrand et al 2013). Wild-to-crop gene flow in plants, for instance, can both provide useful resistance to stresses (e.g. to harsh climate conditions in maize, Hufford et al 2013). On the other hand, crop-to-wild gene flow can generate weeds thriving among crops (as in beets, Ellstrand et al 2013) or decrease the fitness of wild species, possibly leading them to extinction (Todesco et al 2016).

A phenomenon less often recognized despite being at least as concerning is that secondary contact between crops and their wild relatives also foster gene flow between their pathogens. Such gene flow between previously isolated pathogen populations can indeed promote the emergence of new epidemiological characteristics such as expanded host range (Inderbitzin et al 2011), resistance to antimicrobials (Hanage et al 2009) and enhanced virulence (Stukenbrock et al 2012). Secondary contacts between crops and their wild relatives may thus favor disease emergence and are a major threat not only for crop health, but also in agrosystems (Leroy et al 2014, Leroy et al 2016).

Apple is the most cultivated fruit in temperate areas in the world and consequently is very important economically in these areas. During the 1920's, Vavilov explored Central Asian mountains and found forests areas of wild apple trees, *Malus sieversii*, harbouring a great phenotypic diversity. He thus suggested that this area was the center of origin of the domesticated apple tree, *M. domestica*, and that *M. sieversii* was its progenitor (Harris et al 2002). This scenario has been confirmed since, in particular by the analysis of 23 nuclear genes in several *Malus* species (Velasco et al 2010). Archeological and molecular data indicated that the domesticated apple tree was imported in Europe via the Silk Roads, thus crossing the distribution range of different wild apple tree species (Cornille et al 2014). Because of the weak interspecies reproductive barriers in the *Malus* genus, gene flow occurred between *M. domestica* and various wild species. This has led to introgression of the cultivated apple by additional wild species than their initial progenitor, as shown using both nuclear and chloroplastic markers (Amandine Cornille et al 2012, Nikiforova et al 2013). In particular, the European crabapple *M. sylvestris* has been an important contributor to the modern cultivated varieties (Cornille et al. 2012). In turn, *M. sylvestris* is threatened by massive introgressions from *M. domestica* (Cornille et al 2015). More recently, apple varieties have been transported and cultivated all around the world, including in the area surrounding *M. sieversii* forests in Central Asia.

These contacts between cultivated apple trees and their wild relatives have also fostered gene flow in their pathogens. *Venturia inaequalis* is a fungal pathogen of apples with major economic impact, causing apple scab in the form of grey-brown lesions on leaves and fruits. It parasitizes not only *M. domestica* but also various wild apple tree species, including *M. sylvestris* and *M. sieversii*. However, *V. inaequalis* displays genetic subdivision according to host and geography (Gladieux et al 2010). The wild *M. sieversii* trees from Central Asian Mountains were found parasitized by a specific population (called CAM), genetically differentiated from the Central Asian population from the plains (called CAP), occurring on both *M. sieversii* and *M. domestica* (Gladieux et al. 2010). The European strains were found to form a third differentiated population, occurring both on *M. sylvestris* and *M. domestica* (Gladieux et al. 2010). The CAP and European populations have been suggested to have arisen by host tracking following apple domestication (Gladieux et al. 2010). Experimental inoculations revealed that, while the Asian wild apple M. sieversii could be parasitized by all tested *V. inaequalis* populations, *M. domestica* could only be colonized by its own "agricultural" populations either from Europe or Central Asia (CAP; Le Van et al 2012). This suggested a "pestification" of *V. inaequalis* on cultivated apple trees, having evolved new determinants of pathogenicity in addition to keeping the original ones (de Gracia et al 2015).

The fact that both domesticated apple trees and the agricultural form of *V. inaequalis* nowadays occur in close contact with wild populations in Central Asia raises questions about the consequences of this secondary contact, in particular in terms of introgression and spread and severity of the disease. Despite the importance of the wild Asian apple trees both in terms of conservation of endangered wild species and of valuable genetic resources for future breeding programs, these questions have not been studied so far. Apple species are known to have weak interspecific barriers, which allows introgressions into crabapples from cultivated forms. In *V. inaequalis*, host specificity is a strong prezygotic barrier to gene flow because the fungus mates within the apple leaves and therefore can only mate with strains able to colonize the same tree (Giraud et al 2010, Pierre Gladieux et al 2011). This means that genetic differentiation can occur without evolution of intrinsic prezygotic barriers such as mate choice (Giraud et al. 2010, Gladieux et al. 2011, le Gac et al; 2008). Such a lack of intrinsic barriers can lead to gene flow if host specificity barriers are disrupted, e.g., due to hybridization between host trees (Leroy et al 2016), leading to a kind of secondary contact. Disruption of geographical isolation can also lead to a secondary contact favoring gene flow between previously isolated populations. In such cases of secondary contact, gene flow can still be restricted by genetic incompatibilities having evolved during the time when mating was prevented (Coyne and Orr 2004). Such genetic incompatibilities can impair gene flow either along the whole genome, or only around the loci controlling the genetic incompatibilities if recombination is abundant enough and the incompatible loci not too numerous (Roux et al. 2013).

In the present study, we therefore addressed the following questions: 1) Does introduction of domesticated apple trees threaten the genetic integrity of Asian wild apples via introgressions? 2) Does it promote gene flow in *V. inaequalis* pathogens, between the agricultural (CAP) and the wild (CAM) populations? 3) In this case, does it pose another threat on Asian wild apples through the disease? 4) Does the pattern of the disease, i.e., what pathogen genotype can occur on what tree genotype, inform on the existence of ecological barriers to gene flow?

Materials and methods

Sampling

Leaves or fruit skin pieces with visible scab lesions were collected from 245 trees in Kazakhstan in 2006. They came from nine different sampling sites, mostly in wild forests. Two

of these sites were however particular, being an orchard of cultivated apple trees and the forest surrounding this orchard, respectively. This orchard was around 30 or 40 year-old. *Venturia inaequalis* individuals were collected on the plant material that was genotyped here, with in some cases multiple fungal lesions sampled on the same host tree.

Genotyping Malus *trees*

DNA was extracted from plant tissue with the NucleoSpin plant DNA extraction kit II® (Macherey & Nagel). Multiplex microsatellite PCR amplifications were performed with a Multiplex PCR Kit® (QIAGEN, Inc.). We used 33 microsatellite markers spread across the 17 chromosomes using 10 different multiplex reactions as previously described (Cornille et al 2012). The genotyping was done at the *Gentyane* platform (INRA). Only the microsatellite markers with less than a third of missing data were used and only individuals with less than half of missing data were kept for the analyses. The dataset was finally composed of the genotypes for 28 microsatellite markers for 240 apple tree samples.

Genotyping Venturia inaequalis

For each *V. inaequalis* sample, monospore isolation was performed and cultivated on malt-agar milieu during 7 to 10 days at 17°C. DNA was extracted from 30 to 40 mg of mycelium according to the protocol described previously (Leroy et al. 2016). A total of 270 strains were eventually available for genotyping.

All strains were genotyped at one non-synonymous SNP in each of 192 genes among those displaying the highest F_{ST} between agricultural and wild *V. inaequalis* populations in a previous study (Lemaire & de Gracia *et al.* in prep). In addition, SNPs were selected to be located on different scaffolds of the reference genome (B04) in order to avoid any physical linkage effect.

All samples were genotyped at the 192 SNPs using the KASP method (KBioscience competitive allele-specific polymerase chain reaction assay). All SNP with more than 10% missing data were removed from the analysis. The dataset was thus composed of the genotypes for 181 SNPs from 269 fungal samples.

Since fungi were sampled on the genotyped leaves, we had genotypes for host and pathogen pairs. In total, this added up to genotypes obtained for 263 hosts-pathogens pairs, including 197 unique hosts.

Analyses of diversity, differentiation and hybridization

In order to identify the hybrid status of the apple tree samples, we used reference pure genotypes previously assigned to genepools of *M. domestica* (*N*=50) or *M. sieversii* (*N*=25) (Cornille *et al.* 2012). For each host and pathogen genotype, we computed the "hybrid index" implemented in the add-on R-package *introgress* (Gompert and Buerkle 2010). This "hybrid index" is an estimate of the proportion of alleles that were inherited from one of two hybridizing parental populations, i.e. the level of ancestry from one of two populations. In our case, pure *M. domestica* trees and pure agricultural-type (i.e., CAP) pathogen genotypes had a "hybrid index" of one while pure *M. sieversii* and wild-type pathogen (i.e., CAM) genotypes had a hybrid index of zero. This index therefore represents the proportion of assignment to the *M. domestica* or agricultural type pathogen genepool and was thus called hereafter P_{dom}. The genomic cline analysis was also done with this R package.

The F_{ST}, F_{IS}, H_E and H_O were computed with Genepop (Raymond and Rousset 1995; Rousset 2008). Principal component analyses (PCA) were made with the ade4 R package. Allelic richness (A_r) was estimated with ADZE v.1.0 (Szpiech *et al.*, 2008).

Results

Genetic diversity and differentiation in apple trees

The P_{dom} index estimated in the 240 tree samples ranged between one (i.e. "pure" *M. domestica* genotypes) and zero (i.e. "pure" *M. sieversii* genotypes), with also individuals presenting intermediate values (i.e. hybrids); the pattern was however quite L-shaped, with over-representation of pure wild *M. sylvestris* (Fig. 1).

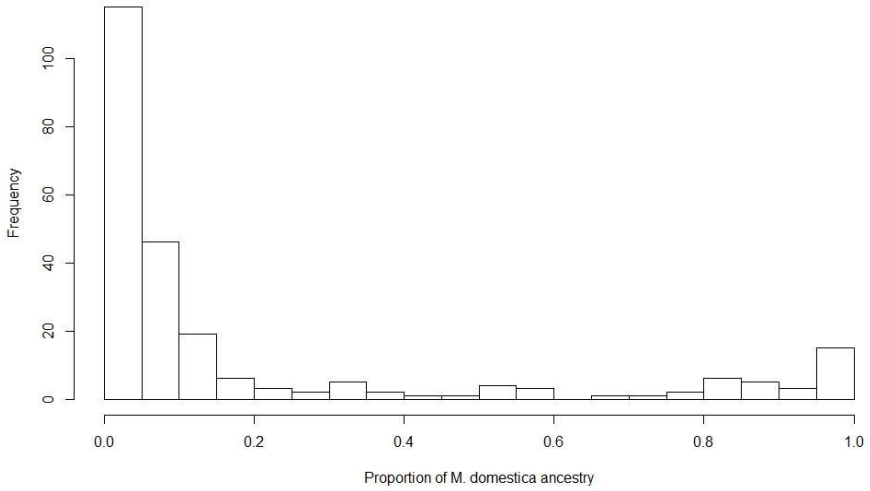


Figure 1: Distribution of proportion of P_{dom} in the 240 tree samples

The first axis of the principal component analysis also separated individuals according to their degree of ancestry in the two group and both analyses gave congruent results (Fig. 2).

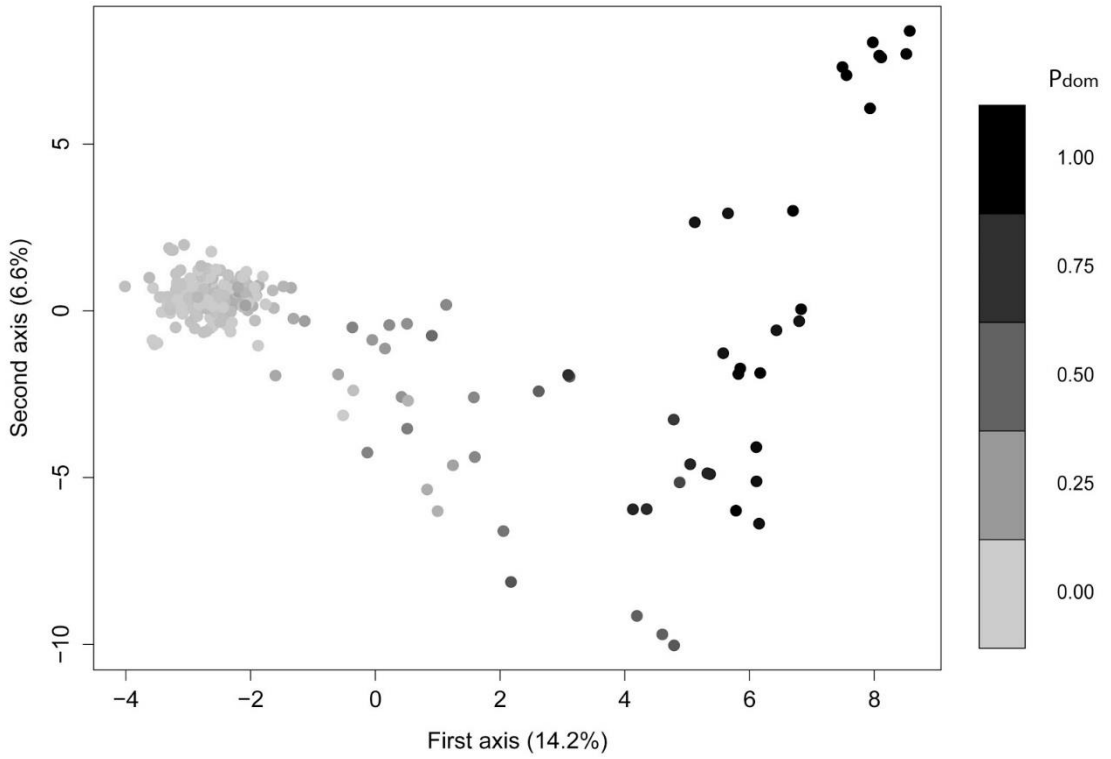


Figure 2: Principal component analysis of the 240 Malus *tree sample. The dots are different intensities of grey according to the measured P_{dom} (proportion of assignation to* M. domestica*) of each corresponding sample.*

Among the 240 tree samples, most had a low proportion of *M. domestica* ancestry (N=186, i.e. 77.5%, with $P_{dom} < 0.2$); three populations in particular included only such pure wild trees (Fig. 3). Only 29 samples (i.e., 12.1%) had a high proportion of *M. domestica* ancestry ($P_{dom} > 0.8$), 28 of them originating from the orchard and only one from another population. This left 25 samples (i.e. 10.4%) identified as hybrids across all the populations and 17 (i.e. 9.04%) when removing the orchard and the forest directly surrounding it.

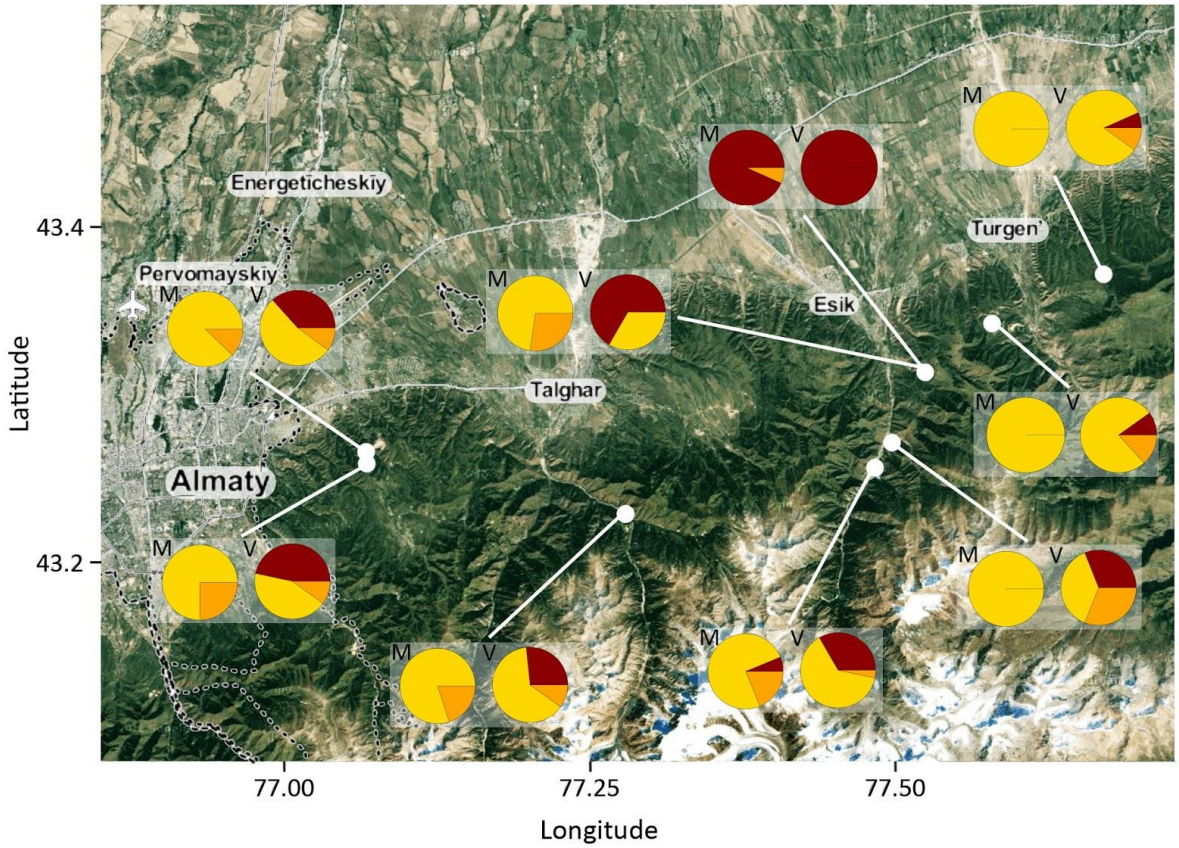


Figure 3: Proportions of wild (yellow), domesticated/agricultural (red) and hybrid (orange) genotypes per sampling site for the Malus *trees (left pie chart in each box) and for its fungal pathogen* Venturia inaequalis *(right pie chart in each box).*

F_{ST} values confirmed that the orchard population was different from the other apple tree populations: the mean F_{ST} between the orchard and the other populations was 0.16, while the mean F_{ST} among the other populations was only 0.02. In the orchard, among the 30 sampled trees, only 13 different genotypes were identified (the two most frequent genotypes being found 8 and 5 times, respectively). This suggests multiple grafting of clonemates.

The genetic diversity estimated by allelic richness standardized for a population size of 10 was the highest in the hybrid genepool and the lowest in the cultivated trees (4.37 for *M. domestica* trees, 4.89 in *M. sieversii* trees and 5.67 in hybrids).

Genetic diversity and differentiation in the fungus Venturia inaequalis

An overall of 270 *V. inaequalis* strains were genotyped at the 181 SNPs. As expected from their design, the 192 selected SNPs had high F_{ST} values (>0.4; see Fig. 4) between the reference (pure) wild (CAM) and agricultural (CAP) populations.

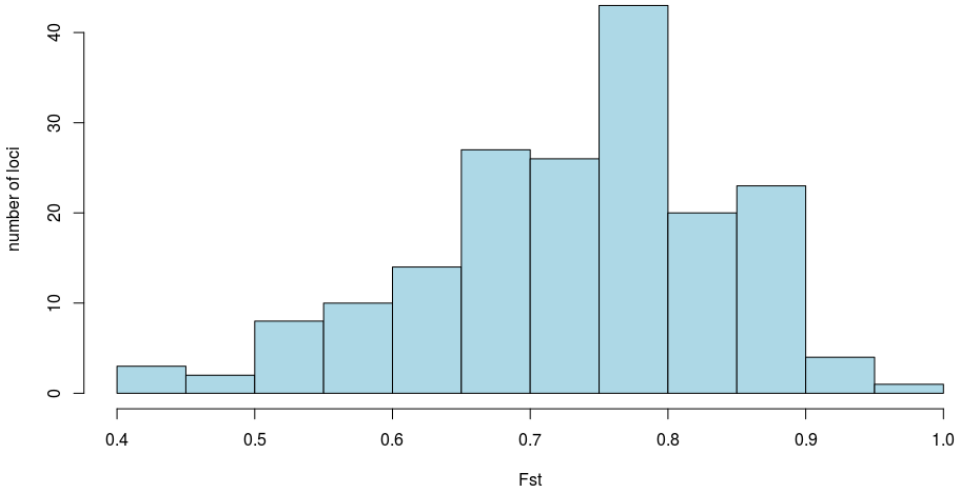


Figure 4: Distribution of F_{ST} between wild and agricultural Venturia inaequalis *reference populations (respectively CAM and CAP) at sites selected for SNP genotyping*

The PCA shows that most of the sampled strains mostly clustered either close to the pure reference wild population, or to the pure the reference agricultural (Fig. 5). Only ca. 30 strains appeared intermediate and were thus likely hybrids.

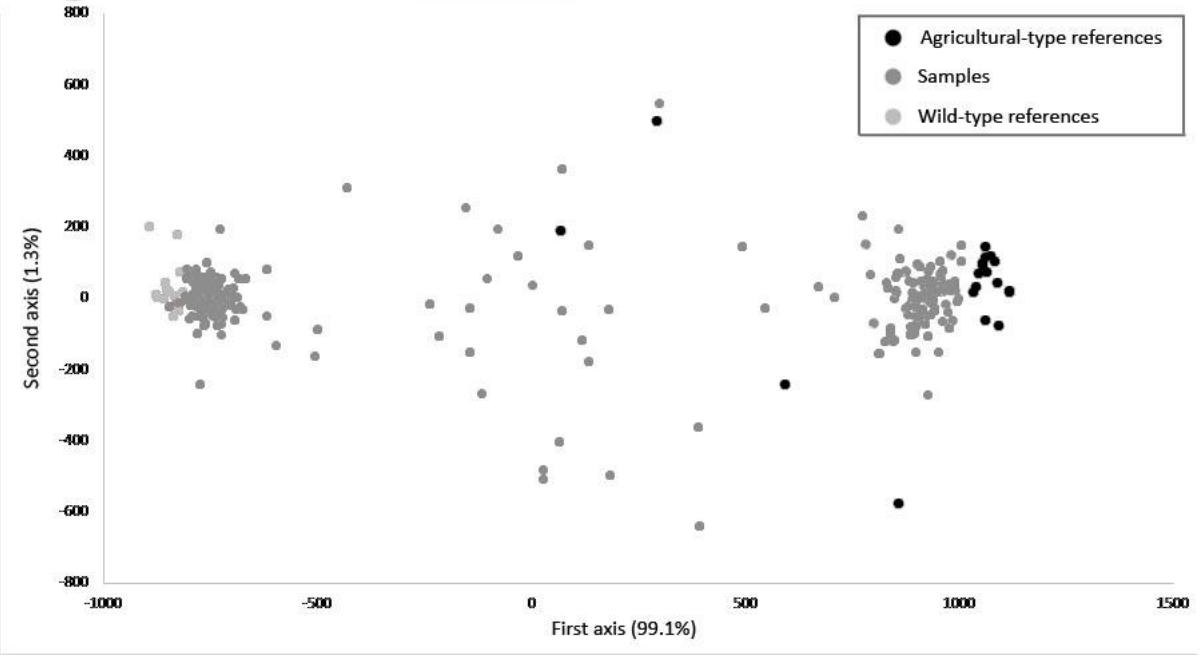


Figure 5: Principal component analysis on the 270 sampled strains of fungal pathogen Venturia inaequalis *and ot the agricultural and wild references (CAP and CAM, in light grey and black, respectively).*

The distribution of the hybrid index estimated using the INTROGRESS package (*i.e.* the proportion of agricultural ancestry) was U-shaped, confirming that only few hybrids were found (Fig. 6). They were found in all the sampling sites, except, interestingly, in the orchard of *M. domestica* and its surrounding forest of *M. sieversii* (Fig. 3). All the strains found in the orchard carried agricultural alleles, but the strains sampled on the surrounding forest belonged either to pure agricultural or wild type.

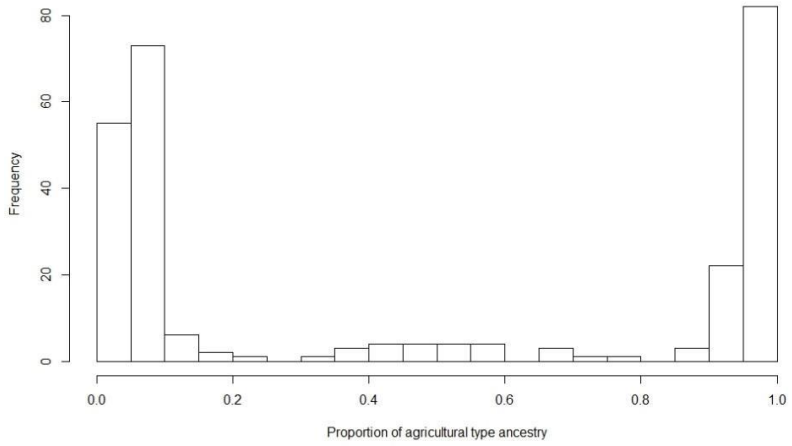


Figure 6: Distribution of proportion of agricultural ancestry among the 270 Venturia inaequalis *strains genotyped at 181 SNPs.*

A genomic cline analysis, assigning each SNP in each strain to one of the two references populations, also supported the lack of widespread gene flow among agricultural and wild type strains (Fig. 7). Indeed, few strains carried both wild and agricultural-type alleles (the *ca.* 30 strains identified above as hybrids). The genomic cline in the right of Fig. 7 indeed shows a sharp change in proportion of agricultural ancestry among strains, indicating little gene flow. In the few hybrids, not all SNPs were of the same agricultural or wild type.

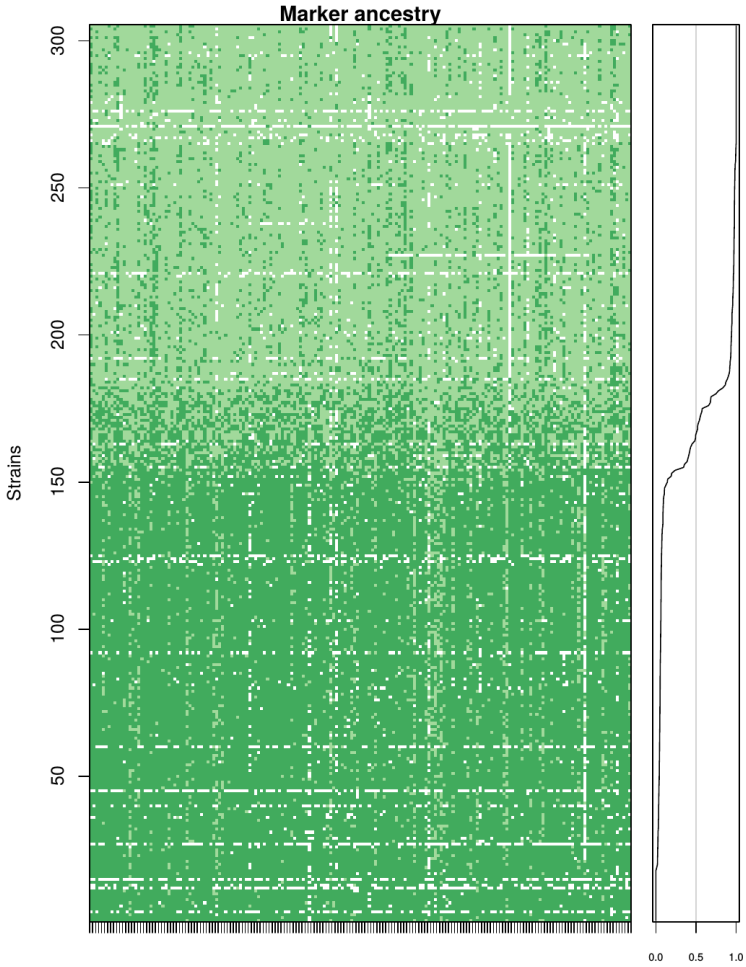


Figure 7: Assignation of each SNP to the two references populations in the 270 Venturia inaequalis *strains sampled in this study and in the strains from the wild and agricultural reference populations (rows) at 181 SNPs (columns). Wild type alleles are indicated in dark green and agricultural ones in light green. A genomic cline is drawn on the right side, giving the distribution of proportion of agricultural-type alleles in strains.*

Comparison of the host/pathogen per pair

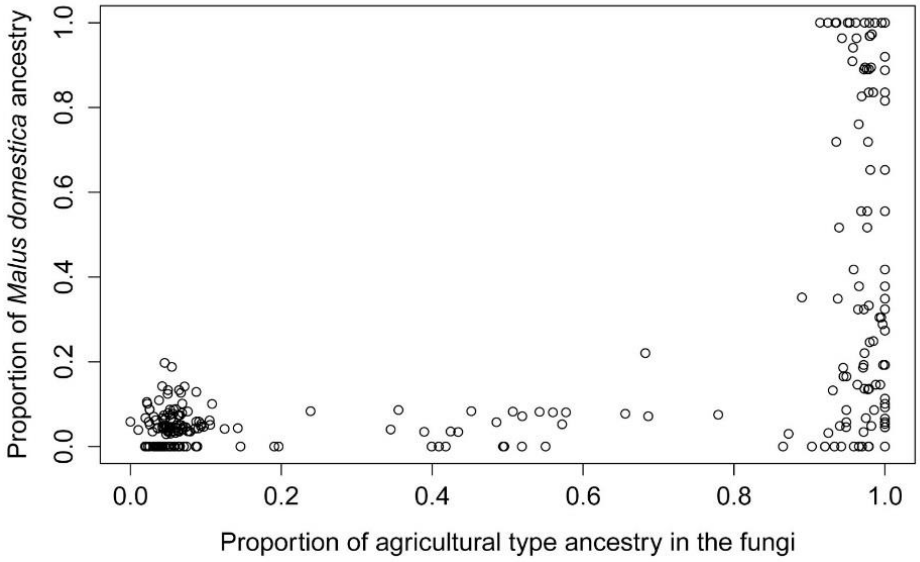


Figure 8: For each host-pathogen pair, proportion of the agricultural type ancestry in the fungal Venturia inaequalis *strain against the proportion of* Malus domestica *ancestry of the host.*

Plotting the P_{dom} values of the host against those in the pathogen for the 263 pairs gave a striking L-shaped pattern (Fig. 8): although hybrids were found in both the trees and the fungi, we did not find a pattern where hybrid pathogens infect hybrid hosts as could have been expected following the example of the hybrid *Melamspora x columbiana* on hybrid poplar (Newcomb et al., 2000). Instead, we found that *M. domestica* trees (i.e. with $P_{dom} > 0.8$) and hybrid apple trees (except for one with a low P_{dom} value) were parasitized only by agricultural-type pathogens (i.e. with $P_{dom} > 0.8$). Wild-type fungi (e.g. with $P_{dom} < 0.2$) and hybrid fungi were only found on wild apple trees.

Two of our populations were of specific interest, corresponding to an orchard of cultivated apple trees and the other to apple trees from the forest directly surrounding the orchard. Interestingly, although this area was the situation of closest contact between wild and agricultural forms in our sampling, we found hybrids only among the trees (25% percent of the forest trees and less than 10% of the orchard trees) but not among the fungi (Fig. 9). Within the orchard, only the agricultural type of fungi was found while both agricultural and wild types were found in the surrounding forest. Therefore, even in this situation of close contact between both host and pathogen wild and agricultural types, only the pure agricultural fungi could infect trees that were not pure *M. sylvestris*.

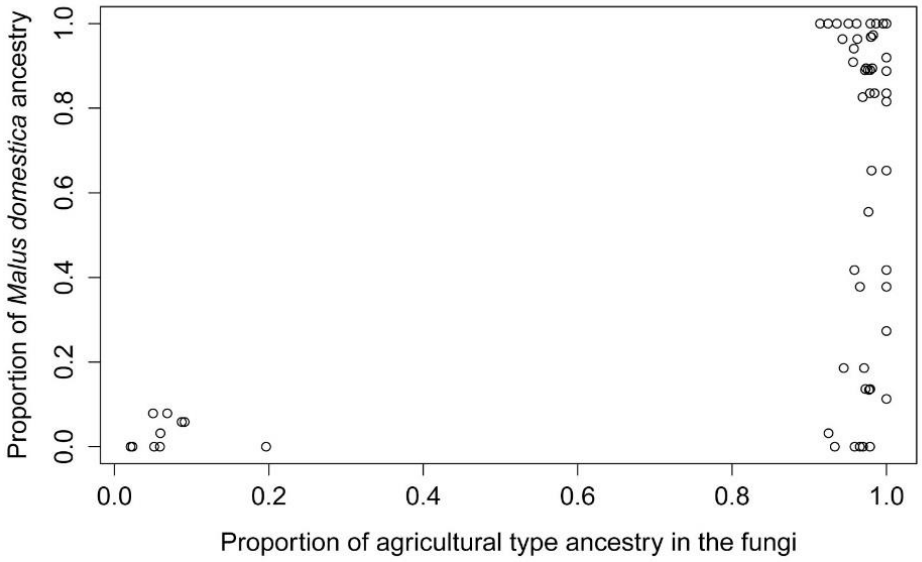


Figure 9: For each host-pathogen pair, proportion of the agricultural type ancestry in the fungal Venturia inaequalis *strain against the proportion of* M. domestica *ancestry of the host, in the case of the orchard and surrounding forest.*

Discussion

Natural populations of M. sieversii *are affected by gene flow from* M. domestica

In the case of apple trees, the only individuals with a high *M. domestica* ancestry (except for one) were found in the orchard. When excluding the orchard and the immediate surrounding forest, hybrids were found in four of the seven sampling sites. In the forest population surrounding the orchard, i.e. the most direct case of secondary contact in our sampling, around a quarter of the trees were identified as hybrid. This suggests lack of reproductive isolation between these apple tree species; the *Malus* genus is indeed known for its weak interspecific reproductive in barriers (Zohary and Hopf 2000).

Overall, it seems that the natural apple tree populations are partially and irregularly affected by gene flow following the secondary contact. This is in contrast to what has been observed in Europe where the European crabapple tree populations have been found massively introgressed by the cultivated apple tree to the point where it may be threatened of extinction (Cornille et al. 2015). The difference between the two situations can probably be explained by several components. First the difference in tree density in populations of the two wild species likely plays a role. Indeed, it has been shown that the pollination distances decrease with density around the pollinated trees in the European wild apple trees and longer pollination distances

increase the risk of interspecific mating with cultivated trees (Reim et al. 2015, Feurtey et al. in prep). *Malus sylvestris* is a light-demanding species and is found at a low density at forest edges, in clearings or in hedges (Schnitzler et al 2014). In contrast, *M. sieversii* grows in denser populations in Central Asian forests. High densities of *M. sieversii* over large areas in mountains would merely prevent pollination from domestic apple trees found here at low densities. In addition, except in the orchard, most cultivated apple trees were grafted on *M. sieversii* rootstocks. Thus dissemination through root sprouts will only produce *M. sieversii* trees, reinforcing isolation of *M. domestica* individuals.

It is also possible that the relative purity of some wild populations in Central Asia can be explained by a difference in timing between the two secondary contacts and the extent of the contact between wild and cultivated species. The cultivated apple tree was introduced in Europe before the 3rd century BC (Cornille et al. 2014) while the secondary contact in our sampling area is probably much more recent (estimated to *ca.* 100 years for *V. inaequalis* in Lemaire & de Gracia et al., in prep). Furthermore, European apple trees are cultivated in gardens and orchard in fragmented agrosystems interspaced with hedges and woods where *M. sylvestris* grows. In contrast, the orchard sampled in Central Asia is only 30 or 40 year-old and the rest of the sampled sites are still isolated from agricultural and heavily inhabited areas. If these differences in time and anthropogenic disturbances of the landscape are part of the explanation for the lack of observed hybridization, then it may be only a question of time before the situation deteriorates in Central Asia.

Invasion of the wild apple tree pathogen populations by agricultural-type V. inaequalis

The outcome of the secondary contact was found different for *V. inaequalis* and for its host. As expected, the wild-type pathogens were found in all the wild-area sampling sites, i.e. all the sites except the orchard where no wild-type pathogens were found despite the occurrence of wild hosts around. More surprisingly considering that most of the sampling sites were quite remote, the agricultural-type pathogens were also found in all of them, with an average of 30% of all pathogens when excluding the orchard.

When comparing host-pathogen pairs an interesting pattern emerged: the wild-type and hybrid fungi only parasitized the *M. sieversii* trees, while the agricultural-type was able to infect all apple trees regardless of their ancestry. Even when considering only the hybrid genotypes, the pattern held: a pathogen needs to be of pure agricultural ancestry to be able to parasite trees

that are even only partly from *M. domestica* ancestry. However, it should be noted that we had few trees with *M. domestica* ancestry and that a further sampling should be performed to assess whether this pattern is general. The pattern we found is likely due to the recurrent selection of resistance genes in cultivated apple trees, against which agricultural pathogens have progressively acquired virulence genes. The history of *M. domestica* domestication was indeed punctuated by introgressions from different wild species along the Silk Roads (Cornille et al 2014). Because of its evolution by host-tracking, it is possible that *V. inaequalis* has had the same history, picking up virulence genes in other pathogens populations as its hosts were acquiring resistance genes from wild species.

Reproductive isolation between the fungal populations without ecological barriers

Despite the pervasive presence of agricultural-type fungi in all the forest populations, which constitute many secondary contact zones, the proportion of hybrid fungi was found relatively low (around 10% over all populations). This suggests that there is at least partial reproductive isolation between the two fungal populations. Such reproductive isolation cannot be explained by geographical isolation since we found both pathogen types in all forest sampling sites. We also observed that both types are able to infect the same host species, *M. sieversii*. We also found instances (N=7) where the difference between two pathogens P_{dom} sampled on the same individual host trees was more than 0.8, indicating that the pathogens were of two different types. The reproductive isolation can thus not be explained by ecological isolation (i.e. which would occur if the two types simply never grew in the same leaves where they mate) since they shared the same hosts (down to the individuals). Reproductive isolation may thus result from intrinsic premating barriers or to post-mating barriers in the form of an accumulation of Dobzhansky-Müller incompatibilities in the fungal populations during their divergence that would prevent F1 individuals to thrive in this haploid fungus.

Overall, we found that the natural populations of *M. sieversii* are partially affected by gene flow from *M. domestica* that was introduced around a century ago in the area, but only a few pure *M. domestica* trees were found in the wild forests. In contrast, the wild populations of *V. inaequalis* are heavily invaded by the agricultural-type. The proportion of hybrids, however, is surprisingly similar, hinting at the existence of reproductive barriers between the two types of pathogens.

References

Cornille, A., Feurtey, A., Gélin, U., Ropars, J., Misvanderbrugge, K., Gladieux, P., & Giraud, T. (2015). Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: A basis for conservation and breeding programs in apples. *Evolutionary Applications*, *8*(4). http://doi.org/10.1111/eva.12250

Cornille, A., Giraud, T., Smulders, M. J. M., Roldan-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, *30*(2), 57–65. Journal Article. http://doi.org/10.1016/j.tig.2013.10.002

Cornille, A., Gladieux, P., Smulders, M. J. M., Roldan-Ruiz, I., Laurens, F., Le Cam, B., … 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 Genetics*, *8*(5). Journal Article. http://doi.org/10.1371/journal.pgen.1002703

Coyne, Jerry A and Orr, H. A. and others. (2004). *Speciation* (Vol. 37). Sinauer Associates Sunderland, MA.

De Gracia, M., Cascales, M., Expert, P., Bellanger, M. N., Le Cam, B., & Lemaire, C. (2015). How Did Host Domestication Modify Life History Traits of Its Pathogens? *Plos One*, *10*(6), e0122909. Journal Article. http://doi.org/10.1371/journal.pone.0122909

Ellstrand, N. C., Meirmans, P., Rong, J., Bartsch, D., Ghosh, A., de Jong, T. J., … Hooftman, D. (2013). Introgression of Crop Alleles into Wild or Weedy Populations. In D. J. Futuyma (Ed.), *Annual Review of Ecology, Evolution, and Systematics, Vol 44* (Vol. 44, pp. 325–345). Book Section. http://doi.org/10.1146/annurev-ecolsys-110512-135840

Gérard, P. R., Husson, C., Pinon, J., & Frey, P. (2006). Comparison of Genetic and Virulence Diversity of Melampsora larici-populina Populations on Wild and Cultivated Poplar and Influence of the Alternate Host. *Phytopathology*, *96*(9), 1027–36. http://doi.org/10.1094/PHYTO-96-1027

Giraud, T., Gladieux, P., & Gavrilets, S. (2010). Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology and Evolution*, *25*(7), 387–395. http://doi.org/10.1016/j.tree.2010.03.006

Gladieux, P., Vercken, E., Fontaine, M. C., Hood, M. E., Jonot, O., Couloux, A., & Giraud, T. (2011). Maintenance of fungal pathogen species that are specialized to different hosts: Allopatric divergence and introgression through secondary contact. *Molecular Biology and Evolution*, *28*(1), 459–471. http://doi.org/10.1093/molbev/msq235

Gladieux, P., Zhang, X. G., Roldan-Ruiz, I., Caffier, V., Leroy, T., Devaux, M., … Le Cam, B.

(2010). Evolution of the population structure of Venturia inaequalis, the apple scab fungus, associated with the domestication of its host. *Mol Ecol*, *19*(4), 658–674. Journal Article. http://doi.org/10.1111/j.1365-294X.2009.04498.x

Gompert, Z., & Buerkle, C. A. (2010). INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, *10*(2), 378–384. Journal Article. http://doi.org/10.1111/j.1755-0998.2009.02733.x

Hanage, W. P., Fraser, C., Tang, J., Connor, T. R., & Corander, J. (2009). Hyper-Recombination, Diversity, and Antibiotic Resistance in Pneumococcus. *Science*, *324*(5933), 1454–1457. http://doi.org/10.1126/science.1171908

Harris, S. A., Robinson, J. P., & Juniper, B. E. (2002). Genetic clues to the origin of the apple. *Trends in Genetics*, *18*(8), 426–430. Journal Article. http://doi.org/10.1016/s0168-9525(02)02689-6

Hufford, M. B., Lubinksy, P., Pyhaejaervi, T., Devengenzo, M. T., Ellstrand, N. C., & Ross-Ibarra, J. (2013). The Genomic Signature of Crop-Wild Introgression in Maize. *Plos Genetics*, *9*(5). Journal Article. http://doi.org/10.1371/journal.pgen.1003477

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*(3). http://doi.org/10.1371/journal.pone.0018260

Le Van, A., Gladieux, P., Lemaire, C., Cornille, A., Giraud, T., Durel, C.-E., … Le Cam, B. (2012). Evolution of pathogenicity traits in the apple scab fungal pathogen in response to the domestication of its host. *Evolutionary Applications*, *5*(7), 694–704. Journal Article. http://doi.org/10.1111/j.1752-4571.2012.00246.x

Leroy, T., Caffier, V., Celton, J.-M., Anger, N., Durel, C.-E., Lemaire, C., & Le Cam, B. (2016). When virulence originates from nonagricultural hosts: evolutionary and epidemiological consequences of introgressions following secondary contacts in Venturia inaequalis. *New Phytologist*, *210*(4), 1443–1452. http://doi.org/10.1111/nph.13873

Leroy, T., Le Cam, B., & Lemaire, C. (2014). When virulence originates from non-agricultural hosts: new insights into plant breeding. *Infect Genet Evol*, *27*, 521–529. Journal Article. http://doi.org/10.1016/j.meegid.2013.12.022

Nikiforova, S. V., Cavalieri, D., Velasco, R., & Goremykin, V. (2013). Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. *Molecular Biology and Evolution*, *30*(8), 1751–1760. http://doi.org/10.1093/molbev/mst092

Raymond, M., & Rousset, F. (1995). GENEPOP (VERSION-1.2) - POPULATION-GENETICS

SOFTWARE FOR EXACT TESTS AND ECUMENICISM. *Journal of Heredity*, *86*(3), 248–249. Journal Article. Retrieved from <Go to ISI>://WOS:A1995RB30200017

Reim, S., Proft, A., Heinz, S., Lochschmidt, F., Höfer, M., Tröber, U., & Wold, H. (2015). Pollen movement in a Malus sylvestris population and conclusions for conservation measures. *Plant Genetic Resources*, *FirstView*, 9. Journal Article. http://doi.org/10.1017/S1479262115000301

Rousset, F. (2008). GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106. Journal Article. http://doi.org/10.1111/j.1471-8286.2007.01931.x

Roux, C., Tsagkogeorga, G., Bierne, N., & Galtier, N. (2013). Crossing the species barrier: Genomic hotspots of introgression between two highly divergent ciona intestinalis species. *Molecular Biology and Evolution*, *30*(7), 1574–1587. http://doi.org/10.1093/molbev/mst066

Saleh, D., Milazzo, J., Adreit, H., Fournier, E., & Tharreau, D. (2014). South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, Magnaporthe oryzae. *New Phytologist*, *201*(4), 1440–1456. http://doi.org/10.1111/nph.12627

Schnitzler, A., Arnold, C., Cornille, A., Bachmann, O., & Schnitzler, C. (2014). Wild European apple (Malus sylvestris (L.) Mill.) Population Dynamics: Insight from Genetics and Ecology in the Rhine Valley. Priorities for a Future Conservation Programme. *Plos One*, *9*(5). Journal Article. http://doi.org/10.1371/journal.pone.0096596

Szpiech, Z. A., Jakobsson, M., & Rosenberg, N. A. (2008). ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, *24*(21), 2498–2504. Journal Article. http://doi.org/10.1093/bioinformatics/btn478

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. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 10954–9. http://doi.org/10.1073/pnas.1201403109

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., … Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*. Journal Article. http://doi.org/10.1111/eva.12367

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., … Viola, R. (2010). The genome of the domesticated apple (Malus × domestica Borkh.). *Nat Genet*, *42*(10), 833–839. http://doi.org/10.1038/ng.654

Zeder, M. A. (2006). Central questions in the domestication of plants and animals. *Evolutionary Anthropology*, *15*(3), 105–117. http://doi.org/10.1002/evan.20101

Zohary, D., & Hopf, M. (2000). *Domestication of Plants in the Old World* (3rd ed.). Oxford University Press.

III. Les silènes et leur pathogène *Microbotryum violaceum*

Introduction aux espèces étudiées : *Microbotryum – Silene*

Le second système hôtes pathogènes que j'ai étudié au cours de ma thèse est le système *Microbotryum – Silene*. *Microbotryum violaceum* est un complexe d'espèces jumelles de champignons pathogènes obligatoires, responsables de la maladie du charbon des anthères sur de nombreuses espèces de plantes appartenant aux Caryophyllaceae (Hood et al., 2010). Les principaux symptômes d'infection sont que les fleurs ne développent pas de structures reproductives femelles, mais seulement des structures mâles, même chez les plantes dioïques comme *S. latifolia* et que, dans les anthères, le pollen est remplacé par les spores noirâtres du champignon, d'où le nom de la maladie, le charbon des anthères.

Le cycle de vie de *M. violaceum* se déroule sur deux années (Fig. 2.5). Ce parasite est, de fait, trouvé exclusivement sur des plantes pérennes et non des annuelles (Hood et al., 2010). La première année, les téliospores germent sur une plante saine et produisent des sporidies, par méiose. Un évènement de conjugaison a lieu entre sporidies de types sexuels différents, fréquemment entre sporidies provenant d'une même tétrade, favorisant de forts taux d'auto-fécondation. Un hyphe dicaryotique se forme alors et pénètre dans les tissus de la plante hôte. Le parasite passe l'hiver dans la plante et, au printemps suivant, colonise les fleurs nouvellement formées et produit des téliospores qui se placent à la place du pollen. Celles-ci sont ensuite transmises de plantes en plantes par les pollinisateurs et/ou le vent et la pluie (López-Villavicencio et al., 2007).

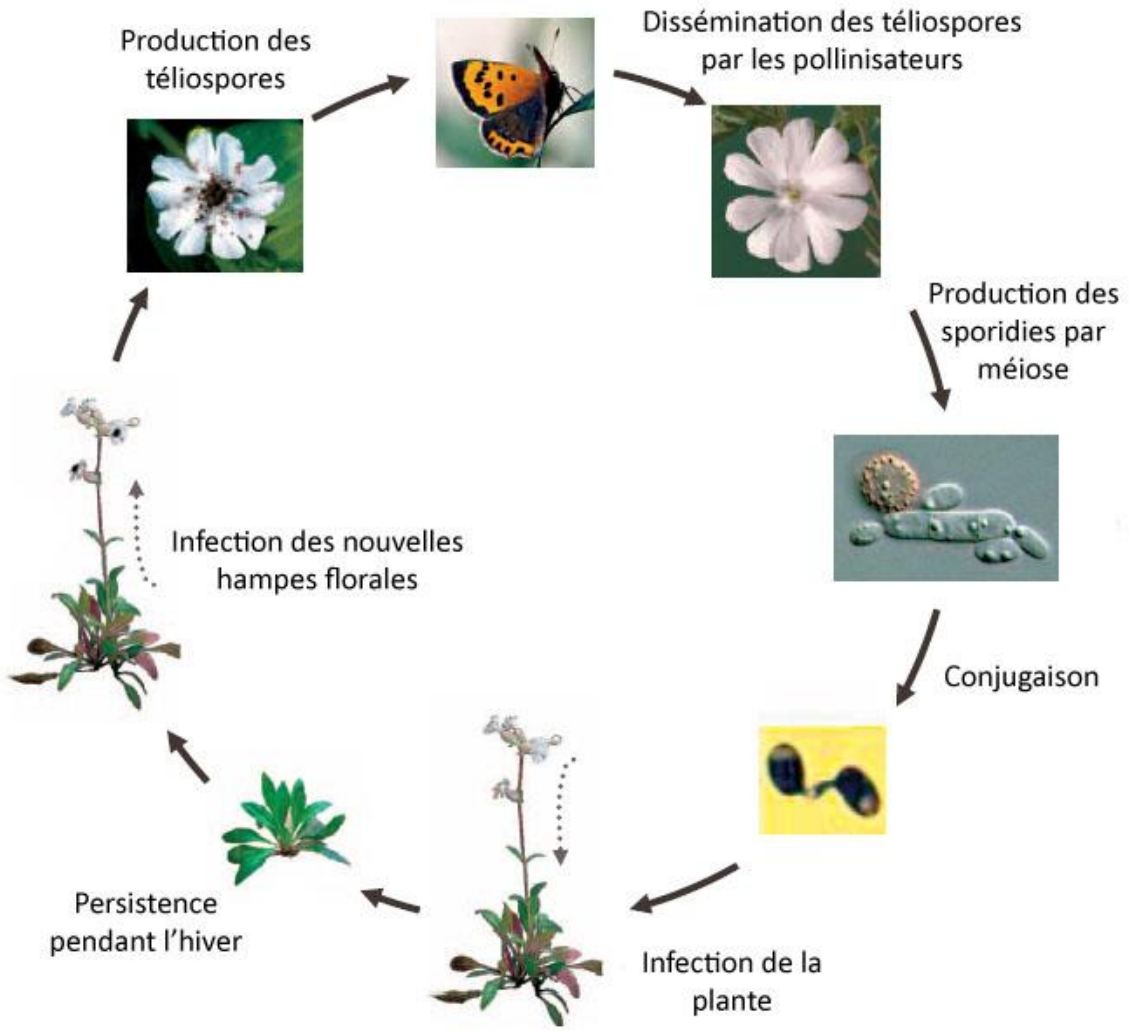


Figure 2.4 : Cycle de vie du pathogène Microbotryum violaceum*. Adapté de* López-Villavicencio et al 2007

Microbotryum violaceum est un modèle d'étude en écologie et génétique évolutive, par exemple en ce qui concerne la spéciation (Pierre Gladieux et al., 2011), la dynamique des épidémies (Brooks, Antonovics, & Keitt, 2008), les infections multiples (López-Villavicencio et al., 2007, 2011), et l'étude des chromosomes sexuels (Baker, 1948; Fontaine, Gladieux, Hood, & Giraud, 2013; Badouin et al. 2015). En ce qui concerne la coévolution, le modèle *M. violaceum* – Caryophyllacées a été étudié en particulier dans le domaine de la spéciation, et il a été montré que les souches présentes sur différentes plantes hôtes constituaient des espèces distinctes hautement spécialisées dans la plupart des cas (Gladieux et al., 2011). De plus, la comparaison des phylogénies des différentes espèces de Caryophyllacées et espèces de *Microbotryum* les infectant montrent que les sauts d'hôte ont été le mode de spéciation le plus fréquent (Fig. 2.6 ;

Refrégier et al 2008), comme pour une grande majorité des systèmes hôtes-parasites (De Vienne et al., 2013).

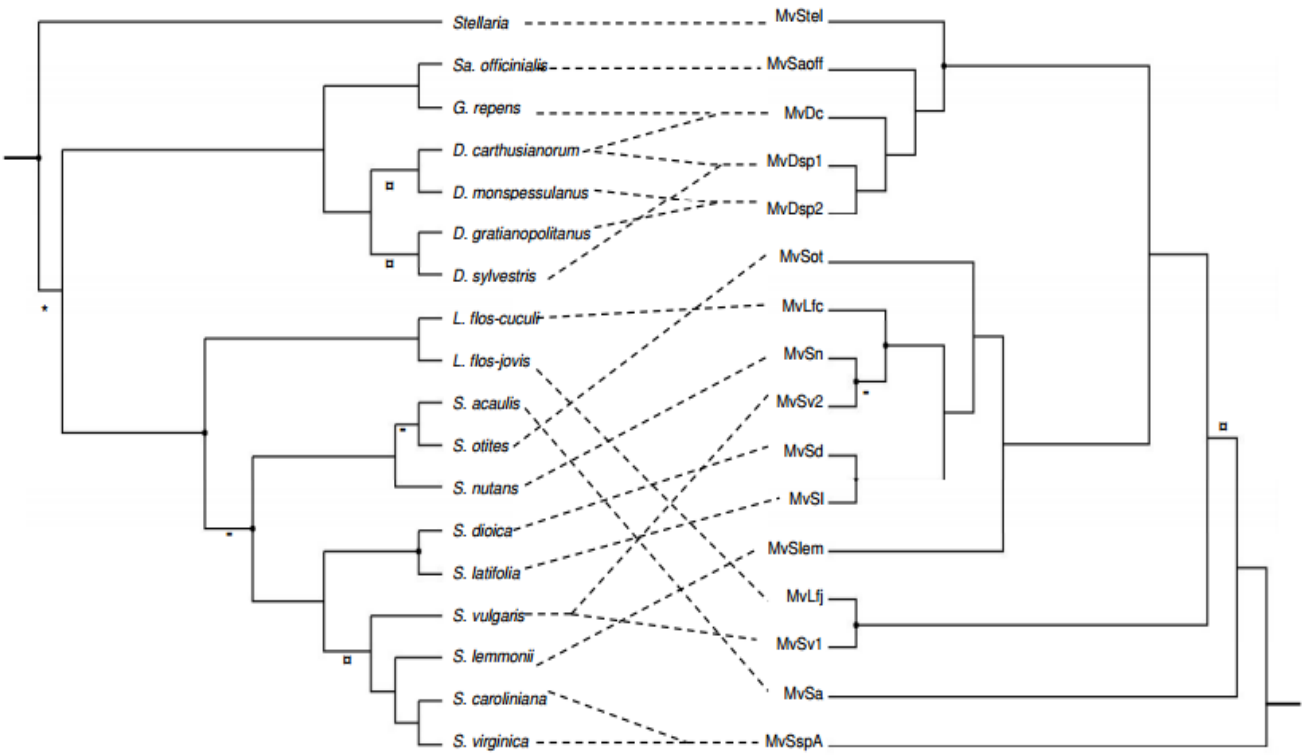


Figure 2.6: Comparaison des phylogénies des espèces de plantes et des espèces de Microbotryum. Figure tirée de Refrégier et al. 2008

Parmi les différentes espèces de Microbotryum, l'une des plus étudiées est Microbotryum lychnidis-dioicae. Ce pathogène est spécialiste du compagnon blanc, Silene latifolia, qui est une plante herbacée portant des fleurs blanches et trouvée le plus souvent dans des habitats modifiés par les activités anthropiques (bords de chemins, bords de champs ou de haies ou encore terrains vagues, Baker 1948, Fontaine et al 2013). C'est sur l'évolution de ce couple hôte-pathogène que je me suis concentrée au cours de ma thèse. J'ai tout d'abord participé à une étude concernant l'invasion des Etats-Unis par M. lychnidis dioicae à la suite de l'invasion de ces territoires par son hôte. J'ai également étudié la co-structure phylogéographique chez ce couple hôte-pathogène en Europe.

Invasion des Etats-Unis par Silene latifolia *et* Microbotryum lychnidis-dioicae

Silene latifolia a été introduite en Amérique du Nord au milieu du XIXème siècle et l'étude des herbiers de cette époque semble indiquer que la première introduction aurait été faite au nord-est des Etats-Unis en Nouvelle-Angleterre (Antonovics, Hood, Thrall, Abrams, & Duthie, 2003). Il a été montré qu'il y a eu une seconde introduction indépendante à l'est et que

les plantes introduites provenaient à la fois d'Europe de l'Ouest et de l'Est (Taylor & Keller, 2007). Des traces de charbons des anthères sur *S. latifolia* ont commencé à être observées beaucoup plus récemment, depuis la seconde moitié du XXème siècle (Fontaine et al., 2013). Une première étude a montré que les souches récoltées aux Etats-Unis sur *S. latifolia* ne sont pas issues d'un saut d'hôte depuis des espèces américaines, mais proviendraient de populations de *M. lychnidis-dioicae* situées en Grande-Bretagne, et plus précisément en Ecosse (Fontaine et al., 2013). L'échantillonnage clairsemé dans cette région n'avait pas permis de pointer vers un lieu plus précis. Une seconde étude avec un échantillonnage plus dense et à laquelle j'ai participé, en réalisant toutes les analyses de génétiques des populations sur *M. lychnidis-dioicae,* a permis d'obtenir une localisation potentielle plus précise (voir l'article en annexe 3).

Structure génétique spatiale chez Silene latifolia *et* Microbotryum lychnidis-dioicae

Au début de ma thèse, les structures génétiques à l'échelle de l'Europe de *S. latifolia* et de *M. lychnidis-dioicae* avaient déjà été étudiées séparément l'une de l'autre. Concernant *S. latifolia*, deux études portant principalement sur l'invasion des Etats-Unis ont également décrit sa structure génétique à l'échelle européenne. La première a utilisé des marqueurs chloroplastiques et a révélé l'existence de groupes phylogéographiques distincts à l'est et à l'ouest de l'Europe (Taylor & Keller, 2007). Cette structure phylogéographique a été interprétée comme résultant de la recolonisation de l'Europe à la fin de l'ère glaciaire à partir de deux refuges glaciaires probablement situés dans les péninsules ibériques et balkaniques. La seconde étude a été réalisée également avec des marqueurs microsatellites et de nouveau des marqueurs chloroplastiques. Elle souligne là encore une structure en deux groupes majeurs, avec les populations italiennes formant un groupe unique avec les populations de l'Ouest de l'Europe, et un second groupe en Europe de l'Est (Keller, Gilbert, Fields, & Taylor, 2012).

La structure génétique de *M. lychnidis-dioicae* avait également été étudiée avant le début de ma thèse. Notamment, Vercken et al (2010) avaient révélé l'existence de trois groupes génétiques majeurs, correspondant à l'Ouest, à l'Est et au Sud de l'Europe (c'est-à-dire à l'Italie pour le dernier). Une structure plus fine avait cependant été détectée également, subdivisant les groupes de l'Ouest et de l'Est de l'Europe. Cette structure a été interprétée comme résultant probablement de la recolonisation depuis les refuges glaciaires des péninsules ibérique, italienne et balkanique, mais aussi depuis un refuge plus éloigné en Eurasie et à partir de deux populations distinctes de la péninsule ibérique. Au cours de ma thèse, la structure de *M. lychnidis-dioicae* a

été étudiée également grâce à l'analyse de génomes complets visant à comprendre l'impact de la sélection, et possiblement de la coévolution, sur les génomes (Badouin, 2016). Là encore, les groupes phylogéographiques à l'Ouest de l'Europe, à l'Est de l'Europe et en Italie ont été identifiés, avec en plus un quatrième groupe constitué de trois souches échantillonnées en Espagne et en France. Les analyses de génomiques des populations ont révélé que 17% du génome de *M. lychnidis-dioicae* montrait des traces de balayages sélectifs récents, suggérant une coévolution récurrente avec son hôte (Badouin, 2016)

Article 5. Costructure chez *Silene latifolia* et *Microbotryum lychnidis-dioicae*

Comme indiqué ci-dessus, la co-divergence chez *M. violaceum* et ses hôtes a été étudiée à l'échelle inter-espèce et a révélé un grand nombre de sauts d'hôte plutôt que de la codivergence. Dans cette étude, nous avons décidé de nous intéresser plutôt à la co-divergence à l'échelle intra-espèce dans le cas du couple hôte-pathogène *S. latifolia* et *M. lychnidis-dioicae*, en lien avec la coévolution.

Les structures phylo-géographiques obtenues séparément dans les études précédentes pour ces deux espèces montraient certaines similitudes. Il semblait exister pour l'hôte, comme pour le pathogène, un groupe génétique couvrant l'Ouest de l'Europe et un autre couvrant l'Est de l'Europe. Cependant, dans le cas de la plante, les populations italiennes semblaient appartenir au groupe génétique de l'Ouest, alors que, pour le champignon, ces populations formaient un groupe génétique distinct.

Les données utilisées dans l'étude de Vercken et al. (2010) étaient constituées uniquement des génotypes des champignons *M. lychnidis-dioicae*, mais ceux-ci étaient issus des fleurs de *S. latifolia*. Ces dernières ayant été conservées, cet échantillonnage constitué des plantes et des pathogènes associés représentait une excellente occasion de comparer très précisément la structure génétique à l'échelle européenne au sein d'un système hôte-pathogène.

Ma contribution à cet article a constitué au génotypage des plantes, à la réalisation des analyses, à la production des figures et à la rédaction du matériel et méthodes, des résultats et d'une partie de la discussion. J'ai également réalisé des analyses d'ABC sur l'hôte et le pathogène afin de déterminer si le scénario évolutif de divergence entre les groupes génétiques détectés dans l'étude était le même pour les deux organismes. En raison du faible nombre de marqueurs microsatellites utilisés, les résultats n'ont pas été concluants puisque les statistiques résumées ne permettaient pas de différencier les simulations des différents scénarios. Ces analyses n'apparaissent donc pas dans l'article suivant. Cet article va être publié dans le « New Phytologist » (au moment de l'écriture de ce manuscrit, l'article est disponible en « early view ») et sera associé à un commentaire par Anna-Liisa Laine et Daniel Croll.

Strong phylogeographic co-structure between the anther-smut fungus and its white campion host

Feurtey Alice, Gladieux Pierre, Hood Michael E., Snirc Alodie, Cornille Amandine, Rosenthal Lisa and Giraud Tatiana

Introduction

Understanding and managing fungal diseases of plants is aided by appreciating the processes that generate and maintain spatial patterns in pathogen diversity (Williams, 2010). Predicting the impacts of coevolving systems, and more generally our ability to manage biological interactions embedded in a complex network of agricultural and natural landscapes, requires knowledge about how spatial heterogeneity, selection and environmental variables contribute to the differentiation of populations (Barrett *et al.*, 2008). The degree of genetic variation in specialized pathogens is particularly influenced by their host's distribution, abundance and genetic make-up (Wilson *et al.*, 2005). This leads to the expectation that the population structure of pathogens would reflect those of their hosts (Criscione *et al.*, 2005). Under a very similar rationale, host and pathogen phylogenies have long been expected to be congruent (Hafner & Page, 1995), whereas it has come as a recent surprise that host shifts are actually the main diversification processes in the vast majority of host–pathogen associations (de Vienne *et al.*, 2013). Unlike the between-species level, however, few studies have addressed the degrees of congruence between host and pathogen genetic structures within-species, although such studies may also reveal unexpected findings, as did the between-species phylogenetic congruence studies (de Vienne *et al.*, 2013). In addition, within-species coevolution and diversification on different hosts corresponds to different evolutionary mechanisms that can, thus, lead to different patterns, that is, with possibly phylogenetic incongruence at the between-species level (de Vienne *et al.*, 2013) but congruence of population genetic structures at the within-species level.

The few published studies on host–pathogen co-structure have been conducted mostly on parasites of animals (Dybdahl & Lively, 1996; McCoy *et al.*, 2005; Miura *et al.*, 2006; Nieberding *et al.*, 2008; Rich *et al.*, 2008; Amico & Nickrent, 2009; Nakao *et al.*, 2009), with a few, notable exceptions in plant parasites (Michalakis *et al.*, 1993; Whiteman & Parker, 2005; Wirth *et al.*, 2005; Tsai & Manos, 2010; Magalhaes *et al.*, 2011). The genetic structure of the pathogens has, in these cases, often been found to be congruent with those of their host. Due to

shorter generation times and the tendency toward selfing mating systems or partial clonality, pathogen populations have nevertheless often been more subdivided than those of their hosts (Nieberding & Olivieri, 2007; Barrett *et al.*, 2008). This may help in reconstructing the host population history (Nieberding & Olivieri, 2007; Barrett *et al.*, 2008), provided that pathogens have followed the same history as their hosts. Indeed, if pathogens show a stronger genetic structure than their hosts, analyses of pathogen population subdivision will be more informative about the joint past demographic histories of the two partners than studies conducted using host data alone. In fungal plant pathogens, however, studies of population structure have focused on agents of crop diseases, where the host population is genetically homogeneous and distributed artificially, and the pathogens are dispersed widely, in part by human transport, and are often recently introduced exotic species (Enjalbert *et al.*, 2005; Barres *et al.*, 2008; Fournier & Giraud, 2008; Gladieux *et al.*, 2008; Saleh *et al.*, 2014).

Here, we report a study on the co-phylogeography of the castrating fungal pathogen *Microbotryum lychnidis-dioicae* (previously named *M. violaceum* or *Ustilago violacea*) and its host plant *Silene latifolia*. Both are model organisms for a variety of topics in ecology and evolution, and, hence, numerous studies are available on their life history, population biology and natural history (Bernasconi *et al.*, 2009; Gladieux *et al.*, 2015). This system is ideal for studying the co-structure between pathogens and their host plants, given that the fungus is an obligate pathogen which is not capable of persisting on its own in the environment (Hood *et al.*, 2010), is highly host-specific in nature (Vercken *et al.*, 2010; Gladieux *et al.*, 2011), the same vectors disperse the fungus and the host pollen (i.e. the pollinators) (Thrall *et al.*, 1995), *Silene* and *Microbotryum* species have identical generation times (i.e. one per year) (Thrall & Jarosz, 1994), and there has been no attempt to modify plant distribution or genotypes or to control the disease because its effects are without economic consequences (Bernasconi *et al.*, 2009). A stronger phylogeographic structure is expected in the fungal pathogen than its host plant because *M. lychnidis-dioicae* is highly selfing (Hood & Antonovics, 2000; Giraud *et al.*, 2005, 2008), whereas *S. latifolia* is dioecious and shows inbreeding depression (Teixeira *et al.*, 2009), because the fungus only disperses via pollinators, whereas gene flow in the plant also occurs via its seeds in addition to pollen, and because pollinators discriminate against diseased plants (Shykoff & Bucheli, 1995), likely reducing spore transport compared to pollen flow. Note that *M. lychnidis-dioicae*, unlike most fungi, cannot proliferate by asexual reproduction beyond a few cells after meiosis, and a sex event (i.e. meiosis followed by syngamy and dikaryotic growth) is required for infecting a new plant. However, the mating system is highly automictic – with mostly diploid selfing by mating among products of the same meiosis (Oudemans *et al.*, 1998;

Hood & Antonovics, 2000; Thomas *et al.*, 2003; Badouin *et al.*, 2015). A stronger population genetic structure has been found in *M. lychnidis-dioicae* than in *S. latifolia,* although co-structure has been investigated only at a very small spatial scale so far (Delmotte *et al.*, 1999). Local adaptation in this system has likewise been investigated at a very small scale to date, with local plant genotypes being more resistant to their local anther-smut pathogen than to pathogens from other nearby populations (Kaltz *et al.*, 1999). Differences were found in terms of the percentage of plants that became diseased after experimental inoculations, whereas no gene-for-gene interactions have been detected in this system.

Previous studies on the population genetics of *M. lychnidis-dioicae* provided one of the most clear-cut examples of phylogeographic structure in pathogens (Vercken *et al.*, 2010; Gladieux *et al.*, 2011, 2013; Fontaine *et al.*, 2013). Clustering analyses revealed the existence of three main genetically distinct groups within *M. lychnidis-dioicae*, with distributions strikingly similar to the major intraspecific lineages identified in Europe for many temperate plant and animal species (Hewitt, 1999). This within-species structure is believed to reflect recolonization from southern refugia after glaciation. The Eastern and Western clusters were each further split into two groups, consistent with colonization from distinct refugia located in the Balkans and further East in Eurasia, and survival of the pathogen in distinct regional refugia in Western Europe.

Phylogeographic analyses of chloroplast genetic data in the host plant, *S. latifolia*, also revealed patterns of population structure consistent with postglacial expansion from distinct Mediterranean refugia, although these were less clear-cut than for the anther-smut fungus, with large geographical overlaps of maternal lineages (Taylor & Keller, 2007). However, thorough comparisons between host and pathogen structures were prohibited by the facts that the plant and pathogen samples were not collected at the same localities, and that a selective sweep in *S. latifolia* chloroplast DNA may have obscured historical patterns (Muir & Filatov, 2007). Other studies using microsatellite markers in the host plant *S. latifolia* were done, but at smaller geographical scales (Delmotte *et al.*, 1999; Barluenga *et al.*, 2011; Magalhaes *et al.*, 2011) or with less dense sampling (Keller *et al.*, 2012). Nevertheless, a clear differentiation was found between Eastern and Western European clusters (Keller *et al.*, 2012).

Thanks to a collection of associated plant/pathogen samples (fungal spores and plant material from the same disease plants) with an unprecedented sampling of the plant in terms of both density and geographical scale in Europe, in the present study we investigated the co-structure of *M. lychnidis-dioicae* and *S. latifolia* using nuclear microsatellite markers for both the plant and the fungus, and we used inoculation experiments to test for the existence of local

adaptation at the cluster level. More specifically, we addressed the following questions: Can we detect footprints of three glacial refugia in the plant population structure as occur in its fungal pathogen? How congruent are the population structures of the host and the pathogen at the European scale? Are pairwise genetic distances between pathogen individuals and host individuals significantly correlated? In this case, does the correlation still hold when removing the effect of isolation-by-distance: are the host and pathogen population genetic structures more congruent than expected if they were just both affected by the same drift–migration equilibrium after similar postglacial recolonization histories? Is there experimental evidence of local adaptation in the host–pathogen association that may contribute to congruent spatial structure at the cluster level?

Materials and methods

Sampling, genotyping and species identification

We used the same diseased *Silene latifolia* plants from which the anther-smut samples had been genotyped previously (Vercken *et al.*, 2010). Because *Microbotryum* spores were collected in flower buds of *S. latifolia*, samples included both materials from the fungus and from the parasitized plants. DNA was extracted using the Nucleospin Plant II kit following manufacturer instructions (Macherey Nagel R). We genotyped the plants using 16 microsatellite markers distributed across the genome (Moccia *et al.*, 2009). Microsatellite markers were amplified by multiplex PCR, with the Multiplex PCR Kit (Qiagen) following manufacturer instructions (for details of the multiplex PCR programs used, see Supporting Information Tables S1 and S2). Genotyping was performed on an ABI Prism X3730XL at the Gentyane Genotyping Platform (INRA, Clermont, France). Alleles were scored with Genemapper v.4.0 (Applied Biosystems).

For the anther-smut samples, we used a previously published dataset of genotypes produced with a set of 11 microsatellite markers (Vercken *et al.*, 2010). The *Microbotryum* dataset was rendered haploid by randomly removing one allele in the rare cases of heterozygosity (Vercken *et al.*, 2010). Indeed, individuals are almost completely homozygous because of high selfing rates (Hood & Antonovics, 2000; Giraud, 2004; Giraud *et al.*, 2005, 2008) and this may bias analyses of population subdivision using Bayesian clustering methods assuming random mating. Apparent heterozygosity may, in principle, result from co-infection of the plant by multiple genotypes (Tollenaere *et al.*, 2012), but previous studies have shown that

a given *S. latifolia* bud typically hosts a single *M. lychnidis-dioica* genotype, even in cases of multiple infections at the plant scale (López-Villavicencio *et al.*, 2007). For all the analyses, we kept only the individuals (both for the fungus and the plant) successfully genotyped at eight loci or more.

Our plant dataset contained individuals that were identified by the samplers as *S. latifolia* (n = 592 individuals), *S. dioica* (n = 74) and putative hybrids between these species (n = 50). The anther-smut fungi dataset initially contained samples that were identified by the samplers as collected on *S. latifolia* (n = 669), on *S. dioica* (n = 371) and on hybrids between the two plant species (n = 30). We first ran analyses to ensure that our samples all belonged to the species studied, *S. latifolia* for the plant and *M. lychnidis-dioicae* for the fungus. Indeed, the sister plant species *S. dioica* and *S. latifolia* have largely overlapping distributions and *S. dioica* is also affected by anther-smut disease caused by *M. silenes-dioicae*, the sister species of *M. lychnidis-dioicae*. Both of the plant and the anther-smut fungi pairs hybridize in nature, and plant and fungal hybrids are difficult to detect morphologically (Karrenberg & Favre, 2008; Vercken *et al.*, 2010; Gladieux *et al.*, 2011). We therefore used Bayesian multilocus genotype assignments to detect potential hybrids in plants and anther-smut fungi (Falush *et al.*, 2003). We identified 578 and 679 genotypes as 'pure' representatives of *S. latifolia* and *M. lychnidis-dioicae*, respectively (Notes S1; Fig. S1). Only these 'pure' individuals were retained for subsequent analyses, from which we could obtain both the genotype of the plant and of its associated anther-smut pathogen for 391 samples.

Population structure

We investigated the population structure within the *M. lychnidis-dioicae* and *S. latifolia* species. First, we used the model-based Bayesian clustering methods implemented in Tess v.2.3 (Guillot, 2009). Analyses were performed anew on the genetic marker dataset created for the fungus in our previous study (Vercken *et al.*, 2010) in order to have the same software tools applied on the plant and fungal datasets. We kept only one representative of each genotype in the dataset for clustering analyses and assigned the repeated genotypes to the same cluster as the corresponding representative genotypes. We used the conditional auto-regressive (CAR) Gaussian model of admixture with 200 000 MCMC iterations after a burn-in of 100 000 iterations. We used the CAR Gaussian model of admixture with linear trend surface, setting the spatial interaction parameter (ρ) to 0.6. These parameters (ρ and trend) affect the weight given to spatial distance when clustering genotypes. Tess outputs were processed with Clumpp v.1.1.2 (Jakobsson & Rosenberg, 2007), to identify distinct clustering solutions in the replicated runs for

each *K* value. We considered that an individual belonged to a given genetic cluster if its membership coefficient for this cluster was superior to 0.8. As described above, repeated genotypes were not included in the analyses that established cluster assignment per population, but they were added back for subsequent analyses.

We also performed population structure analyses not relying on any population genetic model and therefore not affected by deviations from Hardy–Weinberg assumptions, such as those caused by selfing. Principal component analyses (PCA) were performed using the ade4 R package and discriminant analyses of principal component (DAPC) using the adegenet R package with the R software (find.clusters function).

For both the host and the parasite, we tested the existence of isolation-by-distance (IBD) patterns by assessing the significance of the correlations between matrices of genetic and geographic distances, using a Mantel test (100 000 resamples for the null distribution) as implemented in the ncf R package (Paradis *et al.*, 2004). Genetic distances between individuals were calculated using Nei's Da estimator as implemented in Populations v.1.2.32 (Langella, 2002). Geographic distances were computed with Tess (with the option Compute Great Circle Distances). A Mantel test also was performed to evaluate the significance of the correlation between the genetic distances between pathogen individual pairs and plant individual pairs.

In order to investigate the congruence of genetic structure between the host and the pathogen, we performed partial Mantel tests (100 000 resamples for the null distribution) using the *ncf* R package to assess the significance of the correlation of pairwise genetic distances between pathogen individuals and plant individuals, as above, but removing the effect of the correlation between genetic and geographic distances.

The genetic variation for both plants and pathogens was quantified using the allelic richness, *Ar*, calculated with Adze v.1.0 (Szpiech *et al.*, 2008). Expected heterozygosity (H_E), observed heterozygosity (H_O) and F_{IS} were computed using Genepop v.4.3 (Rousset, 2008). The statistical analyses on allelic richness and H_E (correlations and Wilcoxon signed rank tests) were performed with Jmp v.9 (SAS Institute Inc., Cary, NC, USA) using the value of the statistics per marker.

Experimental inoculations

In order to generate host material for the inoculation experiment, seeds of *S. latifolia* were produced under glasshouse conditions from plants originating from field-collected seeds. Each location was represented by a pooled set of seeds from at least three crosses between male and female parental plants. Because sufficient parental material was not available from all

regions (e.g. the Southern cluster), we could not perform a full design inoculation experiment. For each available seed–fungus population pair, inoculations were performed using a local pathogen strain and a strain chosen at random among available strains from a different cluster (Table S3). Inoculum suspensions were prepared from field-collected teliospores of *M. lychnidis-dioicae*, suspended in water at a concentration of 500 spores ml^{-1}. Teliospore germination was confirmed to be > 90% by plating on potato dextrose agar and examination after 24 h incubation.

Inoculation procedures followed prior studies (Hood & Antonovics, 2000). Briefly, seeds were germinated in Petri dishes on 0.8% agar medium containing Murashige and Skoog (MS) basal salt. After 10 d of growth, the apical meristems were exposed by expansion of the cotyledons and 4 µl of inoculum suspension was applied. Inoculated plants were maintained at 15°C for 3 d, and then transplanted to soil and maintained under glasshouse conditions until flowering. Infection status was assessed by the presence of *M. lychnidis-dioicae* spores in the flowers.

For each source of seeds, experimental treatments included inoculation with *M. lychnidis-dioicae* collected from the same location as the source of seeds or with *M. lychnidis-dioicae* from one of the other identified clusters (Table S3). Two hundred seeds per location were used, with one hundred seedlings each assigned randomly to one of the two treatments (local vs intercluster inoculum type). Positions of the plants in the Petri dishes and the glasshouse were randomized. A logistic regression was used to test for differences in disease rates between treatments, performed with Jmp v.9 (SAS Institute).

Results

Population structure and IBD in the host plant Silene latifolia

The population structure of the host plant *S. latifolia* (n = 578) was analysed using the Tess Bayesian clustering method. The Deviance Information Criterion (DIC), a statistical measure of the model prediction abilities for Tess, sharply dropped at K = 3 and decreased much more slowly with further increase in K (Fig. S2). This indicates that the three-cluster structure is the strongest genetic structure. In fact, the barplots showed only three clusters even at higher K values (Fig. S3). Also consistent with this finding, PCA and DAPC clearly separated the three same genetic clusters (Figs 1a, S4), and no further clear and geographically relevant cluster could be identified at higher K using DAPC or looking at further axes in the PCA. A first cluster was found in France, the UK, Belgium and Germany (hereafter called the 'Western' cluster; n = 319),

the second cluster in Italy (hereafter called the 'Southern' cluster; $n = 91$), and the third one in Hungary, Ukraine, Russia and the Czech Republic (hereafter called the 'Eastern' cluster, $n = 132$) (Fig. 1 b & c). We found 36 admixed individuals that could not be assigned to any cluster – genotypes whose highest membership coefficient was < 0.8. Half of these admixed individuals came from locations in contact areas among clusters.

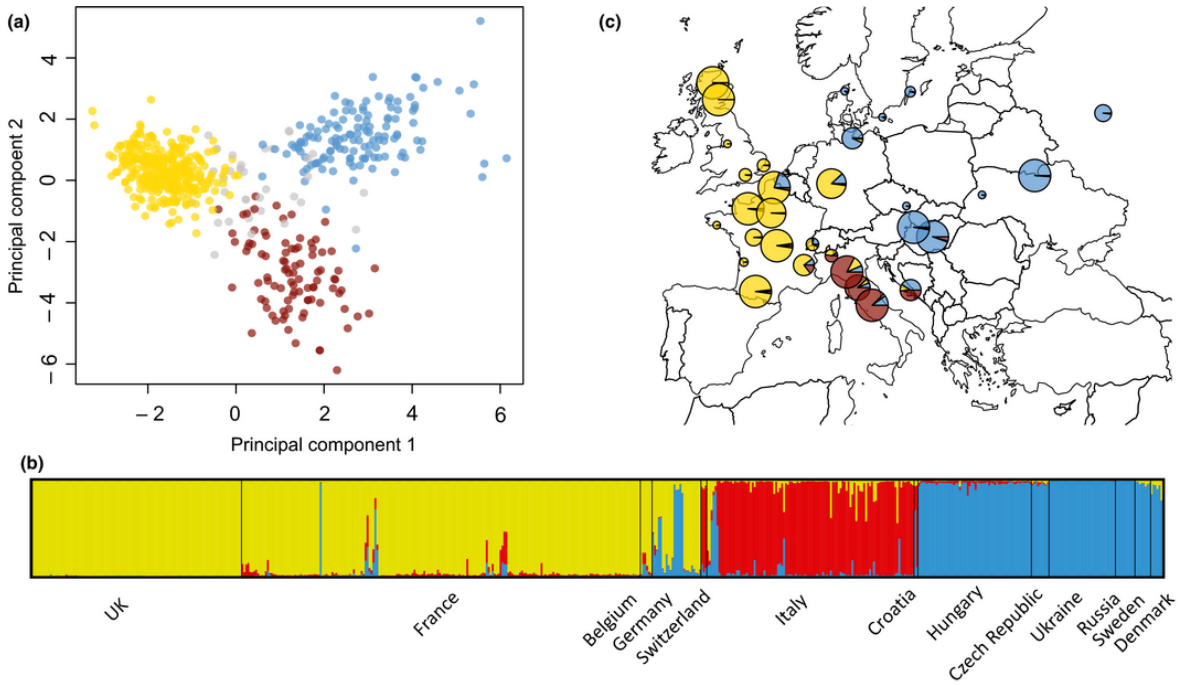


Figure 1: Genetic structure of the host plant, Silene latifolia, representing the three clusters obtained at K = 4 and in the majority of the K = 3 solutions. (a) Principal component analysis on the microsatellite alleles. The dot colour corresponds to the three clusters assigned to each individual by the Bayesian computations. The grey dots are the individuals not assigned to any cluster when using a threshold of 0.8. (b) Bar plots of membership probabilities per individual (as vertical bars) from the Tess analysis. (c) Map of mean cluster membership probabilities per locality inferred with Tess. The pie diameter reflects the number of individuals sampled in the corresponding locality: the smallest circles correspond to $n \leq 3$ *and the largest circles to* $n \geq 16$.

The correlation between the matrices of genetic and geographic distances between pairs of *S. latifolia* individuals was significant, at the European scale ($r = 0.3351$, $P < 0.0001$), as well as within each cluster (Mantel tests; Eastern: $r = 0.1486$, $P < 0.0001$; Southern: $r = 0.2091$, $P < 0.0001$; Western $r = 0.1151$, $P < 0.0001$), indicating IBD patterns at both spatial scales.

Allelic richness (Ar), expected heterozygosity (H_E), observed heterozygosity (H_O) and inbreeding index F_{IS} were computed for the tree genetic clusters (Table S4). The Eastern cluster appeared the most genetically diverse but no significant differences were detected between clusters. F_{IS} values were relatively high for a dioecious plant ($F_{IS} > 0.16$).

Population structure and IBD in the fungal pathogen Microbotryum lychnidis-dioicae

Genetic structure patterns for *M. lychnidis-dioicae* (n = 679) was also investigated using Tess, and gave results very close to the clustering results found in the host plant. Three clusters were differentiated at K = 3: one in Western (n = 308), one in Southern (mainly Italy, n = 160) and one in Eastern Europe (n = 178) (Fig. 2a,b). Only 33 individuals could not be assigned to any cluster, half of which came from contact areas between two clusters. The Deviance Information Criterion (DIC) decreased steadily with increasing number of K, with no sharp drop (Fig. S2b). The PCA, however, showed that the structure at K = 3 was the strongest, with three distinct clusters corresponding to those obtained with the Tess software and a separation of the Western cluster by the first axis, then a subsequent split into Southern and Eastern clusters along the second axis (Fig. 2c).

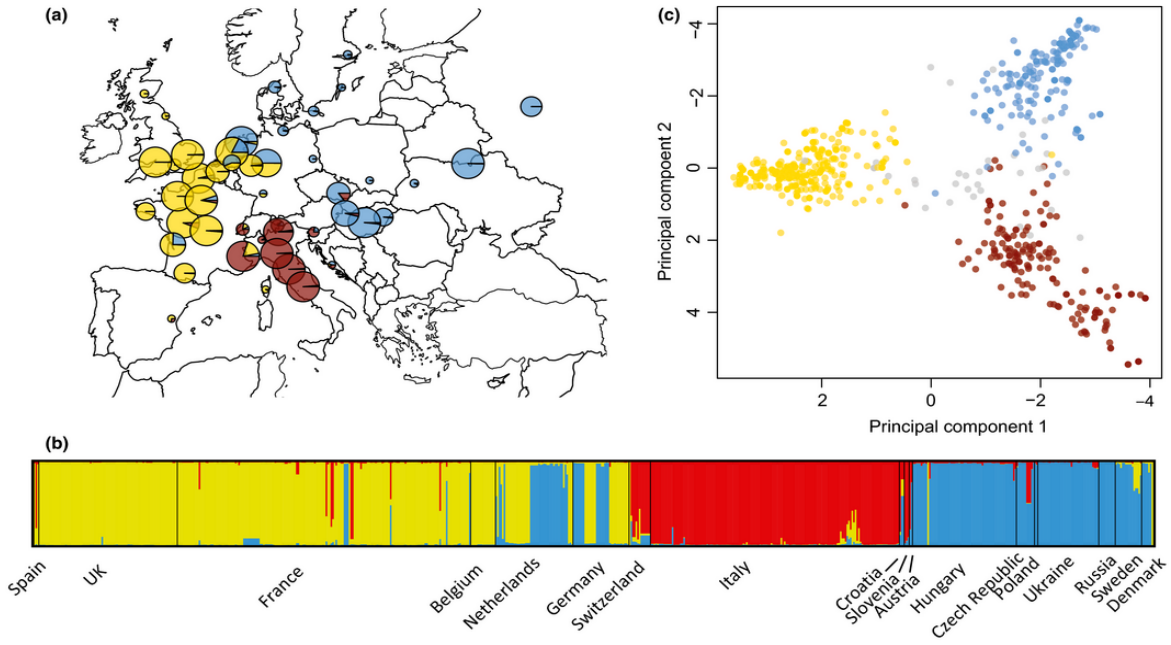


Figure 2 : Genetic structure of the anther-smut fungus, Microbotryum lychnidis-dioicae, at K = 3. (a) Map of mean cluster membership proportions per locality inferred with Tess at K = 3. The pie diameter reflects the number of individuals sampled in the corresponding locality: the smallest circles correspond to n ≤ *3 and the largest circles to* n ≥ *16. (b) Bar plots of membership proportions per individual (as vertical bars) from the Tess analysis for* K = *3. (c). Principal component analysis on the microsatellite alleles. The dot colour corresponds to the cluster assigned for each individual by the Bayesian computations. The grey dots are the individuals not assigned to any cluster when using a threshold of 0.8.*

The correlations between the matrices of genetic and geographic distances between pairs of *M. lychnidis-dioicae* were significant at the European scale as well as within clusters (Mantel tests; European scale, r = 0. 3937, P = 0.0002; Eastern cluster, r = 0. 3459, P = 0.0002; Southern

cluster, $r = 0.3905$, $P = 0.0002$; Western cluster, $r = 0.2347$, $P = 0.0002$). This indicated IBD patterns in *M. lychnidis-dioicae*, both within clusters and at the European scale.

The allelic richness and the expected heterozygosity in *M. lychnidis-dioicae* were much lower than in the plant (Table S5), and no significant differences were detected between clusters. F_{IS} values were very high, as expected for this highly selfing fungus.

In agreement with the steady decline of the DIC value from Tess, we found a finer population subdivision, with well-defined clusters at $K>3$. Seven geographically circumscribed clusters were thus identified by both model-based and nonmodel-based methods (Figs S5, S6).

Co-structure between the host Silene latifolia *and pathogen* Microbotryum lychnidis-dioicae

At $K = 3$, the host and pathogen population structures appeared highly similar (Fig. 3 a & b). In particular, the limits of the *M. lychnidis-dioicae* clusters appeared extremely congruent with those found in *S. latifolia*. Of the 80 sampling localities where at least two joint samples of both pathogen and host were available, only five were assigned to different clusters between the host and the pathogen.

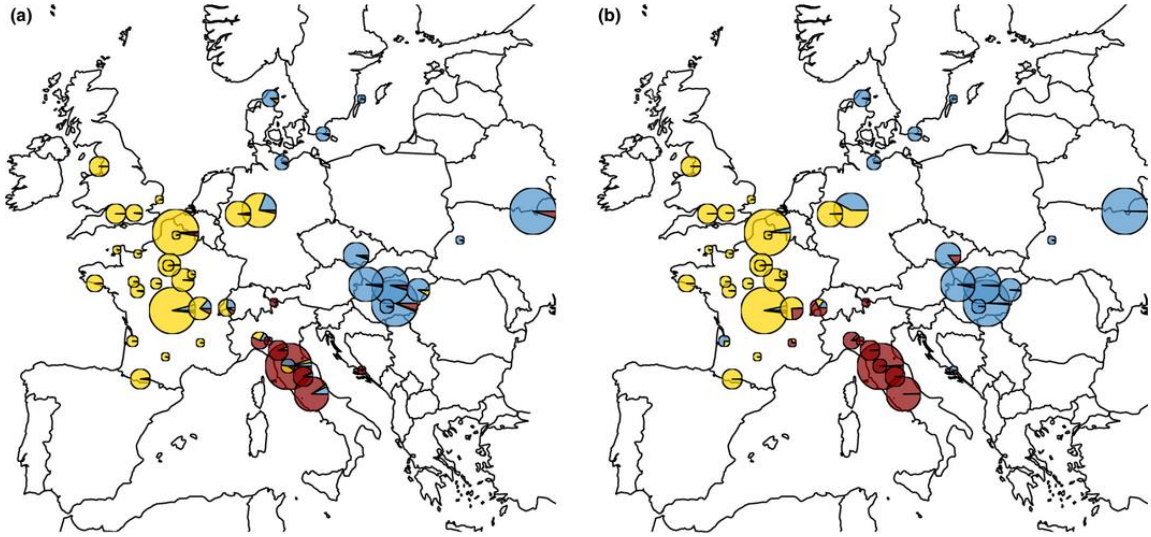


Figure 3: Genetic structure of the host plant, Silene latifolia*, and its pathogen,* Microbotryum lychnidis-dioicae*, from paired samples. (a) Map of mean membership proportions per localities inferred with Tess for the three-cluster structure for* S. latifolia. *(b) Map of mean membership probabilities per locality inferred with Tess for the three-cluster structure for* M. lychnidis-dioicae. *The pie diameter reflects the number of individuals sampled in the corresponding locality and the colours denote different clusters. Only the individuals for which the host–pathogen pair were available are shown here.*

Microsatellite data were available for 391 associated pairs of both hosts and pathogens (i.e. the genotypes from the fungus and its host plant individual). Directly associated host and pathogen samples, collected on the same plant individual, predominantly were assigned to the

same clusters for the host and the pathogen (i.e. Western, Eastern or Southern cluster). In only 8 (2%) host and pathogen pairs, the plant and fungal individuals were assigned to different clusters. In three host and pathogen pairs, the plant and fungal individuals were both admixed (i.e. 8–9% of the admixed plant or fungal individuals), all of these being located at the suture zones between the Western and Southern clusters.

The correlations between genetic distance matrices for the host and the pathogen were significant at the European scale as well as within clusters (Mantel tests; European scale, $r = 0.3505$, $P < 0.0001$, Fig. 4; Eastern, $r = 0.1389$, $P < 0.0001$; Southern, $r = 0.1518$, $P = 0.0003$; Western, $r = 0.0837$, $P = 0.0081$).

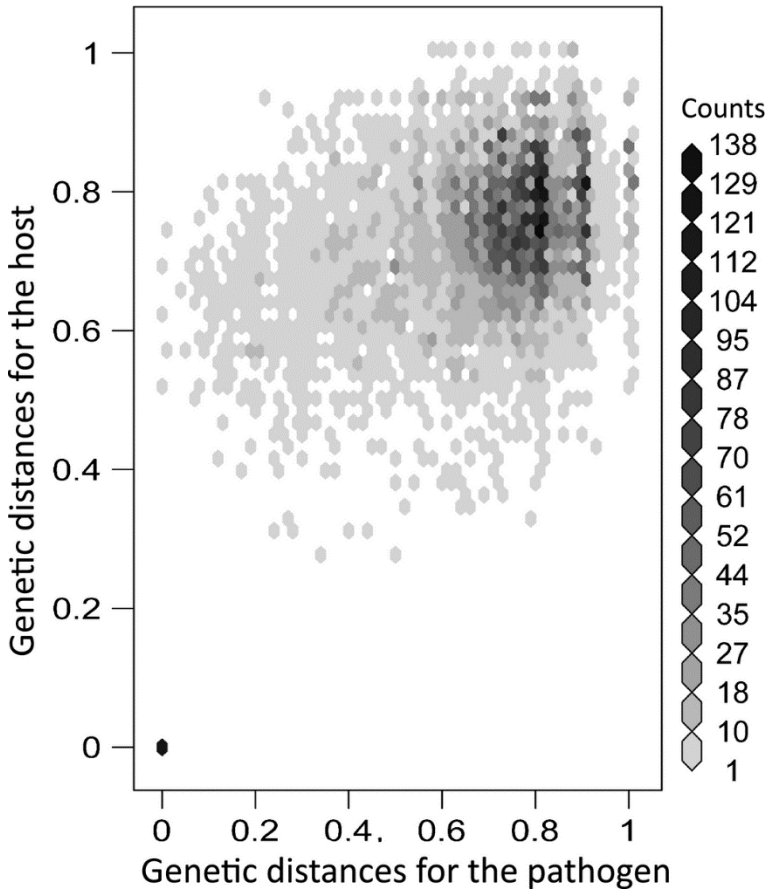


Figure 4: Genetic distances (Nei's Da) between plant (Silene latifolia) *pairs plotted against genetic distances between the corresponding pathogen* (Microbotryum lychnidis-dioicae) *pairs. For clarity we used a hexagon binning representation and the distances per sampling location; the darker the hexagon, the more dots are found in the corresponding area of the plot.*

A partial Mantel test was then performed for testing the congruence between the genetic structures of the host and the pathogen while controlling for the effect of isolation by geographic distance, using the two matrices of genetic distances (host and pathogen) and a matrix of

geographic distance between pairs of samples. The partial Mantel test at the European scale was again significant ($r = 0.2277$; $P < 0.0001$), indicating that the genetically closer two *S. latifolia* individuals were, the genetically closer their anther-smut pathogens were, independently of the shared pattern of IBD. With a threshold of 0.8 for assignation of individuals to a genetic cluster, the partial Mantel test was significant for the Eastern cluster ($r = 0.0840$, $P < 0.0001$) and close to significance for the two other clusters (Western: $r = 0.0450$; $p = 0.0903$; Southern: $r = 0.0710$, $P = 0.0538$). When lowering the threshold of assignation to a cluster to 0.5 (thus increasing the number of sample pairs and adding admixed individuals of plants and pathogens), all the partial Mantel tests were significant (Europe: $r = 0.2277$; $P < 0.0001$; Western: $r = 0.0665$; $P = 0.0229$; Eastern: $r = 0.0889$, $P < 0.0001$; Southern: $r = 0.0987$, $P = 0.0089$).

Inoculation experiments

In order to test the existence of local adaptation in the plant and/or the pathogen at the cluster level, seedlings from one location in the western cluster and three locations in the eastern cluster were inoculated with spores coming either from their local population (hereafter, 'local treatment') or from a population belonging to a different cluster, randomly assigned (hereafter, 'inter-cluster treatment'). The probability of a plant being diseased varied significantly with the treatment (i.e. local vs intercluster; Fig. 5); although the seedling origin was not significant as a main effect, the interaction between the seedling origin and the treatment was significant (Table S6). The disease rates were, however, never lower in inoculations by pathogens from a different cluster than by local pathogens (Fig. 5). Altogether, this means that a given plant better resists its local anther-smut fungi than one from other clusters, but not significantly so in all populations. The results thus indicate that local adaptation by the host plants often occurs.

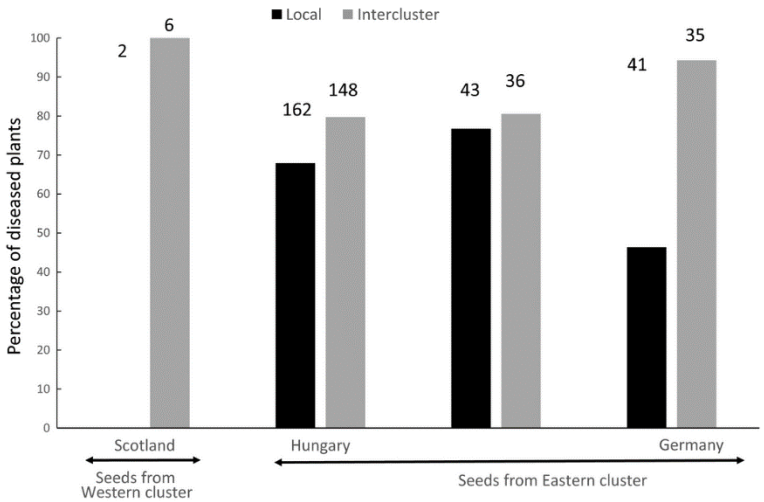


Figure 5: Disease rates obtained in inoculation experiments of Silene latifolia *by* Microbotryum lychnidis-dioicae. *The percentage of diseased plants is represented for each treatment, the total number of plants having flowered being indicated at the top of the bars. Seedlings from one location in the Western cluster (left) and three localities within the Eastern cluster (right) have been inoculated with spores coming either from their local populations (black bars) or from a population belonging to a different cluster grey bars); see Supporting Information Table S3 for a description of the inoculation design.*

Discussion

Footprints of glacial refugia in the population structure of the host Silene latifolia

A sampling of an unprecedented density and spatial scale for *Silene latifolia* enabled the investigation of genetic structure across Europe for this important model in ecology and evolution (Bernasconi *et al.*, 2009). The strongest pattern of microsatellite variation corresponded to three genetic clusters: a Western cluster, a Southern cluster in Italy and an Eastern cluster. The pattern obtained here using microsatellite markers appears much more clear-cut and revealed further geographical structure (i.e. the Italian clusters) than the one obtained previously with chloroplast markers (Taylor & Keller, 2007), which showed large geographical overlaps between lineages. This discrepancy reinforces the idea of a selective sweep having occurred throughout several clusters in the chloroplast DNA (Ironside & Filatov, 2005), whereas the nuclear gene pools from different southern glacial refugia remained differentiated after recolonization of northern Europe. The population structure of *S. latifolia* has been investigated previously using microsatellite markers across Europe, but only the East/West structure was highlighted (Keller *et al.*, 2012); nevertheless that previous study suggested that a finer genetic subdivision was present. In fact, re-analysing this independent dataset, that includes different markers and samples from ours, also reveals the Italian cluster at $K > 2$ (not shown).

The biogeographical pattern we found in *S. latifolia* is typical of the genetic structure originating from contracted ranges in the south of Europe during the last glacial maximum followed by recolonization northward after climate warming at the beginning of the Holocene. The structure thus indicates that, during the last glaciation, *S. latifolia* likely survived in three main glacial refugia (the Iberian, the Italian and the Balkan refugia), and then recolonized Northern Europe without much large-scale eastern or western dispersal that would have admixed the different clusters. A similar pattern has been found in numerous animal and plant species, these three regions being the main European refugia where climatic conditions stayed mild enough for temperate species during the glacial phases (Hewitt, 1999, 2004).

The identification of a new differentiated gene pool in *S. latifolia* will be useful for future studies on the different topics for which this plant is a model, from the study of sex chromosomes and mating systems, to speciation and biotic interactions (Marais *et al.*, 2008; Bernasconi *et al.*, 2009; Blavet *et al.*, 2011; Muyle *et al.*, 2012). The knowledge of diversity and population subdivision is essential for such evolutionary and ecological studies; for example, differentiated

clusters may display different degrees of reproductive isolation with closely related species or different degrees of degeneration or suppressed recombination on sex chromosomes. In addition, the invasion of *S. latifolia* in the United States has been studied only with knowledge of two large genetic clusters in Europe (Keller & Taylor, 2008, 2010; Keller *et al.*, 2009, 2012), and it will be highly informative to include the newly identified Italian cluster in future studies investigating invasion dynamics and the evolutionary and ecological processes underlying invasions.

Co-structure between the host Silene latifolia *and pathogen* Microbotryum lychnidis-dioicae

In *Microbotryum lychnidis-dioicae*, the strongest genetic structure divided the samples into three genetic clusters, as previously reported using the same microsatellite dataset as well as using nuclear gene sequences (Vercken *et al.*, 2010; Gladieux *et al.*, 2011, 2013). These three genetic clusters closely matched those observed in *S. latifolia,* with the same Southern, Western and Eastern clusters. This indicates that the anther-smut pathogen remained during the last glaciation in all of the three refugia of its host (i.e. in the Iberian, Italian and Balkan peninsulas), leaving no major disease-free area.

Our extensive sampling of host–pathogen pairs from the same individual plants allowed comparison of the population structures of the two interacting species at a very fine level across Europe. The population structures were strikingly congruent between the pathogen and its host. It is noteworthy that the population structures of both *S. latifolia* and its pathogen indicates the rarity of recent effective long-distance dispersal events, whereas both are known to be able to disperse across large distances. For instance, both have invaded large ranges of North America within the last centuries (Keller *et al.*, 2012; Fontaine *et al.*, 2013; Gladieux *et al.*, 2015). In Europe, recolonization after glaciation may also have occurred through long-distance dispersal events (Bialozyt *et al.*, 2006). Yet the three main clusters identified in Europe in the plant and the pathogen were clearly circumscribed geographically, suggesting that they remained isolated one from each other for thousands of generations.

One of the factors possibly explaining these discrepancies is local adaptation to the host. A previous study at a smaller geographical scale showed that the plant was adapted locally to its anther-smut fungus: thus, *S. latifolia* individuals were more resistant to their local *M. lychnidis-dioicae* strains than to geographically more distant strains (Kaltz *et al.*, 1999; Kaltz & Shykoff, 2002). The results presented here suggest that this pattern of local adaptation also holds at the

cluster level. Lower resistance to foreign pathogens may cause a form of immigrant sterility by preventing *S. latifolia* genotypes to establish in regions where local *M. lychnidis-dioicae* populations belong to a cluster that is different from the cluster present in their region of origin, as the immigrant plant genotypes would be sterilized more often by local pathogens. The pattern of local adaptation supports suggestions of ongoing cycles of coevolution, with the plant running faster in the arms race thanks to its higher levels of interpopulation migration levels (Kaltz & Shykoff, 2002).

Local adaptation of *S. latifolia* to their anther-smut pathogens could, by contrast, favour long-distance establishment of the fungus, as our results suggest that immunity against foreign pathogens is lower than immunity against local ones. However, another factor may limit pathogen migration and contribute to the relative isolation among geographically-circumscribed genetic clusters of the anther-smut fungus. A competitive exclusion of distantly related genotypes by resident genotypes has been shown previously to occur in *M. lychnidis-dioicae* (Hood, 2003; López-Villavicencio *et al.*, 2007, 2011; Buono *et al.*, 2014). This relatedness-dependent competitive exclusion may prevent long-distance migrants to establish in genetically different clusters. Moreover, for both the host and the pathogen, local adaptation to climatic conditions may also play a role in the low effective dispersal across Europe. If the different genetic clusters are locally adapted to abiotic conditions, migrants may be prevented from establishment by being poorer competitors in abiotic conditions to which they are ill-suited.

The significant correlations between genetic distance matrices for the host and the pathogen further revealed that their population structures were congruent even at the within-cluster scale. Such highly congruent population structures are also likely due, in part, to the anther-smut fungus being an obligate pathogen (i.e. being unable to grow outside its host plant). In addition, the pathogen spores are dispersed by the same insect vectors serving as pollinators for the plant (Jennersten, 1988), increasing the probability that the host plant and its anther-smut fungus followed the same recolonization pathways at fine scales. It would be interesting in future studies to also assess the genetic co-structure between the plants the fungus and the pollinators. In fact, a previous work at a smaller geographical scale found a correlation between among-population genetic distances of *S. latifolia* and *Hadena biscuris,* a pollinator and seed-eater of the plant (Magalhaes *et al.*, 2011).

The congruence of population structures between *S. latifolia* and its *M. lychnidis-dioicae* pathogen appeared even stronger than what could be expected by geography and dispersal alone. Indeed, the partial Mantel tests revealed that the correlations between genetic distances of pairs of individuals between the host and the pathogen remained significant or close to significant

when controlling for isolation by distance effect. This suggests either that similar dispersal pathways have been taken by the host and the pathogen, that were not linear with geographical distance, and/or that coevolution has played a role in the congruence of the population structures. Indeed, the anther-smut fungus may more frequently disperse to host populations that carry similar host genotypes as its population-of-origin on these particular host genotypes. As explained above, genotype-by-genotype effects on disease probability have actually been reported in experiments in the *S. latifolia–M. lychnidis-dioicae* system, showing mostly local adaptation in the host more than in the pathogen (Kaltz *et al.*, 1999; Kaltz & Shykoff, 2002).

Stronger genetic subdivision in the pathogen Microbotryum lychnidis-dioicae *than in the host* Silene latifolia

The strongest genetic structure in *M. lychnidis-dioicae* was the three-cluster structure across Europe, with the Eastern, Northern and Southern clusters. However, Tess detected further subdivisions in the fungus. As suggested previously (Vercken *et al.*, 2010), these subdivisions may be a clue to a more complex history than the classic model of the Iberian, Italian and Balkan refugia. The split in the Eastern cluster is consistent with recolonization from two distinct refugia, one in the Balkans and the other further East. The lack of such subdivision in the host *S. latifolia,* despite more markers being used than in the fungus, and despite the markers being more polymorphic, might suggest that the host plant distribution was larger than that of its anther-smut pathogen during the last glaciation, spanning across both the Balkans and further East, whereas the distribution of *M. lychnidis-dioicae* would have been more fragmented. This would be consistent with contemporaneous observations that large areas colonized by *S. latifolia* are disease-free (T. Giraud, pers. obs.).

An alternative, although not exclusive, interpretation is that several life-history traits of both host and pathogen may contribute to a higher genetic subdivision in the fungus than in the plant, even if their distribution ranges have remained quite similar in the southern refugia. First, *S. latifolia* gene flow is mediated both by dispersal of its seed and pollination, whereas *M. lychnidis-dioicae* gene flow occurs only via spore transport by pollinators as seeds do not carry the fungus (Baker, 1947). In addition, the pollinators discriminate against the infected flowers (Shykoff & Bucheli, 1995) which may further reduce the extent of pathogen gene flow compared to its host. Finally, the mating systems are important life-history traits that may explain lower levels of gene flow among populations in the fungus than in the plant. Although F_{IS} values obtained here and previously suggest that some inbreeding occurs in this dioecious

plant (Delmotte *et al.*, 1999; Magalhaes *et al.*, 2011; Keller *et al.*, 2012) despite dioecy and inbreeding depression (Teixeira *et al.*, 2009), the fungus has a much more closed mating system, being highly selfing (Delmotte *et al.*, 1999; Hood & Antonovics, 2000; Giraud, 2004; Giraud *et al.*, 2008). Therefore, it may well be that the stronger subdivision in the anther-smut fungus reflects fragmented ranges in southern refugia and/or more northern refugia, but that the higher level of gene flow in the plant has erased their footprints by homogenizing allele frequencies.

Conclusion

In fungal plant pathogens, studies of genetic structures at continental scales have been performed almost exclusively for crop diseases, where the host population is often genetically homogeneous across the landscape and pathogen dispersal is impacted heavily by anthropogenic introductions and regional transport (Enjalbert *et al.*, 2005; Barres *et al.*, 2008; Fournier & Giraud, 2008; Gladieux *et al.*, 2008; Saleh *et al.*, 2014). The pathogen population structures in these cases were either very weak at a continental scale (Barres *et al.*, 2008; Fournier & Giraud, 2008; Saleh *et al.*, 2014) or reflected temperature adaptation (Enjalbert *et al.*, 2005; Mboup *et al.*, 2012). The present study therefore brings novel insights into the co-structure and coevolutionary patterns of plants and their pathogens, in a natural system, revealing strikingly strong congruent subdivisions. This study is one of the first to be conducted at this spatial scale with such comprehensive and paired sampling, associated with inoculation experiments, and thus provides a benchmark to develop future studies. Studies on a variety of other systems are needed to assess whether congruence population structure between host and pathogen is a general phenomenon. The knowledge of the co-structure is essential for understanding process of co-local adaptation between hosts and pathogens (Gandon *et al.*, 1996).

Acknowledgements

We thank the *Plateforme de Génotypage* GENTYANE INRA UMR1095 and William Fontanaud for assistance with genotyping. We thank all those who helped with sampling seeds, plant material and pathogens (Vercken *et al.*, 2010; Gladieux *et al.*, 2015). This work was supported by the European Research Council (starting grant GenomeFun 309403 to T.G.), the Agence Nationale de la Recherche (grants GANDALF ANR-12-ADAP-0009 and Emerfundis 07-BDIV-003), and the National Science Foundation (grant DEB-1115765 to M.E.H.).

Author contributions

T.G. and M.E.H. designed and supervised the study; A.F. performed the population genetics analyses with the help of T.G., P.G. and A.C.; M.E.H. and L.R. performed the inoculation experiments; A.F. and A.S. genotyped the plants; T.G. and M.E.H. contributed to the funding of the study.

References

Amico GC, Nickrent DL. 2009. Population structure and phylogeography of the mistletoes *Tristerix corymbosus* and *T. aphyllus* (Loranthaceae) using chloroplast DNA sequence variation. *American Journal of Botany* 96(8): 1571-1580.

Badouin H, Hood M, Gouzy J, Aguileta G, Siguenza S, Perlin M, Cuomo C, Fairhead C, Branca A, Giraud T. 2015. Chaos of rearrangements in in the mating-type chromosomes of the anther-smut fungus *Microbotryum lychnidis-dioicae*. *Genetics* 200: 1275-1284.

Baker HG. 1947. Infection of species of *Melandrium* by Ustilago violacea (Pers.) Fuckel and the transmission of the resultant disease. *Annals of Botany* 11(43): 333-348.

Barluenga M, Austerlitz F, Elzinga J, Teixeira S, Goudet J, Bernasconi G. 2011. Fine-scale spatial genetic structure and gene dispersal in *Silene latifolia*. *Heredity* 106: 13-24.

Barres B, Halkett F, Dutech C, Andrieux A, Pinon J, Frey P. 2008. Genetic structure of the poplar rust fungus Melampsora larici-populina: Evidence for isolation by distance in Europe and recent founder effects overseas. *Infection Genetics and Evolution* 8(5): 577-587.

Barrett LG, Thrall PH, Burdon JJ, Linde CC. 2008. Life history determines genetic structure and evolutionary potential of host-parasite interactions. *Trends in Ecology & Evolution* 23(12): 678-685.

Bernasconi G, Antonovics J, Biere A, Charlesworth D, Delph LF, Filatov D, Giraud T, Hood ME, Marais GAB, McCauley D, et al. 2009. Silene as a model system in ecology and evolution. *Heredity* 103(1): 5-14.

Bialozyt R, Ziegenhagen B, Petit RJ. 2006. Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology* 19(1): 12-20.

Blavet N, Charif D, Oger-Desfeux C, Marais G, Widmer A. 2011. Comparative high-throughput transcriptome sequencing and development of SiESTa, the *Silene* EST annotation database. *BMC Genomics* 12(1): 376.

Buono L, López-Villavicencio M, Shykoff J, Snirc A, Giraud T. 2014. Influence of multiple infection and relatedness on virulence: disease dynamics in an experimental plant population and its castrating parasite. *PLoS One* 9: e98526.

Criscione CD, Poulin R, Blouin MS. 2005. Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Mol Ecol* 14: 2247-2258.

de Vienne DM, Refregier G, Lopez-Villavicencio M, Tellier A, Hood ME, Giraud T. 2013. Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist* 198(2): 347-385.

Delmotte F, Bucheli E, Shykoff JA. 1999. Host and parasite population structure in a natural plant-pathogen system. *Heredity* 82: 300-308.

Dybdahl MF, Lively CM. 1996. The geography of coevolution: Comparative population structures for a snail and its trematode parasite. *Evolution* 50(6): 2264-2275.

Enjalbert J, Duan X, Leconte M, Hovmoller M, De Vallavieille-Pope C. 2005 Genetic evidence of local adaptation of wheat yellow rust *(Puccinia striiformis f. sp tritici)* within France *Mol Ecol* 14 2065-2073

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

Fontaine MC, Gladieux P, Hood ME, Giraud T. 2013. History of the invasion of the anther smut pathogen on Silene latifolia in North America. *New Phytologist* 198(3): 946-956.

Fournier E, Giraud T. 2008. Sympatric genetic differentiation of populations of a generalist pathogenic fungus, Botrytis cinerea, on two different host plants, grapevine and bramble. *J Evol Biol* 21: 122-132.

Gandon S, Capowiez Y, Dubois Y, Michalakis Y, Olivieri I. 1996. Local adaptation and gene-for-gene coevolution in a metapopulation model. *Proceedings of the Royal Society of London Series B-Biological Sciences* 263(1373): 1003-1009.

Giraud T. 2004. Patterns of within population dispersion and mating of the fungus *Microbotryum violaceum* parasitising the plant *Silene latifolia*. *Heredity* 93: 559-565.

Giraud T, Jonot O, Shykoff JA. 2005. Selfing propensity under choice conditions in a parasitic fungus, Microbotryum violaceum, and parameters influencing infection success in artificial inoculations *Int. J. Plant Sci.* 166: 649-657.

Giraud T, Yockteng R, Lopez-Villavicencio M, Refregier G, Hood ME. 2008. The mating system of the anther smut fungus, *Microbotryum violaceum*: selfing under heterothallism. *Euk. Cell* 7: 765-775.

Gladieux P, Devier B, Aguileta G, Cruaud C, Giraud T. 2013. Purifying selection after episodes of recurrent adaptive diversification in fungal pathogens. *Infection Genetics and Evolution* 17: 123-131.

Gladieux P, Feurtey A, Hood M, Snirc A, Clavel J, Dutech C, Roy M, Giraud T. 2015. The population biology of fungal invasions. *Mol Ecol.* 24:1969-86.

Gladieux P, Vercken E, Fontaine M, Hood M, Jonot O, Couloux A, Giraud T. 2011. Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Mol. Biol. Evol.* 28: 459-471.

Gladieux P, Zhang X-G, Afoufa-Bastien D, Sanhueza R-MV, Sbaghi M, Le Cam B. 2008. On the origin and spread of the scab disease of apple: out of Central Asia. *Plos One* 3:e1455.

Guillot G. 2009. On the inference of spatial structure from population genetics data using the Tess program. *BIOINFORMATICS* 25:1796-801.

Hafner MS, Page RDM. 1995. Molecular phylogenies and host-parasite cospeciation: gophers and lice as a model system. *Philos Trans R Soc London* 349: 77-83.

Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* 68: 87-112.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 359(1442): 183-195.

Hood M, Mena-Alí J, Gibson A, Oxelman B, Giraud T, Yockteng R, Arroyo M, Conti F, Pedersen A, Gladieux P, et al. 2010. Distribution of the anther-smut pathogen Microbotryum on species of the Caryophyllaceae assessed from natural history collections. *New Phytol* 187: 217-229.

Hood ME. 2003. Dynamics of multiple infection and within-host competition by the anther-smut pathogen. *The American Naturalist* 162(1): 122-133.

Hood ME, Antonovics J. 2000. Intratetrad mating, heterozygosity, and the maintenance of deleterious alleles in *Microbotryum violaceum (=Ustilago violacea). Heredity* 85: 231-241.

Ironside JE, Filatov DA. 2005. Extreme population structure and high interspecific divergence of the Silene Y chromosome. *Genetics* 171(2): 705-713.

Jakobsson M, Rosenberg N. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801-1806.

Jennersten O. 1988. Insect dispersal of fungal disease: effects of Ustilago infection on pollinator attraction in Viscaria vulgaris. *Oikos* 51: 163-170.

Kaltz O, Gandon S, Michalakis Y, Shykoff J. 1999. Local maladaptation of the plant pathogen *Microbotryum violaceum* to its host *Silene latifolia* : Evidence from a cross-inoculation experiment. *Evolution* 53: 395-407.

Kaltz O, Shykoff J. 2002. Within- and among-population variation in infectivity, latency and spore production in a host-pathogen system. *JEB* 15: 850-860.

Karrenberg S, Favre A. 2008. Genetic and ecological differentiation in the hybridizing campions *Silene dioica* and *S. latifolia*. *Evolution* 62(4): 763-773.

keller S, Taylor D. 2008. History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. *Ecology Letters* 11: 852-866.

Keller SR, Gilbert KJ, Fields PD, Taylor DR. 2012. Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, Silene latifolia. *Molecular Ecology* 21(19): 4721-4734.

Keller SR, Sowell DR, Neiman M, Wolfe LM, Taylor DR. 2009. Adaptation and colonization history affect the evolution of clines in two introduced species. *New Phytologist* 183(3):678-90.

Keller SR, Taylor DR. 2010. Genomic admixture increases fitness during a biological invasion. *Journal of Evolutionary Biology* 23(8): 1720-1731.

Langella O. 2002. POPULATIONS 1.2.28. Population genetic software (individuals or populations distances, phylogenetic trees). Available from http://bioinformatics.org/~tryphon/populations/.

López-Villavicencio M, Courjol F, Gibson A, Hood M, Jonot O, Shykoff J, Giraud T. 2011. Competition, cooperation among kin and virulence in multiple infections. *Evolution* 65:1357-1366.

Lopez-Villavicencio M, Jonot O, Coantic A, Hood M, Enjalbert J, Giraud T. 2007. Multiple infections by the anther smut pathogen are frequent and involve related strains. *PloS Pathogens* 3: e176.

Magalhaes IS, Gleiser G, Labouche AM, Bernasconi G. 2011. Comparative population genetic structure in a plant-pollinator/seed predator system. *Molecular Ecology* 20(22): 4618-4630.

Marais GAB, Nicolas M, Bergero R, Chambrier P, Kejnovsky E, Moneger F, Hobza R, Widmer A, Charlesworth D. 2008. Evidence for degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. *Current Biology* 18(7): 545-549.

Mboup M, Bahri B, Leconte M, De Vallavieille-Pope C, Kaltz O, Enjalbert J. 2012. Genetic structure and local adaptation of European wheat yellow rust populations: the role of temperature-specific adaptation. *Evolutionary Applications* 5(4): 341-352.

McCOY KD, BOULINIER T, TIRARD C. 2005. Comparative hostparasite population structures: disentangling prospecting and dispersal in the black-legged kittiwake *Rissa tridactyla*. *Mol Ecol* 14: 2825-2838.

Michalakis Y, Sheppard AW, Noel V, Olivieri I. 1993. POPULATION-STRUCTURE OF A HERBIVOROUS INSECT AND ITS HOST-PLANT ON A MICROGEOGRAPHIC SCALE. *Evolution* 47(5): 1611-1616.

Miura O, Torchin ME, Kuris AM, Hechinger RF, Chiba S. 2006. Introduced cryptic species of parasites exhibit different invasion pathways. *Proceedings of the National Academy of Sciences of the United States of America* 103(52): 19818-19823.

Moccia MD, Oger-Desfeux C, Marais GAB, Widmer A. 2009. A White Campion *(Silene latifolia)* floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping. *Bmc Genomics* 10: 243.

Muir G, Filatov D. 2007. A Selective Sweep in the Chloroplast DNA of Dioecious Silene (Section Elisanthe). *Genetics* 177(2): 1239-1247.

Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB. 2012. Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *Plos Biology* 10(4): e1001308.

Nakao M, Xiao N, Okamoto M, Yanagida T, Sako Y, Ito A. 2009. Geographic pattern of genetic variation in the fox tapeworm Echinococcus multilocularis. *Parasitology International* 58(4): 384-389.

Nieberding CM, Durette-Desset MC, Vanderpoorten A, Casanova JC, Ribas A, Deffontaine V, Feliu C, Morand S, Libois R, Michaux JR. 2008. Geography and host biogeography matter for understanding the phylogeography of a parasite. *Molecular Phylogenetics and Evolution* 47(2): 538-554.

Nieberding CM, Olivieri I. 2007. Parasites: proxies for host genealogy and ecology? *Trends in Ecology & Evolution* 22(3): 156-165.

Oudemans PV, Alexander HM, Antonovics J, Altizer S, Thrall PH, Rose L. 1998. The distribution of mating-type bias in natural populations of the anther-smut *Ustilago violacea* on *Silene alba* in Virginia. *Mycologia* 90: 372-381.

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20(2): 289-290.

Rich KA, Thompson JN, Fernandez CC. 2008. Diverse historical processes shape deep phylogeographical divergence in the pollinating seed parasite Greya politella. *Molecular Ecology* 17(10): 2430-2448.

Rousset F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol. Ecol. Resources* 8: 103-106.

Saleh D, Milazzo J, Adreit H, Fournier E, Tharreau D. 2014. South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, Magnaporthe oryzae. *New Phytologist* 201(4): 1440-1456.

Shykoff JA, Bucheli E. 1995. Pollinator visitation patterns, floral rewards and the probability of transmission of *Microbotryum violaceum,* a veneral disease plant. *Journal of Ecology* 83: 189-198.

Szpiech Z, Jakobsson M, Rosenberg N. 2008. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* 24: 2498-2504.

Taylor DR, Keller SR. 2007. Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution* 61(2): 334-345.

Teixeira S, Foerster K, Bernasconi G. 2009. Evidence for inbreeding depression and post-pollination selection against inbreeding in the dioecious plant Silene latifolia. *Heredity* 102(2): 101-112.

Thomas A, Shykoff J, Jonot O, Giraud T. 2003. Mating-type ratio bias in populations of the phytopathogenic fungus *Microbotryum violaceum* from several host species. *Int. J. Plant Sci.* 164: 641-647.

Thrall PH, Biere A, Uyenoyama MK. 1995. Frequency-dependant disease transmission and the dynamics of the Silene-Ustilago host-pathogen system. *Am. Nat.* 145(1): 43-62.

Thrall PH, Jarosz AM. 1994. Host-pathogen dynamics in experimental populations of Silene alba and Ustilago violacea. I. Ecological and genetic determinants of disease spread. *Journal of Ecology* 82: 549-559.

Tollenaere C, Susi H, Nokso-Koivisto J, Koskinen P, Tack A, Auvinen P, Paulin L, Frilander MJ, Lehtonen R, Laine AL. 2012. SNP design from 454 sequencing of *Podosphaera plantaginis* transcriptome reveals a genetically diverse pathogen metapopulation with high levels of mixed-genotype infection. *Plos One* 7(12): e52492.

Tsai Y-HE, Manos PS. 2010. Host density drives the postglacial migration of the tree parasite, Epifagus virginiana. *Proceedings of the National Academy of Sciences of the United States of America* 107(39): 17035-17040.

Vercken E, Fontaine M, Gladieux P, Hood M, Jonot O, Giraud T. 2010. Glacial refugia in pathogens: European genetic structure of anther smut pathogens on Silene latifolia and S. dioica. *PloS Pathogens* 6: e1001229.

Whiteman NK, Parker PG. 2005. Using parasites to infer host population history: a new rationale for parasite conservation. *Animal Conservation* 8: 175-181.

Williams PD. 2010. Darwinian interventions: taming pathogens through evolutionary ecology. *Trends in Parasitology* 26(2): 83-92.

Wilson DJ, Falush D, McVean G. 2005. Germs, genomes and genealogies. *Trends in Ecology & Evolution* 20(1): 39-45.

Wirth T, Meyer A, Achtman M. 2005. Deciphering host migrations and origins by means of their microbes. *Molecular Ecology* 14(11): 3289-3306.

Supplementary Information

Note S1: Identification of species

Our plant dataset initially contained individuals that were identified by the samplers as *S. latifolia* (n=592 individuals), *S. dioica* (*n*=74) and putative hybrids between these species (*n*=50). The initial anther-smut fungi dataset initially contained samples that were identified by the samplers as collected on *S. latifolia* (n=669), on *S. dioica* (371) and on hybrids between the two plant species (n=30).

We checked species identity with a STRUCTURE analysis (Falush *et al.*, 2003) on the complete datasets, for both plants and pathogens. This clustering software is based on Markov Chain Monte Carlo (MCMC) simulations and group multilocus genotypes into K gene pools in order to minimize deviations from Hardy-Weinberg and linkage equilibrium within gene pools. Each run consisted of 500,000 MCMC iterations after a burn-in of 50,000 iterations using a model allowing for admixture between gene pools and correlated allele frequencies. As this software is based on the assumption of Hardy-Weinberg equilibrium, the presence in the dataset of individuals sharing the same genotypes due to high levels of relatedness may lead to the assignment of these highly related individuals in spurious separate clusters. To avoid such a problem with our highly selfing pathogen, only one representative of each genotype was kept in the dataset for STRUCTURE analyses. The removed genotypes were then assigned to the same cluster as the genotype present in the STRUCTURE analyses to which they were identical.

The individuals assigned with a membership coefficient >0.9 in one gene pool were considered as "pure" representative of their species. Previous studies have shown that these parameters yielded robust species assignment, congruent and more finely discriminating than gene sequences or host plant collection (Vercken et al., 2010; Gladieux et al., 2011). This left 578 and 679 genotypes as 'pure' representative of *S. latifolia* and *M. lychnidis-dioicae*, respectively (Fig. S1), representing the individuals retained for subsequent analyses. Among these pure genotypes, we could obtain both the genotype of the plant and of its associated anther-smut pathogen for 391 samples.

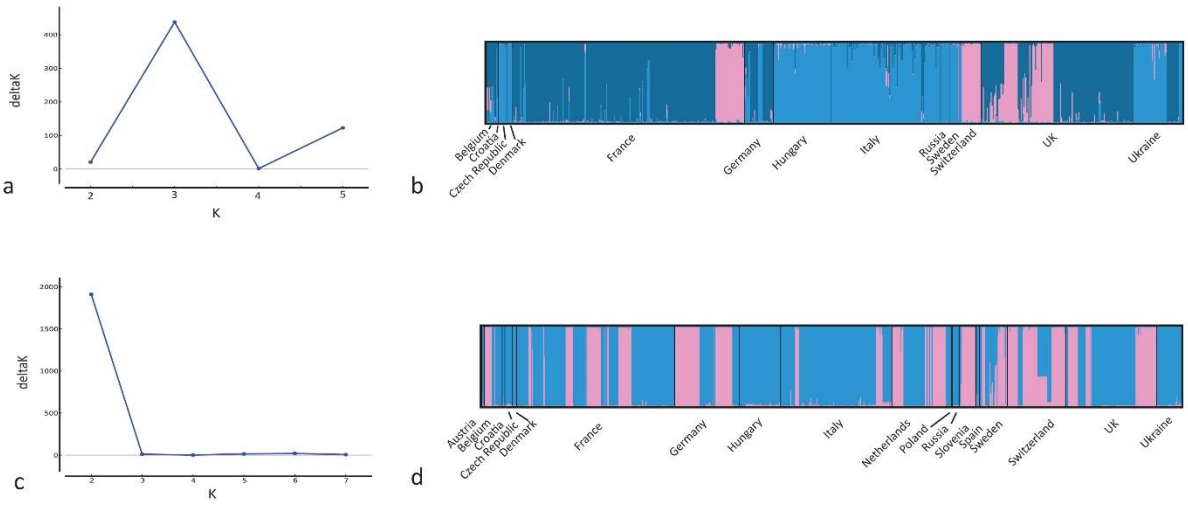


Figure S1. Identification of species. (a) ΔK statistic for the STRUCTURE analysis on the *Silene latifolia* and *Silene dioica* datasets (see Supplementary Text for detailed methods); (b) Membership proportions per individual (as vertical bars) from the STRUCTURE results at K=3 on the *Silene latifolia* (in light and dark blue) and *Silene dioica* (in pink) dataset. (c) ΔK statistic for the STRUCTURE analysis on the *Microbotryum lychnidis-dioicae* and *Microbotryum silenes-dioicae* dataset; (d) Membership proportions per individual (as vertical bars) inferred with STRUCTURE at K=2 on the *Microbotryum lychnidis-dioicae* (in blue) and *Microbotryum silenes-dioicae* (in pink) dataset.

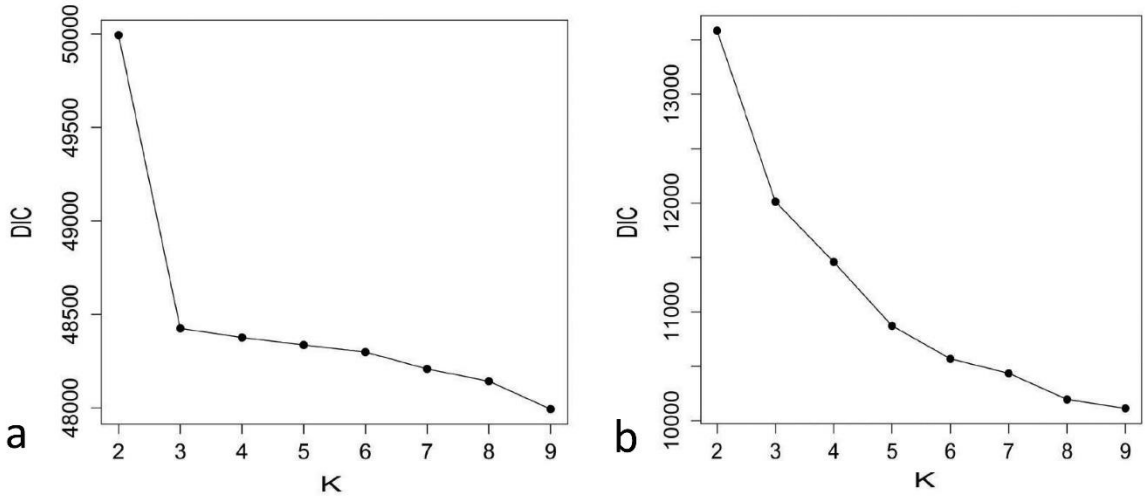


Figure S2. Plot of the deviance information criterion (DIC) given by the software TESS used to estimate the number of clusters corresponding to the strongest genetic structure for (a) *Silene latifolia* and (b) *Microbotryum lychnidis-dioicae*.

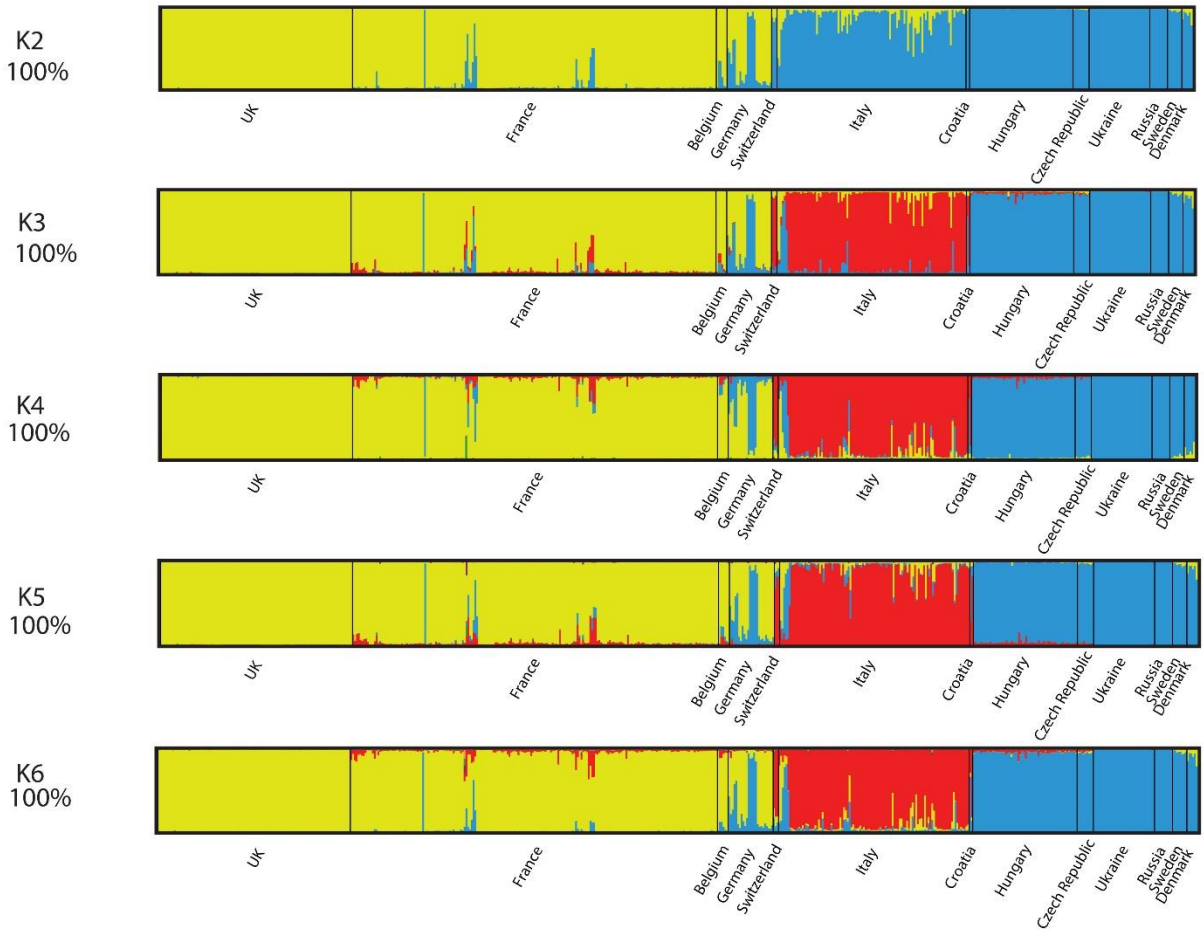


Figure S3. Membership proportions per *Silene latifolia* individual (as vertical bars) from the TESS results, from K=2 to K=6. Only the major solutions (found in the majority of runs) are shown.

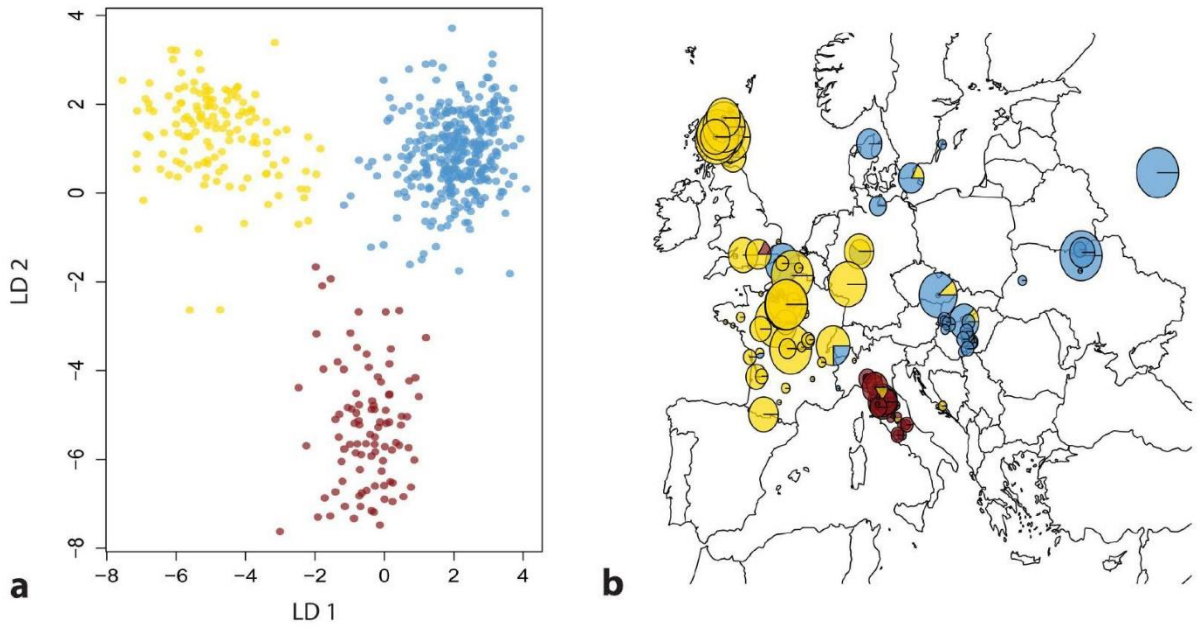


Figure S4. Plots of the results from the discriminant analyses of principal component (DAPC) on the *Silene latifolia* dataset. (a) Plot of the first two axes. The dot colors correspond to the groups found with the function find.clusters (adegenet R package). (b) Map of the mean assignment of individuals to the groups found with the function find.clusters (adegenet R package).

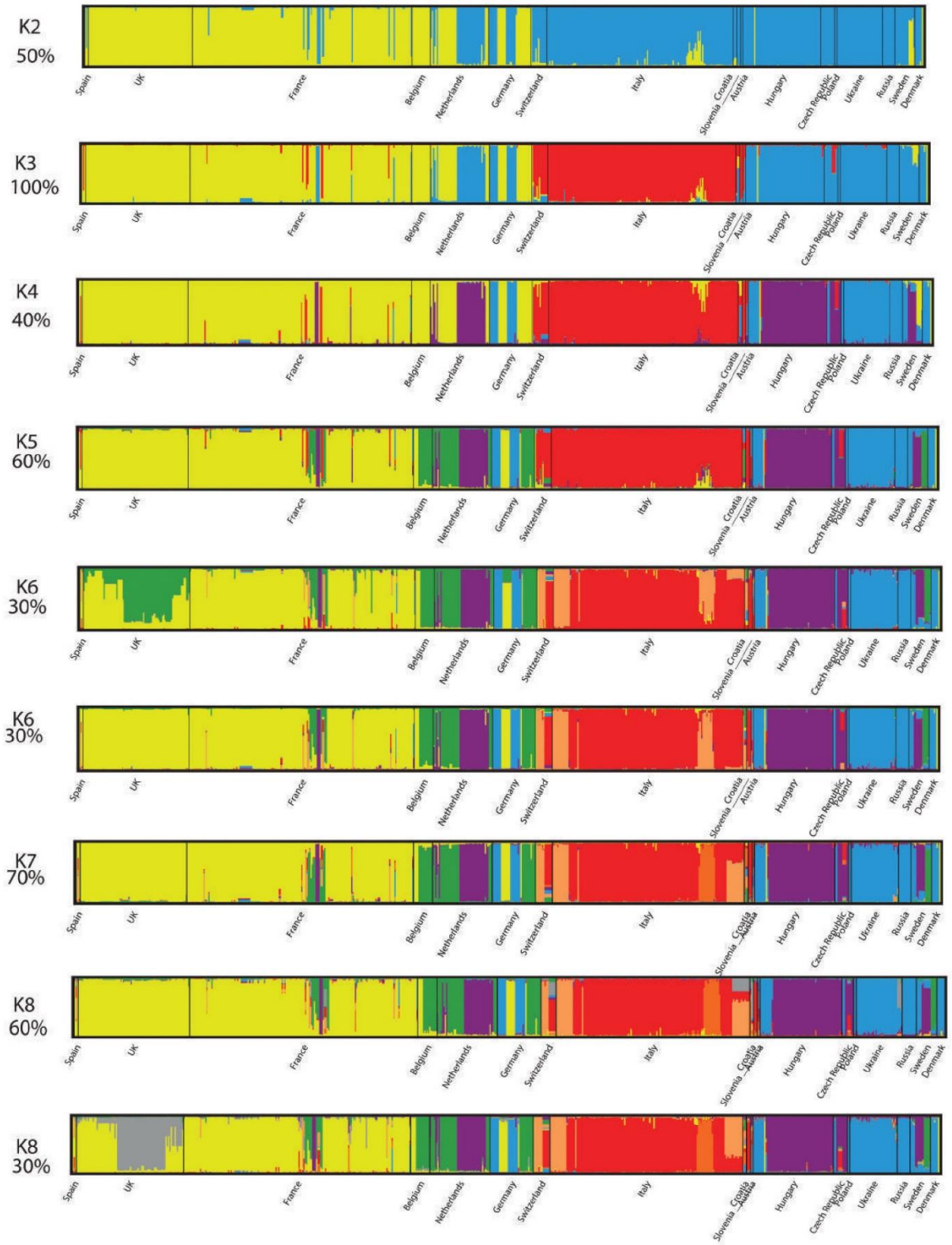


Figure S5. Membership proportions per *Microbotryum lychnidis-dioicae* individual (as vertical bars) inferred with TESS from K=2 to K=8.

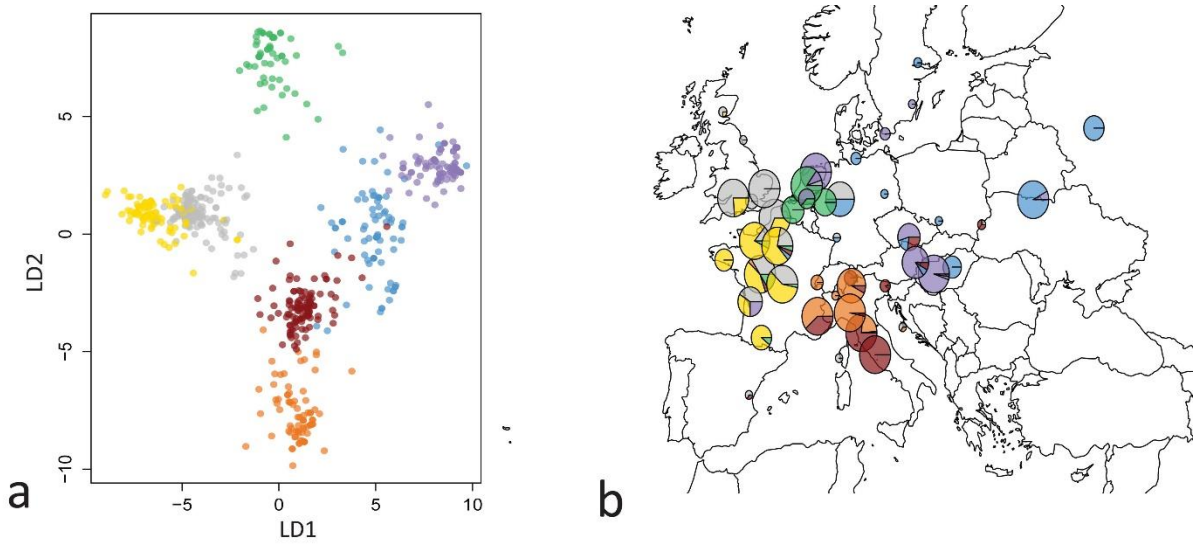


Figure S6. Discriminant analyses of principal component (DAPC) on the *Microbotryum lychnidis-dioicae* dataset. (a) Plot of the first two linear discriminants. The dot colors correspond to the groups found with the function find.clusters (adegenet R package). (b) Map of the mean assignment of individuals to the groups found with the function find.clusters (adegenet R package). The placement of the clusters found with TESS was extremely similar.

Table S1. Details of the multiplex PCR used to amplify microsatellite markers in Silene latifolia.

	Dye	Repeat	Range
SIL1			
SLeSSR06	FAM	(AAT)9	141-193
SLeSSR20	ATTO 565	(TGA)5	179-194
SLeSSR23	ATTO 550	(CCCAAT)4.7	186-219
SIL2			
SLeSSR12	FAM	(TC)8	144-190
SLeSSR13	ATTO 565	(AG)11	180-213
SLeSSR29	ATTO 550	(AAG)8	196-296
SIL3			
SLeSSR07	FAM	(TCAAT)9	127-185
SLeSSR14	ATTO 550	(AGC)9.7	303-351
SLeSSR15	ATTO 565	(ATCAAG)4.8	176-244
SIL4			
SLeSSR01	ATTO 565	(AACCCA)4.7	198-227
SLeSSR27	FAM	(CATCCT)6.2	221-248
SLeSSR10	ATTO 550	(GT)9	310-344
SLeSSR28	Yakima Yellow	(AGA)5	172-199
SIL5			
SILeSSR02	ATTO 565	(TAATT)5	193-221
SILeSSR21	ATTO 550	(AAG)7	221-229
SILeSSR22	FAM	(ACA)8.3	144-189

Table S2. Details of the PCR program used to amplify microsatellite markers in *Silene latifolia*

	TEMPERATURE (°C)	TIME
FIRST DENATURATION AND ACTIVATION OF THE POLYMERASE	95	15 min
TOUCH-DOWN AMPLIFICATION CYCLES: 10 CYCLES	94	30s
	60 to 50	1 min 30
	72	1 min
AMPLIFICATION CYCLES: 25 CYCLES	94	30s
	52	1 min 30
	72	1 min 30
FINAL ELONGATION	72	15 min

Table S3. Cross-inoculation design, origin of the *Microbotryum lychnidis-dioicae* strains and *Silene latifolia* seeds (country and GPS coordinates).

Host Collection	Cluster	Local Seed/Pathogen, Latitude, Longitude	Inter-Cluster Pathogen, Latitude, Longitude
Scotland	Western	55.90, -2.27	Eastern, 51.18, 14.31
Hungary	Eastern	47.46, 19.14	Western, 55.90, -2.27
Czech Republic	Eastern	49.59, 17.25	Western, 48.70, 2.17
Germany	Eastern	51.18, 14.31	Southern, 44.32, 7.68

Table S4. Summary statistics for *Silene latifolia* for each of the three TESS genetic clusters. H_O: observed heterozygosity, H_E:expected heterozygosity, F_{IS} and allelic richness A_r estimated for a subsample of size 10.

Cluster	H_O	H_E	F_{IS}	A_r (N=10)
Eastern	0.4847	0.6340	0.2354	4.4100
Western	0.5151	0.6136	0.1606	3.8540
Southern	0.4757	0.5992	0.2061	3.9243

Table S5. Summary statistics for *Microbotryum lychnidis-dioicae* for each of the three TESS genetic clusters. H_O: observed heterozygosity, H_E:expected heterozygosity, F_{IS} and allelic richness A_r estimated for a subsample of size 10.

Cluster	H_O	H_E	F_{IS}	A_r (N=10)
Eastern	0.0251	0.5604	0.9553	3.3041
Western	0.0559	0.5210	0.8927	3.2504
Southern	0.0366	0.6389	0.9427	3.5399

Table S6. Logistic regression testing local adaptation at the cluster level in the system *Microbotryum lychnidis-dioicae* - *Silene latifolia*. See Table S3 for the inoculation design. Seedling origin is not significant as a main effect, but the treatment (local vs. inter-cluster) and its interaction with the seedling origin was significant.

	d.f.	ChiSquare	P-value
Seedling origin	3	0.0631	0.9959
Treatment (local vs. inter-cluster)	1	17.2545	<0.0001*
Seedling origin x Treatment	3	17.3114	0.0006*

References

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

Gladieux P, Vercken E, Fontaine M, Hood M, Jonot O, Couloux A, Giraud T. 2011. Maintenance of fungal pathogen species that are specialized to different hosts: allopatric divergence and introgression through secondary contact. *Mol. Biol. Evol.* 28: 459-471.

Vercken E, Fontaine M, Gladieux P, Hood M, Jonot O, Giraud T. 2010. Glacial refugia in pathogens: European genetic structure of anther smut pathogens on Silene latifolia and S. dioica. *PloS Pathogens* 6: e1001229.

IV. Discussion de la section 2

Les deux systèmes hôtes-pathogènes que j'ai étudiés comprennent une espèce végétale pour hôte et une espèce fongique comme pathogène. Néanmoins, les deux systèmes sont très différents puisque hôtes comme pathogènes ont des traits d'histoire de vie différents. Ainsi, les compagnons blancs ont un temps de génération d'une année et vivent 2-6 ans, tandis que les pommiers peuvent vivre plusieurs dizaines d'années et ne se reproduisent pas avant plusieurs années. Les pathogènes diffèrent notamment sur le mode de dispersion des propagules puisque *V. inaequalis* n'est dispersé qu'à très petite distance via des gouttelettes d'eau alors que *M. lychnidis-dioicae* est dispersé plus largement par les pollinisateurs (J K Bowen et al., 2011; López-Villavicencio et al., 2007).

L'influence anthropique distingue également les deux systèmes étudiés. En effet, l'évolution de l'un des deux systèmes a été fortement influencée par l'Homme puisque le pommier cultivé a subi un processus de domestication et que *V. inaequalis* est sujet à de nombreux traitements anti-fongiques dans les vergers. Dans le cas du système *Silene/Microbotryum*, toute influence humaine est inconsciente et se limite à la création/destruction d'habitats ou la dispersion de propagules telle que celle ayant permis à la plante comme au pathogène d'envahir l'Amérique du Nord (Fontaine et al., 2013). Dans le cas du système pommier/ *V.inaequalis*, à l'inverse, l'influence humaine est prononcée et volontaire. Du côté de l'hôte, *M. domestica* a été domestiqué à partir de *M. sieversii* en Asie Centrale puis transporté vers l'Europe, sélectionné puis re-transporté vers sa région d'origine (Amandine Cornille et al., 2014). Du côté du pathogène, la tavelure est un pathogène du pommier économiquement important et contre lequel de nombreux traitements sont effectués (MacHardy et al., 2001).

Venturia - Malus

L'histoire présumée du pommier implique une domestication à partir de *M. sieversii*. Le pommier cultivé a ensuite été importé dans toutes les régions tempérées du monde, créant des contacts secondaires avec différentes espèces de pommiers sauvages. Dans la première section de ce manuscrit, j'ai étudié les conséquences de la culture du pommier *M. domestica* en Europe dans l'aire de répartition de *M. sylvestris*. Cette culture semble dater d'au moins l'époque classique puisque les romains et les grecs font mention de ce fruit. Le contact secondaire que j'ai étudié dans la seconde partie de mon manuscrit est beaucoup plus récent. En effet, traditionnellement, les Kazakhs sont un peuple nomade, chasseurs et éleveurs. L'agriculture n'était donc pas développée dans cette région du monde avant l'annexion de ces territoires par l'Empire Russe au cours du XVIIIème siècle et l'évolution du mode de vie des populations locales (De Gracia et al., 2015).

Conséquences du contact secondaire chez le pommier

Comme nous l'avons vu dans la première section de ce manuscrit, les conséquences du contact entre pommier cultivé et pommier sauvage en Europe sont sévères pour *M. sylvestris*. A travers toute l'Europe, des hybrides entre les deux espèces ont été détectés et les populations naturelles, déjà clairsemées, notamment en raison des modifications des paysages engendrés par l'anthropisation des paysages, sont maintenant en danger de perdre leur intégrité génétique.

Dans les populations d'Asie centrale que nous avons étudiées, la situation parait différente. Ainsi, nous n'avons détecté que très peu de *M. domestica* dans les populations naturelles. Nous avons cependant identifié environ 10% d'hybrides répartis dans différentes populations avec pourtant trois sites ne comprenant aucun hybride ni aucun *M. domestica*. Il semblerait donc que les populations naturelles de pommiers dans cette région soient encore relativement épargnées par l'invasion de *M. domestica*.

Cette relative pureté des populations naturelles de montagnes kazakhes ne peut pas être expliquée par des barrières reproductives entre *M. sieversii* et *M. domestica*, puisque dans le seul cas de contact direct entre les deux espèces (un verger entouré de forêt), nous avons identifié un quart d'arbres hybrides. Il est donc possible que les deux cas de contacts secondaires étudiés dans ce manuscrit soient en fait deux étapes d'un même processus et que ce ne soit qu'une question de temps avant que les populations naturelles de *M. sieversii* soient elles-aussi menacées par les flux de gènes avec *M. domestica*.

Conséquences du contact secondaire chez Venturia inaequalis

Le contact secondaire dans les montagnes d'Asie centrale a apparemment eu des conséquences très différentes pour *V. inaequalis* et pour son hôte. Nous avons identifié la présence de pathogènes de type sauvage dans toutes les populations de forêt, comme nous l'attendions. De façon plus surprenante, nous avons également détecté des pathogènes de type agricole dans toutes les populations y compris celles dans lesquelles aucun arbre *M. domestica* ou hybride n'a été identifié. Cela signifie que, contrairement à son hôte, le pathogène ayant évolué en réponse à la domestication a envahi les populations naturelles des montagnes du Kazakhstan. Cette différence est étonnante lorsque l'on considère que les capacités de dispersion de *V. inaequalis* sont supposées être assez limitées. Après être projetées sur de très faibles distances (de l'ordre de quelques millimètres), les ascospores sont dispersées sur des distances inférieures à 100 m par le vent (MacHardy et al., 2001). Les meilleures chances de dispersion à longue distance proviennent du déplacement de fruits ou de feuilles infectées par l'Homme. Le pommier, en revanche, a des capacités de dispersion plus importante. Son pollen peut être transporté par les pollinisateurs sur des distances allant jusqu'à plusieurs kilomètres comme nous l'avons vu dans la section précédente.

Cette situation est à mettre en regard d'un autre contact secondaire entre deux populations de *V. inaequalis* étudié par Leroy et al (2016). Depuis 1914, le pommier japonais *M. floribunda* a été utilisé comme source de résistance à la tavelure pour l'amélioration variétale, pour son gène de résistance Rvi6, anciennement nommé *Vf*. Des souches de *V. inaequalis* virulentes contre les pommiers portant ce gène ont été observées dans les vergers et leur origine a récemment été identifiée comme provenant d'un saut d'hôte depuis *M. floribunda* (Lemaire et al., 2016). Deux vergers contenant des pommiers Rvi6 et des pommiers non-Rvi6 ont été étudiés comme zones de contact secondaire. Dans le premier verger (site 1), les hôtes sont les variétés « Judeline », porteuse de Rvi6, et « Antoinette », non-porteuse de Rvi6, mais portant d'autres gènes de résistance. Dans le second verger (site 2), les hôtes sont issus d'un croisement entre le cultivar « Prima », porteur de Rvi1 et de Rvi6, et le pommier TN10-8, ne portant aucune de ces résistances. Les conséquences du contact secondaire pour les populations *V. inaequalis* virulentes ou non sur les pommiers portant le gène Rvi6 dans ces deux vergers furent drastiquement différentes. Sur le premier site, très peu d'hybridations ont été observées entre les deux types de pathogènes. Dans le second verger, à l'inverse, toutes les souches, qu'elles aient été collectées sur des hôtes Rvi6 ou non, sont des souches hybrides entre les deux types de *V. inaequalis*. En effet, les résistances additionnelles de la variété « Antoinette » représentaient

certainement une barrière pour l'infection des pathogènes Rvi6 infectant « Judeline » et, puisque *V. inaequalis* se reproduit dans les feuilles, créaient ainsi une forte barrière écologique aux flux de gènes. A l'inverse, parmi les hôtes hybrides du second verger, il se trouvait certainement des hôtes sensibles aux deux types de pathogènes, ce qui a permis les croisements.

Nous savions que nous n'étions pas dans un cas d'isolement géographique en Asie centrale puisque nous avons détecté la présence des types agricole et sauvage de *V. inaequalis* dans tous les sites d'échantillonnage (à l'exception du verger où seul le type agricole a été identifié). Malgré cela, le pourcentage d'hybrides n'est pas extrêmement élevé, mais d'environ 10%, ce qui semble indiquer l'existence de barrières aux flux de gènes. Nous avons tout d'abord considéré la possibilité de barrières écologiques, mais les deux types de pathogènes partagent la même espèce hôte, *M. sieversii*, et même dans certains cas le même individu hôte. Pour comparer cette situation à l'étude concernant le contact secondaire entre les pathogènes de *M. domestica* et de *M. floribunda*, nous serions donc plutôt dans un cas similaire au verger numéro 2. Et pourtant nous n'observons pas la même issue au contact secondaire, puisqu'au lieu des 100% d'hybrides observés dans l'étude de Leroy et al (2016), nous n'en observons que 10%. De plus, le verger que nous avons échantillonné en Asie centrale existe depuis environ 30 ou 40 ans ce qui devrait être assez de temps pour permettre des croisements entre les *V. inaequalis* de type agricole et de type sauvage, si le pathogène a été apporté avec son hôte cultivé. Puisqu'il n'est pas envisageable d'expliquer l'isolement reproductif entre les deux souches par des barrières écologiques, il doit certainement exister d'autre types de barrières qui n'étaient pas présentes dans le cas des souches étudiés par Leroy et al (2016). Il faut noter que les populations de *V. inaequalis* étudiées dans le contact secondaire en Europe par Leroy et al. (2016) ne sont pas les mêmes populations que celles étudiées ici et ont donc eu effectivement des histoires évolutives différentes.

Les barrières en jeu dans le cas du contact secondaire en Asie centrale étudié ici pourraient être pré-zygotiques intrinsèques (choix actif d'un partenaire sexuel similaire) ou post-zygotiques comme des incompatibilités de Dobzhansky-Muller (Fishman & Willis, 2001). Il existe encore peu de cas pour lesquels les gène impliqués dans ce type d'incompatibilités ont été identifiés (principalement chez la drosophile, Brideau et al 2006). Identifier ce type de gènes chez *V. inaequalis* pourrait donc être une piste de recherche intéressante. Malheureusement, le trop faible nombre d'individus hybrides dans notre jeu de données ne nous donnera probablement pas la possibilité de tester cette hypothèse ici, et encore moins d'identifier des gènes potentiellement impliqués dans ce type d'interactions. Il est également possible que les barrières post-zygotiques soient plus complexes que le simple modèle de Dobzhansky-Muller, comme cela

semble être le cas pour les incompatibilités hybrides observées entre *Saccharomyces cerevisiae* et *S. paradoxus* (Kao, Schwartz, & Sherlock, 2010). Dans ce cas, un nombre encore plus important d'hybrides serait nécessaire pour détecter les gènes impliqués (une centaine d'hybrides analysés n'ont pas suffit dans l'étude de Kao, Schwartz, & Sherlock 2010). Comprendre les raisons de cet isolement reproducteur ainsi que les causes des différences observées entre les contacts secondaires étudiés ici et par Leroy et al (2016) serait une piste de recherche très intéressante pour l'avenir.

Capacités d'infection de V. inaequalis

Notre échantillonnage contenant des paires d'hôtes et de pathogènes, nous avons pu étudier les capacités d'infection de *V. inaequalis* en conditions naturelles. Nos observations ont confirmé les résultats issus des inoculations expérimentales de Le Van et al. (2012) : les pathogènes de type agricole sont capables de parasiter à la fois *M. sieversii* et *M. domestica* alors que les pathogènes de type sauvage ne sont capables d'infecter que les pommiers sauvages. De façon intéressante, les pathogènes hybrides n'ont eux-mêmes été récoltés que sur des pommiers sauvages et dans aucun cas nous n'avons trouvé de pathogène hybrides ni sur un pommier cultivé, ni sur un pommier hybride. Cette observation va à l'encontre de ce que d'autres cas de pathogènes hybrides auraient pu faire penser. Ainsi, le pathogène de l'oïdium des céréales capable d'infecter le triticale, hybride entre le blé et le seigle, est lui –même un hybride entre les pathogènes des deux progéniteurs (Menardo et al., 2015). Chez les arbres, un cas similaire a été observé avec le pathogène, *Melampsora x columbiana*, parasitant les peupliers issus d'un croisement entre *Populus trichocarpa* et *P. deltoides* et provenant lui-même d'une hybridation entre les pathogènes de ces deux espèces d'arbres (Newcombe, Stirling, McDonald, & Bradshaw, 2000).

Afin de déterminer si l'absence des hybrides de *V. inaequalis* sur les pommiers cultivés reflète réellement leurs capacités d'infection, nous avons décidé d'inoculer artificiellement les hybrides sur M. *domestica*. En effet, les sites que nous avons étudiés, à l'exception du verger, ne contiennent que peu de pommiers identifiés comme *M. domestica*, et nous voulons donc déterminer si l'absence de parasites hybrides infectant le pommier cultivé est dû à la faible quantité de ce type d'hôtes ou bien si toutes les souches hybrides échantillonnées sont réellement incapables d'infecter les pommiers cultivés. Cette étude expérimentale est toujours en cours au moment de la rédaction de ce manuscrit et les résultats ne sont donc pas présentés ici. Dans le cas où certaines souches hybrides seraient capables d'infecter des pommiers cultivés, nous pourrons peut-être les utiliser afin de détecter des loci liés à ces capacités d'infection.

Microbotryum lychnidis-dioicae -Silene latifolia

Une structure phylogéographique commune au pathogène et à son hôte

En étudiant la structure phylogéographique de *M. lychnidis-dioicae*, j'ai identifié trois groupes génétiques principaux, en accord avec l'analyse de ces données par Vercken et al. (2010), mais aussi une sous-structuration formée de sept groupes génétiques cohérents géographiquement. La structure phylogeographique chez le compagnon blanc semble plus simple que celle de son pathogène : nous n'avons détecté que trois groupes génétiques distincts. Par rapport aux études précédentes (Keller et al., 2012; Taylor & Keller, 2007), nous avons identifié la présence d'un groupe supplémentaire important en Europe, couvrant l'ensemble de l'Italie. Or les trois groupes principaux détectés chez *M. lychnidis-dioicae* correspondent de façon remarquable à ceux que nous avons identifiés pour le compagnon blanc, avec seulement 2% de paires hôtes-pathogènes dans lesquelles les deux espèces étaient assignées à des groupes géographiques différents.

La forte similitude entre les deux structures suggère une histoire commune des deux espèces : une contraction de leur aire de répartition dans les refuges glaciaires des péninsules ibérique, italienne et balkanique durant la dernière glaciation, suivie d'une recolonisation de l'Europe à la fin de l'ère glaciaire. L'existence de trois groupes génétiques couvrant l'Europe et résultant de la différenciation des populations isolées dans les péninsules du sud de l'Europe durant l'ère glaciaire est classique et a été trouvée dans le cas de nombreux organismes, bien que les schémas de recolonisation puissent varier (Hewitt, 1999; Hewitt, 2004). Dans le cas qui nous occupe ici, la recolonisation a été limitée, dans le cas des populations italiennes, par la barrière formée par les Alpes (ce qui fut également le cas pour le pommier sauvage étudié dans la première section de ce manuscrit) et s'est interrompue pour les deux autres groupes avec la formation d'une zone d'admixture en Allemagne. La congruence entre les structures de l'hôte et du pathogène signifie notamment que, bien qu'il existe aujourd'hui des zones où le compagnon blanc est présent sans infection par *M. lychnidis-dioicae*, il semblerait que le parasite ait été présent dans ces trois refuges durant la dernière glaciation.

Des flux de gènes limités entre groupes génétiques

Les trois groupes génétiques sont très bien définis géographiquement et nous n'avons identifié que peu d'hybrides entre groupes (environ 5% pour le pathogène et 6% pour la plante), ce qui indique une certaine rareté des évènements de dispersion à longue distance qui auraient pour effet d'homogénéiser les groupes. Ce faible nombre d'hybrides ne semble pas être un artefact résultant d'un nombre trop limité de marqueurs disponibles dans notre étude, au moins dans le cas de *M. lychnidis-dioicae*. En effet, ces trois groupes génétiques ont été identifiés également avec des SNPs couvrant l'ensemble du génome par (Badouin, 2016) et, même avec ce type de données, la structuration reste forte, avec des valeurs de F_{ST} entre groupes très élevées (avec un minimum de 0.56), et aucune admixture n'est détectée, confirmant les faibles niveaux de flux de gènes entre ces groupes. Les expériences d'inoculations artificielles que nous présentons dans l'article donnent un premier élément de réponse quant à cette rareté des évènements de dispersion à longue distance. En effet, nos résultats suggèrent que la plante est localement adaptée à son pathogène, ce qui pourrait limiter son implantation dans des zones géographiques où se trouvent les parasites d'un groupe génétique différent. En effet, les plantes pourraient y être stérilisées plus souvent que les plantes locales et ne pas parvenir à s'établir durablement. Dans ce cadre, il est surprenant de constater que *M. lychnidis-dioicae* n'est pas, lui, dispersé de façon efficace sur de longue distance, puisqu'il est capable de mieux infecter les plantes des groupes génétiques voisins. Il est possible que l'exclusion par le pathogène de génotypes trop différents du sien lors des infections multiples (López-Villavicencio et al., 2007) puisse limiter les événements de dispersion entre régions.

Une sous-structuration présente chez le pathogène, mais pas chez son hôte

Alors qu'aucun groupe génétique supplémentaire n'a été détecté chez la plante, une sous-structuration a été identifiée chez *M. lychnidis-dioicae*, avec l'existence de sept groupes génétiques cohérents géographiquement. Les résultats obtenus avec les génomes complets suggèrent également l'existence d'un groupe génétique supplémentaire dans la péninsule ibérique (Badouin, 2016), que nous n'avons pas détecté dans notre étude, très probablement en raison d'un trou dans notre échantillonnage dû à la rareté de la maladie dans cette région.

Cette sous-structure pourrait suggérer l'existence d'une histoire évolutive plus complexe que celle indiquée par l'existence de trois grands groupes correspondant aux refuges glaciaires des péninsules ibérique, italienne et balkanique. Comme dans le cas du pommier sauvage exposé dans la première section, la distribution fragmentée du pathogène pourrait être expliquée par

l'existence de « refuges à l'intérieur des refuges » dans lesquels le pathogène se serait différencié (Gómez & Lunt, 2007).

Une autre hypothèse implique les différences de traits d'histoire de vie entre *S. latifolia* et *M. lychnidis-dioicae*, comme les capacités de dispersion ou les systèmes de reproduction. Cette seconde hypothèse n'exclue pas nécessairement celle de refuges plus fragmentés. Par exemple, les flux de gènes entre les plantes sont médiés à la fois par la dispersion de leur pollen, transportés par les pollinisateurs dont profitent également les spores du champignon, mais aussi par leurs graines. Or, des capacités de dispersion plus importantes pourraient lisser chez la plante des différences subtiles entre populations provenant d'une histoire faisant intervenir de nombreux micro-refuges. De plus, le champignon a un système de reproduction très autogame ce qui pourrait favoriser l'acquisition plus rapide d'une structuration forte et locale, puisque ce type de reproduction accroît la dérive génétique et réduit les flux de gènes (Nieberding & Olivieri, 2007).

Plus largement, cette étude est une des premières étudiant la costructure d'un système hôte-pathogène à une échelle géographique aussi grande et un échantillonnage aussi dense. Les études de costructures entre hôtes et pathogènes restent d'ailleurs assez rare. Cette étude pourra stimuler des recherches dans cette direction sur d'autres systèmes, qui permettront de déterminer si nos résultats sont généraux, ou s'ils sont dus à des traits d'histoire de vie spécifiques, et dans ce cas, lesquels.

Conclusion

Au cours de mon projet de thèse, j'ai étudié l'histoire de la domestication du pommier cultivé et l'impact de cette histoire sur l'évolution d'autres pommiers du genre *Malus*. L'histoire du pommier cultivé implique une domestication en Asie Centrale à partie de *M. sieversii* puis une importation vers l'Europe durant ou précédant l'époque classique qui se serait accompagnée d'introgressions avec différentes espèces sauvages. Il existait, au début de ma thèse, un débat sur l'importance des introgressions chez le pommier cultivé par l'espèce sauvage de pommiers endémique en Europe, *M. sylvestris*. Grâce à l'analyse de génomes reséquencés des trois espèces de pommiers impliqués dans l'histoire de la domestication, j'ai montré que des introgressions de *M. sylvestris* ont effectivement contribué au génome de certaines variétés actuelles, mais que certaines variétés semblent, à l'inverse, toujours très proches de l'espèce progénitrice, *M. sieversii*. Les variétés introgressées, elles, forment à présent un groupe génétique distinct de l'espèce progénitrice. Ce sont principalement des variétés européennes (à l'exclusion du sud de l'Europe) et elles sont elles-mêmes divisées en deux groupes génétiques à l'est et à l'ouest de l'Europe. Un approfondissement de cette étude permettra peut-être d'éclaircir le scénario évolutif ayant conduit à l'existence de ces groupes génétiques de variétés cultivés et notamment de préciser le rôle de *M. sylvestris*.

L'importation du pommier cultivé en Europe et sa culture répandue au sein de l'aire de répartition de *M. sylvestris* n'ont pas seulement eu des conséquences au niveau génomique pour *M. domestica*, mais elles ont également impacté l'évolution de *M. sylvestris*. En effet, deux études successives auxquelles j'ai participé ont permis de montrer que l'espèce sauvage, naturellement présente en faible densité, est actuellement introgressée par l'espèce cultivée, même dans son environnement naturel forestier. L'intensité des hybridations est liée aux activités anthropiques telles que l'abondance des vergers. Cette observation associée à l'absence de réduction de valeur sélective chez les hybrides nous a conduits à nous intéresser à la conservation de cette espèce

européenne. En nous penchant sur les filières de vente de *M. sylvestris*, nous avons montré que le système actuel ne permet pas de replanter avec certitude des individus appartenant à l'espèce sauvage et que, même lorsque c'est le cas, les arbres replantés ont une diversité génétique très faible. Dans le but d'améliorer la conservation du pommier sauvage en France, nous avons participé à la mise en place d'une règlementation de cette espèce forestière au niveau national. Nous avons, pour cela, utilisé notre étude phylogéographique sur toute aire de répartition, ce qui nous a permis de définir deux unités de conservation différentes à l'échelle de la France. L'avenir de ce projet inclut la mise en place de nouveaux vergers à graines respectant la nouvelle réglementation grâce aux arbres que nous avons identifiés comme n'étant pas introgressés à travers le pays.

Après avoir étudié les conséquences de l'introduction du pommier cultivé en Europe, je me suis également intéressée à la ré-introduction du pommier cultivé en Asie Centrale où a eu lieu sa domestication. Ainsi, les montagnes du Kazakhstan abritent encore des forêts incluant diverses espèces d'arbres fruitiers dont le pommier sauvage asiatique, *M. sieversii*. Au cours des deux derniers siècles, la culture du pommier s'est répandue dans cette zone géographique traditionnellement habitée par des peuples nomades. Le contact secondaire entre les deux espèces n'a pas encore eu, dans les populations naturelles étudiées dans cette thèse, l'effet qu'il a en Europe puisque nous avons identifié des populations sauvages encore intactes. Nous avons néanmoins observé l'existence de d'hybrides dans certaines populations, notamment celle en contact direct avec un verger de pommiers cultivés, indiquant que le même processus d'introgression que celui ayant eu lieu en Europe est peut-être en marche en Asie Centrale.

Lorsque le pommier cultivé a été ré-introduit dans sa région d'origine, il semblerait qu'il ait apporté avec lui une souche de tavelure qui lui était propre, qu'on a appelée ici *V. inaequalis* de type agricole, créant alors un contact secondaire entre cette souche et la souche naturelle parasitant les pommiers sauvages asiatiques. L'issue du contact secondaire chez le pathogène est très différente de celle observée chez son hôte : nous avons identifié la présence du pathogène de type agricole (ainsi que du type sauvage) dans toutes les populations forestières échantillonnées, y compris celles qui apparaissaient exemptes d'arbres *M. domestica* et hybrides. Le taux d'hybrides entre les deux types de pathogènes est comparativement bas, surtout en regard d'autres cas de contacts secondaires observés chez ce pathogène en Europe, entre d'autres populations. Nous avons donc cherché à identifier les barrières à la reproduction érigées entre ces deux types de pathogènes. Notre étude montrent que ces barrières ne sont pas dues à un isolement géographique où écologique puisque, bien que seules les souches de type agricoles soient capables d'infecter les hôtes hybrides ou *M. domestica*, les pommiers sauvages abritent

les deux types de pathogènes (et ce jusqu'au niveau de l'individu, avec certains hôtes infectés à la fois par les deux types). Il est donc possible que les barrières reproductives, dans ce cas, soient plutôt post-zygotiques et cela pourrait être une piste de recherche intéressante pour l'avenir.

Au cours de ma thèse, j'ai également étudié un autre type de système hôte-pathogène, *M. lychnidis-dioicae* causant le charbon des anthères chez le compagnon blanc (*S. latifolia*). Je me suis intéressée à la structure génétique de l'hôte et de son pathogène à l'échelle de l'Europe, utilisant des techniques et concepts identiques à ceux m'ayant permis d'étudier la structure génétique chez le pommier sauvage européen. La comparaison des structures phylogéographiques entre *M. lychnidis-dioicae* et *S. latifolia* s'est révélée extrêmement congruente, révélant l'existence de trois grands groupes génétiques correspondant aux refuges glaciaires situés dans les péninsules ibériques, italienne et balkanique. Une adéquation très fine entre les structures a pu être montrée puisque nos jeux de données incluent les génotypes pour les pathogènes et pour les plantes sur lesquelles ils ont été collectés, nous permettant donc de comparer les assignations aux groupes génétiques au niveau de l'individu. Une structure génétique plus fragmentée a néanmoins pu être montrée chez *M. lychnidis-dioicae* reflétant peut-être une histoire évolutive plus complexe lors de la dernière ère glaciaire et de la recolonisation ultérieure de l'Europe dont les traces auraient été effacées chez la plante. Il est ainsi possible qu'il ait existé des populations multiples de ces espèces dans les grandes zones de refuges glaciaires ou bien des refuges cryptiques situés plus au nord.

Plus largement, le travail présenté dans cette thèse a permis d'apporter un éclairage sur l'impact des activités anthropiques sur les trajectoires évolutives des espèces. Les actions humaines influencent le devenir de très nombreuses (si ce n'est de toutes les) espèces comme le montre l'exemple du compagnon blanc et de son pathogène *M. lychnidis-dioicae* qui, bien qu'ils n'aient aucune implication économique, ont tous les deux envahi l'Amérique du Nord suite à une introduction involontaire par l'Homme. Mais, dans cette thèse, les conséquences de l'implication humaine sont les plus visibles chez les espèces de pommiers et le pathogène responsable de la tavelure. Notre civilisation est une civilisation agricole et le parcours évolutif de certaines espèces comme le pommier cultivé a été transformé par l'action de l'Homme qui a cherché à obtenir des cultures plus productives, mondialisées et sans traces de pathogènes. Mais, là encore, les espèces impactées ne sont pas seulement celles que nous cherchons volontairement à contrôler, puisque les actions humaines peuvent également infléchir le devenir des apparentés sauvages de ces cultures ou des espèces avec lesquelles elles co-évoluent.

Bibliographie

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. http://doi.org/10.1111/j.1420-9101.2012.02599.x

Antonovics, J., Hood, M. E., Thrall, P. H., Abrams, J. Y., & Duthie, G. M. (2003). Herbarium studies on the distribution of anther-smut fungus (Microbotryum violaceum) and Silene species (Caryophyllaceae) in the eastern united states. *American Journal of Botany*, *90*(10), 1522–1531. http://doi.org/10.3732/ajb.90.10.1522

Arrigo, N., Guadagnuolo, R., Lappe, S., Pasche, S., Parisod, C., & Felber, F. (2011). Gene flow between wheat and wild relatives: empirical evidence from Aegilops geniculata, Ae. neglecta and Ae. triuncialis. *Evolutionary Applications*, *4*(5), 685–695. Journal Article. http://doi.org/10.1111/j.1752-4571.2011.00191.x

Arriola, P. E., & Ellstrand, N. C. (1997). Fitness of interspecific hybrids in the genus sorghum: Persistence of crop genes in wild populations. *Ecological Applications*, *7*(2), 512–518. http://doi.org/10.1890/1051-0761(1997)007[0512:FOIHIT]2.0.CO;2

Bacilieri, R., Lacombe, T., Le Cunff, L., Di Vecchi-Staraz, M., Laucou, V., Genna, B., … Boursiquot, J.-M. (2013). Genetic structure in cultivated grapevines is linked to geography and human selection. *BMC Plant Biology*, *13*(1), 25. http://doi.org/10.1186/1471-2229-13-25

Badouin, H. (2016). *Génomique évolutive chez les champignons Microbotryum : adaptation à l'hôte et chromosomes de types sexuels*. Université Paris-Saclay.

Baker, H. (1948). Stages in invasion and replacement demonstrated by species of Melandrium. *Journal of Ecology*, (36), 96–119.

Balsemin, E., & Collin, E. (2004). Conservation in situ des ressources génétiques des arbres forestiers en France métropolitaine. *Ingénieries*, *40*, 51–60. Journal Article.

Begon, M. (2009). Part II.9 Ecological Epidemiology. In S. A. Levin, S. R. Carpenter, H. C. J. Godfray, A. P. Kinzig, M. Loreau, J. B. Losos, … D. S. Wilcove (Eds.), *The Princeton Guide to Ecology*. PRINCETON UNIVERSITY PRESS.

Bendahmane, M., Dubois, A., Raymond, O., & Bris, M. Le. (2013). Genetics and genomics of flower initiation and development in roses. *Journal of Experimental Botany*, *64*(4), 847–857. http://doi.org/10.1093/jxb/ers387

Bennett, A. F., Radford, J. Q., & Haslem, A. (2006). Properties of land mosaics: Implications for nature conservation in agricultural environments. *Biological Conservation*, *133*(2), 250–264. http://doi.org/10.1016/j.biocon.2006.06.008

Bergstrom, C. T., & Dugatkin, L. A. (2012). *Evolution*. W. W. Norton & Company.

Bomblies, K., & Weigel, D. (2007). Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nature Reviews Genetics*, *8*(5), 382–393. http://doi.org/10.1038/nrg2082

Bowen, J. K., Mesarich, C. H., Bus, V. G., Beresford, R. M., Plummer, K. M., & Templeton, M. D. (2011). Venturia inaequalis: the causal agent of apple scab. *Mol Plant Pathol*, *12*(2), 105–122. Journal Article. http://doi.org/10.1111/j.1364-3703.2010.00656.x

Bowen, J. K., Mesarich, C. H., Bus, V. G. M., Beresford, R. M., Plummer, K. M., & Templeton, M. D. (2011). Venturia inaequalis: The causal agent of apple scab. *Molecular Plant Pathology*. http://doi.org/10.1111/j.1364-3703.2010.00656.x

Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E., & Leung, H. (2013). Plant-pathogen interactions: Disease resistance in modern agriculture. *Trends in Genetics*, *29*(4), 233–240. http://doi.org/10.1016/j.tig.2012.10.011

Brideau, N. J., Flores, H. A., Wang, J., Maheshwari, S., Wang, X., & Barbash, D. A. (2006). Lethality in Drosophila, *314*(NOVEMBER), 1292–1295.

Brooks, C. P., Antonovics, J., & Keitt, T. H. (2008). Spatial and Temporal Heterogeneity Explain Disease Dynamics in a Spatially Explicit Network Model. *The American Naturalist*, *172*(2), 149–159. http://doi.org/10.1086/589451

Brown, J. K., & Tellier, A. (2011). Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual Review of Phytopathology*, *49*, 345–367. http://doi.org/10.1146/annurev-phyto-072910-095301

Cai, L., Giraud, T., Zhang, N., Begerow, D., Cai, G., & Shivas, R. G. (2011). The evolution of species concepts and species recognition criteria in plant pathogenic fungi. *Fungal Diversity*, *50*, 121–133. http://doi.org/10.1007/s13225-011-0127-8

Charmet, G. (2011). Wheat domestication: Lessons for the future. *Comptes Rendus - Biologies*, *334*(3), 212–220. http://doi.org/10.1016/j.crvi.2010.12.013

Cheddadi, R., Vendramin, G. G., Litt, T., François, L., Kageyama, M., Lorentz, S., … Lunt, D. (2006). Imprints of glacial refugia in the modern genetic diversity of Pinus sylvestris. *Global Ecology and Biogeography*, *15*(3), 271–282. http://doi.org/10.1111/j.1466-822X.2006.00226.x

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*(8), 2171–2182. Journal Article. http://doi.org/10.1111/j.1365-294X.2006.02924.x

Coart, E., Vekemans, X., Smulders, M. J. M., Wagner, I., Van Huylenbroeck, J., Van Bockstaele, E., & Roldan-Ruiz, I. (2003). Genetic variation in the endangered wild apple (Malus sylvestris (L.) Mill.) in Belgium as revealed by amplified fragment length polymorphism and microsatellite markers. *Molecular Ecology*, *12*(4), 845–857. Journal Article. http://doi.org/10.1046/j.1365-294X.2003.01778.x

Cornille, A., Feurtey, A., Gélin, U., Ropars, J., Misvanderbrugge, K., Gladieux, P., & Giraud, T. (2015). Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: A basis for conservation and breeding programs in apples. *Evolutionary Applications*, *8*(4). http://doi.org/10.1111/eva.12250

Cornille, A., Giraud, T., Bellard, C., Tellier, A., Le Cam, B., Smulders, M. J. M., … Gladieux, P. (2013a). Postglacial recolonization history of the European crabapple (Malus sylvestris Mill.), a wild contributor to the domesticated apple. *Molecular Ecology*, *22*(8), 2249–2263. Journal Article. http://doi.org/10.1111/mec.12231

Cornille, A., Giraud, T., Smulders, M. J. M., Roldán-Ruiz, I., & Gladieux, P. (2014). The domestication and evolutionary ecology of apples. *Trends in Genetics*, *30*(2), 57–65. Journal Article. http://doi.org/10.1016/j.tig.2013.10.002

Cornille, A., Gladieux, P., & Giraud, T. (2013b). Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *Evolutionary Applications*, *6*(5), 737–748. Journal Article. http://doi.org/10.1111/eva.12059

Cornille, A., Gladieux, P., Smulders, M. J. M., Roldan-Ruiz, I., Laurens, F., Le Cam, B., … 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 Genetics*, *8*(5). Journal Article. http://doi.org/10.1371/journal.pgen.1002703

Couch, B. C. (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*(2), 613–630. http://doi.org/10.1534/genetics.105.041780

De Gracia, M., Cascales, M., Expert, P., Bellanger, M. N., Le Cam, B., & Lemaire, C. (2015). How Did Host Domestication Modify Life History Traits of Its Pathogens? *Plos One*, *10*(6), e0122909. Journal Article. http://doi.org/10.1371/journal.pone.0122909

de Guia, A. P. O., & Saitoh, T. (2007). The gap between the concept and definitions in the Evolutionarily Significant Unit: the need to integrate neutral genetic variation and adaptive variation. *Ecological Research*, *22*(4), 604–612. http://doi.org/10.1007/s11284-006-0059-z

de Lafontaine, G., Amasifuen Guerra, C. A., Ducousso, A., & Petit, R. J. (2014). Cryptic no more: soil macrofossils uncover Pleistocene forest microrefugia within a periglacial desert. *New Phytologist*, *204*(3), 715–729. http://doi.org/10.1111/nph.12833

de Lafontaine, G., Amasifuen Guerra, C. A., Ducousso, A., Sánchez-Goñi, M.-F., & Petit, R. J. (2014). Beyond skepticism: uncovering cryptic refugia using multiple lines of evidence. *New Phytologist*, *204*(3), 450–454. http://doi.org/10.1111/nph.13089

De Vienne, D. M., Refrégier, G., López-Villavicencio, M., Tellier, A., Hood, M. E., & Giraud, T. (2013). Cospeciation vs host-shift speciation: Methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist*, *198*(2), 347–385. http://doi.org/10.1111/nph.12150

Decroocq, S., Cornille, A., Tricon, D., Babayeva, S., Chague, A., Eyquard, J.-P., … Decroocq, V. (2016). New insights into the history of domesticated and wild apricot and its contribution to *Plum pox virus* resistance. *Molecular Ecology*, 1–18. http://doi.org/10.1111/mec.13772

Deguines, N., Julliard, R., de Flores, M., & Fontaine, C. (2012). The Whereabouts of Flower Visitors: Contrasting Land-Use Preferences Revealed by a Country-Wide Survey Based on Citizen Science. *PLoS ONE*, *7*(9), e45822. http://doi.org/10.1371/journal.pone.0045822

Dybdahl, M. F., & Storfer, A. (2003). Parasite local adaptation: Red Queen versus Suicide King. *Trends in Ecology and Evolution*, *18*(10), 523–530. http://doi.org/10.1016/S0169-5347(03)00223-4

Ellstrand, N. C. (2014). Is gene flow the most important evolutionary force in plants? *American Journal of Botany*, *101*(5), 737–753. http://doi.org/10.3732/ajb.1400024

Ellstrand, N. C., Meirmans, P., Rong, J., Bartsch, D., Ghosh, A., de Jong, T. J., … Hooftman, D. (2013). Introgression of Crop Alleles into Wild or Weedy Populations. In D. J. Futuyma (Ed.), *Annual Review of Ecology, Evolution, and Systematics, Vol 44* (Vol. 44, pp. 325–345). Book Section. http://doi.org/10.1146/annurev-ecolsys-110512-135840

Ellstrand, N. C., Prentice, H. C., & James, F. (1999). Gene Flow and Introgression from Domesticated Plants into their Wild Relatives. *Annual Review of Ecology and Systematics*, *30*, 539–563.

Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, *23*(7), 347–351. http://doi.org/10.1016/j.tree.2008.04.004

Feulner, P. G. D., Gratten, J., Kijas, J. W., Visscher, P. M., Pemberton, J. M., & Slate, J. (2013). Introgression and the fate of domesticated genes in a wild mammal population. *Molecular Ecology*, *22*(16), 4210–4221. http://doi.org/10.1111/mec.12378

Fishman, L., & Willis, J. H. (2001). Evidence for Dobzhansky-Muller Incompatibilites Contributing To the Sterility of Hybrids Between Mimulus Guttatus and M. Nasutus. *Evolution*, *55*(10), 1932–1942. http://doi.org/10.1111/j.0014-3820.2001.tb01311.x

Flor, H. H. (1971). Current Status of the Gene-For-Gene Concept. *Annual Review of Phytopathology*, *9*(1), 275–296. http://doi.org/10.1146/annurev.py.09.090171.001423

Fontaine, M. C., Gladieux, P., Hood, M. E., & Giraud, T. (2013). History of the invasion of the anther smut pathogen on Silene latifolia in North America. *New Phytologist*, *198*(3), 946–956. http://doi.org/10.1111/nph.12177

Friesen, T. L., Stukenbrock, E. H., Liu, Z., Meinhardt, S., Ling, H., Faris, J. D., … Oliver, R. P. (2006). Emergence of a new disease as a result of interspecific virulence gene transfer. *Nature Genetics*, *38*(8), 953–956. http://doi.org/10.1038/ng1839

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L., & Nielsen, R. (2011). Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genetics*, *7*(11). http://doi.org/10.1371/journal.pgen.1002355

Giraud, T., Refrégier, G., Le Gac, M., de Vienne, D. M., & Hood, M. E. (2008). Speciation in fungi. *Fungal Genetics and Biology*, *45*(6), 791–802. http://doi.org/10.1016/j.fgb.2008.02.001

Gladieux, P., Feurtey, A., Hood, M. E., Snirc, A., Clavel, J., Dutech, C., … Giraud, T. (2015). The population biology of fungal invasions. *Molecular Ecology*, *24*(9). http://doi.org/10.1111/mec.13028

Gladieux, P., Vercken, E., Fontaine, M. C., Hood, M. E., Jonot, O., Couloux, A., & Giraud, T. (2011). Maintenance of Fungal Pathogen Species That Are Specialized to Different Hosts: Allopatric Divergence and Introgression through Secondary Contact. *Molecular Biology and Evolution*, *28*(1), 459–471. http://doi.org/10.1093/molbev/msq235

Gladieux, P., Zhang, X. G., Roldan-Ruiz, I., Caffier, V., Leroy, T., Devaux, M., … Le Cam, B. (2010). Evolution of the population structure of Venturia inaequalis, the apple scab fungus, associated with the domestication of its host. *Mol Ecol*, *19*(4), 658–674. Journal Article. http://doi.org/10.1111/j.1365-294X.2009.04498.x

Gómez, A., & Lunt, D. H. (2007). Refugia within Refugia: Patterns of Phylogeographic Concordance in the Iberian Peninsula. In *Phylogeography of Southern European Refugia* (pp. 155–188). Dordrecht: Springer Netherlands. http://doi.org/10.1007/1-4020-4904-8_5

Greischar, M. A., & Koskella, B. (2007). A synthesis of experimental work on parasite local adaptation. *Ecology Letters*, *10*(5), 418–434. http://doi.org/10.1111/j.1461-0248.2007.01028.x

Gururani, M. A., Venkatesh, J., Upadhyaya, C. P., Nookaraju, A., Pandey, S. K., & Park, S. W. (2012). Plant disease resistance genes: Current status and future directions. *Physiological and Molecular Plant Pathology*, *78*(January 2012), 51–65. http://doi.org/10.1016/j.pmpp.2012.01.002

Gutierrez, A., Cantamutto, M., & Poverene, M. (2011). Persistence of sunflower crop traits and fitness in Helianthus petiolaris populations. *Plant Biology*, *13*(5), 821–830. Journal Article. http://doi.org/10.1111/j.1438-8677.2010.00433.x

Hajjar, R., & Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica*, *156*(1–2), 1–13. http://doi.org/10.1007/s10681-007-9363-0

Hancock, J. F., Luby, J. J., Brown, S. K., & Lobos, G. A. (2008). Apples. *Temperate Fruit Crop Breeding*, 1–37.

Harris, S. A., Robinson, J. P., & Juniper, B. E. (2002). Genetic clues to the origin of the apple. *Trends in Genetics*, *18*(8), 426–430. Journal Article. http://doi.org/10.1016/s0168-9525(02)02689-6

Harris, S. A., Robinson, J. P., & Juniper, B. E. (2002). Genetic clues to the origin of the apple. *Trends in Genetics*, *18*(8), 426–430. http://doi.org/10.1016/S0168-9525(02)02689-6

Harrison, N., & Harrison, R. J. (2011). On the evolutionary history of the domesticated apple. *Nature Genetics*, *43*(11), 1043–1044. http://doi.org/10.1038/ng.935

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, *105*(S1), 795–809. http://doi.org/10.1093/jhered/esu033

Hewitt, G. (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, *68*(1–2), 87–112. http://doi.org/10.1111/j.1095-8312.1999.tb01160.x

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1442), 183–195; discussion 195. http://doi.org/10.1098/rstb.2003.1388

Hood, M. E., Mena-Alí, J. I., Gibson, A. K., Oxelman, B., Giraud, T., Yockteng, R., … Antonovics, J. (2010). Distribution of the anther-smut pathogen Microbotryum on species of the Caryophyllaceae. *New Phytologist*, *187*(1), 217–229. http://doi.org/10.1111/j.1469-8137.2010.03268.x

Hooftman, D. A. P., Oostermeijer, J. G. B., Jacobs, M. M. J., & Den Nijs, H. C. M. (2005). Demographic vital rates determine the performance advantage of crop-wild hybrids in lettuce. *Journal of Applied Ecology*, *42*(6), 1086–1095. Journal Article. http://doi.org/10.1111/j.1365-2664.2005.01086.x

Hovick, S. M., Campbell, L. G., Snow, A. A., & Whitney, K. D. (2012). Hybridization Alters Early Life-History Traits and Increases Plant Colonization Success in a Novel Region. *American Naturalist*, *179*(2), 192–203. Journal Article. http://doi.org/10.1086/663684

Hufford, M. B., Lubinksy, P., Pyhaejaervi, T., Devengenzo, M. T., Ellstrand, N. C., & Ross-Ibarra, J. (2013). The Genomic Signature of Crop-Wild Introgression in Maize. *Plos Genetics*, *9*(5). Journal Article. http://doi.org/10.1371/journal.pgen.1003477

Jackson, J. E. (2003). *Biology of apples and pears*. Cambridge University Press.

Janick, J. (2005). The Origin of Fruits, Fruit Frowing and Fruit Breeding. *Plant Breeding Reviews*, *25*, 255–320.

Johnson, N. (2008). Hybrid incompatibility and speciation. *Nature Education*, *1(1):20*.

Juniper, B. E., & Mabberley, D. J. (2006). *The Story of the Apple*. Timber Press.

Kao, K. C., Schwartz, K., & Sherlock, G. (2010). A genome-wide analysis reveals no nuclear Dobzhansky-Muller pairs of determinants of speciation between S. cerevisiae and S. paradoxus, but suggests more complex incompatibilities. *PLoS Genetics*, *6*(7), 1–12. http://doi.org/10.1371/journal.pgen.1001038

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241. http://doi.org/10.1111/j.1461-0248.2004.00684.x

Keller, S. R., Gilbert, K. J., Fields, P. D., & Taylor, D. R. (2012). Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, Silene latifolia. *Molecular Ecology*, *21*(19), 4721–4734. http://doi.org/10.1111/j.1365-294X.2012.05751.x

Koskela, J., Lefevre, F., Schueler, S., Kraigher, H., Olrik, D. C., Hubert, J., … Ditlevsen, B. (2013). Translating conservation genetics into management: Pan-European minimum requirements

for dynamic conservation units of forest tree genetic diversity. *Biological Conservation*, *157*, 39–49. Journal Article. http://doi.org/10.1016/j.biocon.2012.07.023

Kremen, C., Williams, N. M., Aizen, M. A., Gemmill-Herren, B., LeBuhn, G., Minckley, R., … Ricketts, T. H. (2007). Pollination and other ecosystem services produced by mobile organisms: a conceptual framework for the effects of land-use change. *Ecology Letters*, *10*(4), 299–314. http://doi.org/10.1111/j.1461-0248.2007.01018.x

Larsen, A. S., & Kjaer, E. D. (2009). Pollen mediated gene flow in a native population of Malus sylvestris and its implications for contemporary gene conservation management. *Conservation Genetics*, *10*(6), 1637–1646. Journal Article. http://doi.org/10.1007/s10592-008-9713-z

Le Van, A., Gladieux, P., Lemaire, C., Cornille, A., Giraud, T., Durel, C.-E., … Le Cam, B. (2012). Evolution of pathogenicity traits in the apple scab fungal pathogen in response to the domestication of its host. *Evolutionary Applications*, *5*(7), 694–704. Journal Article. http://doi.org/10.1111/j.1752-4571.2012.00246.x

Lemaire, C., De Gracia, M., Leroy, T., Michalecka, M., Lindhard-Pedersen, H., Guerin, F., … Le Cam, B. (2016). Emergence of new virulent populations of apple scab from nonagricultural disease reservoirs. *New Phytol*, *209*(3), 1220–1229. Journal Article. http://doi.org/10.1111/nph.13658

Leroy, T., Caffier, V., Celton, J.-M., Anger, N., Durel, C.-E., Lemaire, C., & Le Cam, B. (2016). When virulence originates from nonagricultural hosts: evolutionary and epidemiological consequences of introgressions following secondary contacts in Venturia inaequalis. *New Phytologist*, *210*(4), 1443–1452. http://doi.org/10.1111/nph.13873

López-Villavicencio, M., Courjol, F., Gibson, A. K., Hood, M. E., Jonot, O., Shykoff, J. A., & Giraud, T. (2011). Competition, cooperation among kin, and virulence in multiple infections. *Evolution*, *65*(5), 1357–1366. http://doi.org/10.1111/j.1558-5646.2010.01207.x

López-Villavicencio, M., Jonot, O., Coantic, A., Hood, M. E., Enjalbert, J., & Giraud, T. (2007). Multiple infections by the anther smut pathogen are frequent and involve related strains. *PLoS Pathogens*, *3*(11), 1710–1715. http://doi.org/10.1371/journal.ppat.0030176

MacHardy, W. E., Gadoury, D. M., & Gessler, C. (2001). Parasitic and Biological Fitness of Venturia inaequalis : Relationship to Disease Management Strategies. *Plant Disease*, *85*(10), 1036–1051. http://doi.org/10.1094/PDIS.2001.85.10.1036

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, *20*(5), 229–237. http://doi.org/10.1016/j.tree.2005.02.010

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*(7133), 279–283. http://doi.org/10.1038/nature05706

McDonald, B. A., & Linde, C. (2002). Pathogen Population Genetics, Evolutionary potential, and Durable Resistance. *Annual Review of Phytopathology*, *40*(1), 349–379. http://doi.org/10.1146/annurev.phyto.40.120501.101443

Menardo, F., Praz, C., Wyder, S., Bourras, S. . A., McNally, K. E., Parlange, F., … Keller, B. (2015). Hybridization of powdery mildew strains gives raise to pathogens on novel agricultural crop species. *Nature Genetics*, *48*(January), 1–24. http://doi.org/10.1038/ng.3485

Micheletti, D., Troggio, M., Salamini, F., Viola, R., Velasco, R., & Salvi, S. (2011). On the evolutionary history of the domesticated apple. *Nature Genetics*, *43*(11), 1044–1045. http://doi.org/10.1038/ng.983

Miller, A. J., & Gross, B. L. (2011). From forest to field: Perennial fruit crop domestication. *American Journal of Botany*, *98*(9), 1389–1414. http://doi.org/10.3732/ajb.1000522

Naomi F. Miller. (2013). Symbols of Fertility and Abundance in the Royal Cemetery at Ur, Iraq. *American Journal of Archaeology*, *117*(1), 127. http://doi.org/10.3764/aja.117.1.0127

Newcombe, G., Stirling, B., McDonald, S., & Bradshaw, H. D. (2000). Melampsora × columbiana, a natural hybrid of M. medusae and M. occidentalis. *Mycological Research*, *104*(3), 261–274. http://doi.org/10.1017/S0953756299001665

Newton, A. C., Allnutt, T. R., Gillies, A. C. M., Lowe, A. J., & Ennos, R. A. (1999). Molecular phylogeography, intraspecific variation and the conservation of tree species. *Trends in Ecology and Evolution*, *14*(4), 140–145. http://doi.org/10.1016/S0169-5347(98)01555-9

Nieberding, C. M., & Olivieri, I. (2007). Parasites: proxies for host genealogy and ecology? *Trends in Ecology and Evolution*, *22*(3), 156–165. http://doi.org/10.1016/j.tree.2006.11.012

Nikiforova, S. V., Cavalieri, D., Velasco, R., & Goremykin, V. (2013). Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. *Molecular Biology and Evolution*, *30*(8), 1751–1760. http://doi.org/10.1093/molbev/mst092

Petit, R. J., Csaikl, U. M., Bordács, S., Burg, K., Coart, E., Cottrell, J., … Kremer, A. (2002). Chloroplast DNA variation in European white oaks. *Forest Ecology and Management*, *156*(1–3), 5–26. http://doi.org/10.1016/S0378-1127(01)00645-4

Petit, R. J., & Hampe, A. (2006). Some evolutionary consequences of being a tree. In *Annual Review of Ecology Evolution and Systematics* (Vol. 37, pp. 187–214). Book Section. http://doi.org/10.1146/annurev.ecolsys.37.091305.110215

Provan, J., & Bennett, K. D. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, *23*(10), 564–571. http://doi.org/10.1016/j.tree.2008.06.010

Pusadee, T., Schaal, B. A., Rerkasem, B., & Jamjod, S. (2013). Population structure of the primary gene pool of Oryza sativa in Thailand. *Genetic Resources and Crop Evolution*, *60*(1), 335–353. Journal Article. http://doi.org/10.1007/s10722-012-9839-1

Queiroz, K. De. (2007). Species concepts and species delimitation. *Systematic Botany*, *56*(6), 879–886. http://doi.org/10.1080/10635150701701083

Quilot, B., Wu, B. H., Kervella, J., Génard, M., Foulongne, M., & Moreau, K. (2004). QTL analysis of quality traits in an advanced backcross between Prunus persica cultivars and the wild relative species P. davidiana. *Theoretical and Applied Genetics*, *109*(4), 884–897. http://doi.org/10.1007/s00122-004-1703-z

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, *197*(2), 573–589. http://doi.org/10.1534/genetics.114.164350

Refrégier, G., Le Gac, M., Jabbour, F., Widmer, A., Shykoff, J. A., Yockteng, R., … Giraud, T. (2008). Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. *BMC Evolutionary Biology*, *8*(1), 100. http://doi.org/10.1186/1471-2148-8-100

Reim, S., Proft, A., Heinz, S., Lochschmidt, F., Höfer, M., Tröber, U., & Wold, H. (2015). Pollen movement in a Malus sylvestris population and conclusions for conservation measures. *Plant Genetic Resources*, *FirstView*, 9. Journal Article. http://doi.org/10.1017/S1479262115000301

Saleh, D., Milazzo, J., Adreit, H., Fournier, E., & Tharreau, D. (2014). South-East Asia is the center of origin, diversity and dispersion of the rice blast fungus, Magnaporthe oryzae. *New Phytologist*, *201*(4), 1440–1456. http://doi.org/10.1111/nph.12627

Sansavini, S., Donati, F., Costa, F., & Tartarini, S. (2004). Advances in apple breeding for enhanced fruit quality and resistance to biotic stresses: new varieties for the european market. *Journal of Fruit and Ornamental Plant Research*, *12*(Special ed.), 13–52.

Sasaki, A. (2000). Host-parasite coevolution in a multilocus gene-for-gene system. *Proceedings. Biological Sciences / The Royal Society*, *267*(1458), 2183–8. http://doi.org/10.1098/rspb.2000.1267

Schnitzler, A., Arnold, C., Cornille, A., Bachmann, O., & Schnitzler, C. (2014). Wild European apple (Malus sylvestris (L.) Mill.) Population Dynamics: Insight from Genetics and Ecology

in the Rhine Valley. Priorities for a Future Conservation Programme. *Plos One*, *9*(5). Journal Article. http://doi.org/10.1371/journal.pone.0096596

Snow, A. A. (2012). Illegal gene flow from transgenic creeping bentgrass: the saga continues. *Molecular Ecology*, *21*(19), 4663–4664. http://doi.org/10.1111/j.1365-294X.2012.05695.x

Soulé, M. E. (1985). What is conservation biology? *BioScience*. http://doi.org/10.2307/1310054

Stephan, B. R., Wagner, I., & Kleinschmit, J. (2003). *Technical guidelines for genetic conservation and use for wild apple and pear (Malus sylvestris and Pyrus pyraster)*. International Plant Genetic Resources Insitute, Rome, Italy.

Stewart, J. R., Lister, A. M., Barnes, I., & Dalén, L. (2010). Refugia revisited: individualistic responses of species in space and time. *Proceedings. Biological Sciences / The Royal Society*, *277*(1682), 661–71. http://doi.org/10.1098/rspb.2009.1272

Stukenbrock, E. H., & McDonald, B. A. (2008). The Origins of Plant Pathogens in Agro-Ecosystems. *Annual Review of Phytopathology*, *46*(1), 75–100. http://doi.org/10.1146/annurev.phyto.010708.154114

Summers, K., McKeon, S., Sellars, J., Keusenkothen, M., Morris, J., Gloeckner, D., … Snow, H. (2003). Parasitic exploitation as an engine of diversity. *Biological Reviews of the Cambridge Philosophical Society*, *78*, 639–675. http://doi.org/10.1017/S146479310300616X

Taylor, D. R., & Keller, S. R. (2007). Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution*, *61*(2), 334–345. http://doi.org/10.1111/j.1558-5646.2007.00037.x

Todesco, M., Pascual, M. A., Owens, G. L., Ostevik, K. L., Moyers, B. T., Hübner, S., … Rieseberg, L. H. (2016). Hybridization and extinction. *Evolutionary Applications*. Journal Article. http://doi.org/10.1111/eva.12367

Urrestarazu, J., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., … Durel, C.-E. (2016). Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. *BMC Plant Biology*, *16*(1), 130. http://doi.org/10.1186/s12870-016-0818-0

van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory*. http://doi.org/10.1038/344864a0

Vavilov, N. I. (1926). Studies on the origin of cultivated plants. *Trudy Byuro. Prikl. Bot.*, (16), 139–245.

Vekemans, X., & Hardy, O. J. (2004). New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, *13*(4), 921–935. http://doi.org/10.1046/j.1365-294X.2004.02076.x

Velasco, R., & Licciardello, C. (2014). A genealogy of the citrus family. *Nature Biotechnology*, *32*(7), 640–2. http://doi.org/10.1038/nbt.2954

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., … Viola, R. (2010). The genome of the domesticated apple (Malus × domestica Borkh.). *Nat Genet*, *42*(10), 833–839. http://doi.org/10.1038/ng.654

Vercken, E., Fontaine, M. C., Gladieux, P., Hood, M. E., Jonot, O., & Giraud, T. (2010). Glacial refugia in pathogens: European genetic structure of anther smut pathogens on silene latifolia and silene dioica. *PLoS Pathogens*, *6*(12). http://doi.org/10.1371/journal.ppat.1001229

Williams, P. D. (2010). Darwinian interventions: taming pathogens through evolutionary ecology. *Trends in Parasitology*, *26*(2), 83–92. http://doi.org/10.1016/j.pt.2009.11.009

Wolf, D. E., Takebayashi, N., & Rieseberg, L. H. (2001). Predicting the risk of extinction through hybridization. *Conservation Biology*, *15*(4), 1039–1053. Journal Article. http://doi.org/10.1046/j.1523-1739.2001.0150041039.x

Wu, C.-I., & Ting, C.-T. (2004). Genes and speciation. *Nature Reviews. Genetics*, *5*(2), 114–122. http://doi.org/10.1038/nrg1269

Wu, G. A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., … Rokhsar, D. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology*, *32*(7), 656–662. http://doi.org/10.1038/nbt.2906

Zohary, D., & Hopf, M. (2000). *Domestication of Plants in the Old World* (3rd ed.). Oxford University Press.

Annexes

Annexe 1 : Article de vulgarisation « Domestication, une révolution à l'origine de la civilisation »

Durant les deux premières années de ma thèse, j'ai réalisé une mission doctorale au Palais de la Découverte. Ce musée et centre culturel a pour but de rendre la science accessible en présentant ses résultats et méthodes à un public varié incluant le grand public et le public scolaire (du primaire au lycée). Ceci est réalisé à travers plusieurs approches : la présentation d'expositions permanentes, la création d'expositions temporaires, la création d'exposés de vulgarisation présentant des concepts scientifiques et des expériences réalisées face au public et l'édition d'une revue de vulgarisation. Au cours de ma mission doctorale chez eux, j'ai présenté un exposé intitulé la Loterie de l'Hérédité et un atelier destiné aux enfants sur les sens. J'ai également rédigé un article sur la domestication. En effet, le Palais de la Découverte publie une revue de vulgarisation scientifique bimestrielle. Ma proposition de sujet a été acceptée et l'article a été publié dans le numéro 402 (Janvier-février 2016).

Pour des raisons de copyright, la reproduction de l'article n'est pas autorisée, mais une copie de la revue Découverte contenant l'article sera mise à disposition des membres du jury.

Annexe 2 : Copie des conseils mis à disposition par le ministère de l'agriculture concernant *Malus sylvestris*

Malus sylvestris (L.) Mill.
Pommier sauvage
Wild apple

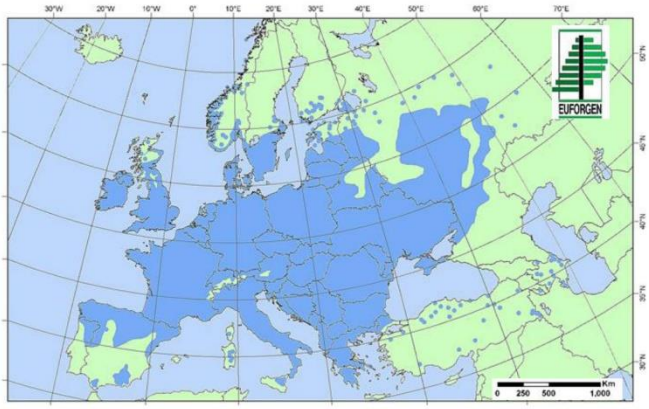
Caractéristiques générales de l'espèce

Aire naturelle

Le pommier sauvage est autochtone dans la plupart des pays européens, de la péninsule ibérique à l'Oural et des côtes norvégiennes aux Balkans. Il est présent de manière éparse sous forme d'individus isolés ou de petits groupes (Stephan *et al.*, 2003).

Suite aux glaciations du quaternaire, le pommier sauvage a recolonisé son aire actuelle à partir de zones refuges du sud de l'Europe. La diversité génétique actuelle de l'espèce, structurée en cinq grands groupes (Europe de l'Est et Balkans, nord de l'Europe, Italie, Europe Centrale dont l'est de la France et Europe de l'Ouest), reflète ces migrations (Cornille *et al.*, 2014).

L'hybridation du pommier sauvage par les variétés cultivées de pommier (*Malus domestica*) est fréquente en milieu naturel, rendant difficile l'identification des deux espèces (Cornille *et al.*, 2014).



Aire de distribution naturelle du pommier sauvage en Europe (EUFORGEN, 2008)

Observations de pommier sauvage en France (IGN, 2014)

Répartition du pommier sauvage en France

Le pommier sauvage est assez commun en France, sauf en région méditerranéenne. On le trouve de l'étage collinéen jusqu'à 1300 m d'altitude. Essence d'accompagnement, sa densité de population excède rarement un arbre/ha, sauf dans certaines forêts alluviales de la vallée du Rhin (Cornille *et al.*, 2014. Lamant *et al.*, 2005).

Version du 20/04/2016. Les informations et préconisations contenues dans cette fiche sont celles qu'il était possible de formuler à la date de rédaction, dans un contexte de forte incertitude sur les évolutions du climat et des aires de répartition des espèces. Il convient donc de s'assurer qu'aucune version plus récente n'a été publiée.
NB : les préconisations de cette fiche ne s'appliquent qu'aux reboisements et ne concernent pas la régénération naturelle.

Auteurs principaux : Anne Pierangelo, Eric Collin (Irstea)

1

Autécologie

Le pommier sauvage est une espèce à large amplitude écologique, qui supporte les climats froids, résiste aux gelées précoces comme tardives et présente une relative tolérance à la sécheresse (Gonin *et al.*, 2013). Héliophile et peu compétitif, on le rencontre particulièrement en milieu ouvert, dans les haies, les friches, en lisière de forêt et dans les parcelles forestières récemment exploitées (Cornille *et al.*, 2014). Il tolère cependant l'ombre, mais sa croissance y est très lente et ses fructifications quasi absentes (Lamant *et al.*, 2005).

Espèce mésophile à très large amplitude, le pommier sauvage pousse sur des sols variés, aussi bien argileux que limoneux, avec plus ou moins d'éléments grossiers. Il est plus rare sur station acidiphile, et craint l'engorgement. Les textures très sableuses lui sont également défavorables, et une forte compacité limite sa croissance (Gonin *et al.*, 2013). Son optimum écologique se trouve sur les sols frais, épais, riches et à réserve en eau importante, mais la concurrence avec les autres espèces le restreint aux stations plus contraignantes, à la limite humide de la forêt ou en conditions ripicoles telles que les prairies humides alluviales (Lamant *et al.*, 2005).

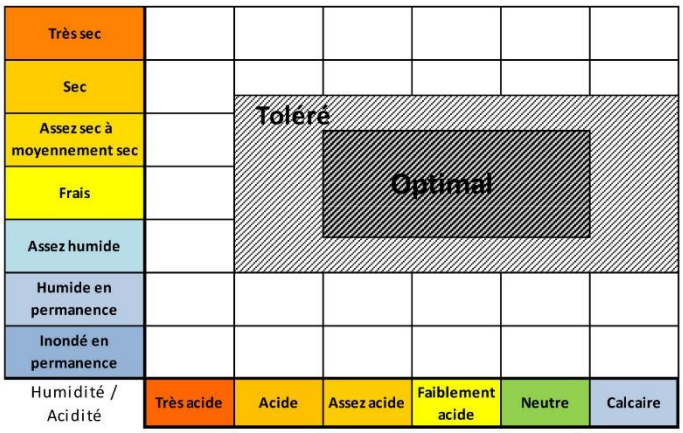


Diagramme de répartition de l'espèce selon les gradients trophiques et hydriques définis dans la Flore forestière française, tome 1. Rameau et al. 1989

Sensibilité aux maladies et ravageurs

Un peuplement forestier situé dans une station adaptée aux exigences de l'espèce présentera une moindre vulnérabilité à certains aléas sanitaires.

Concernant les maladies endémiques affectant les variétés cultivées de pommier, comme la tavelure (*Venturia inaequalis*) et l'oïdium (*Podosphaera leucotricha*), on peut supposer que la cohabitation ancienne de ces maladies avec le pommier sauvage a permis aux populations actuelles de s'adapter via la pression de sélection. En revanche, concernant les maladies d'introduction récente, comme le feu bactérien (*Erwinia amylovora*), et les insectes ravageurs des fruitiers de culture, on peut craindre une sensibilité générale du pommier sauvage avec des niveaux de résistance extrêmement variables d'un individu à l'autre. La dispersion des tiges et le mélange avec d'autres essences limitent cependant les risques d'attaques des maladies à dispersion aérienne (Lévêque *et al.*, 2005).

Effets supposés du changement climatique sur les boisements

Le choix d'une essence de reboisement doit être raisonné en fonction des contraintes climatiques qui apparaîtront successivement durant la vie du boisement. Malgré les incertitudes sur les modèles climatiques, il est nécessaire d'anticiper au mieux les effets directs et indirects des changements climatiques tels que la fréquence accrue et la durée plus longue des sécheresses ou l'augmentation des températures.

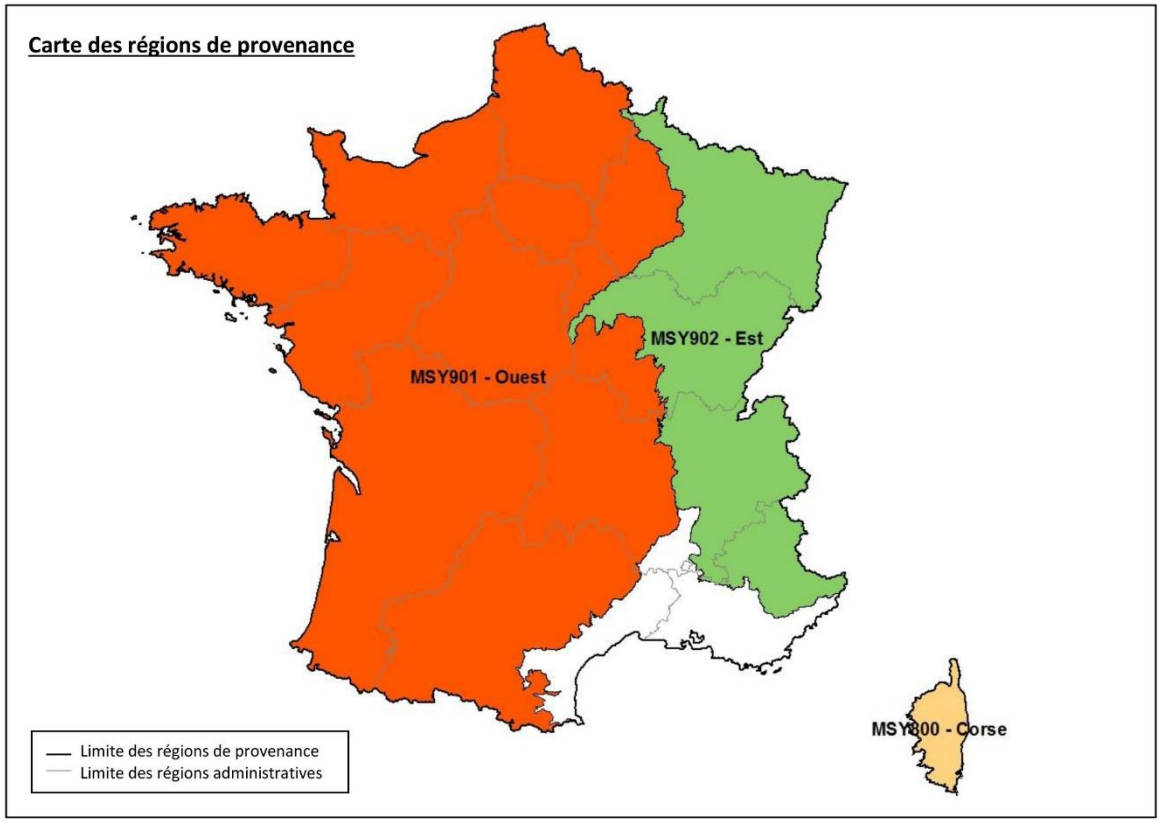
Le pommier sauvage présente une relative tolérance à la sécheresse, et peut s'installer dans des conditions écologiques très variées (Gonin *et al.*, 2013). De ce fait, et compte tenu du faible enjeu de production lié à cette espèce, les effets attendus du changement climatique devraient être peu marqués.

Description des matériels de base

Les matériels forestiers de reproduction (MFR) sont issus des matériels de base. Dans le cas du pommier sauvage, ces derniers sont des sources de graines de catégorie identifiée. Leur code d'identification peut indifféremment se rapporter au matériel commercialisable (MFR), au matériel de base dont il est issu, ou à sa région de provenance.

Des études récentes sur la structuration génétique du pommier sauvage montrent que trois lignées différentes sont présentes en France : deux lignées Ouest et Est issues d'un refuge post-glaciaire de la Péninsule ibérique, et une lignée Sud issue d'un refuge italien (Cornille, 2012). Cette dernière n'est cependant présente qu'à l'état d'allèles introduits dans le génome des individus de deux populations méditerranéennes de *Malus sylvestris*.

Deux régions de provenances sont donc créées : une provenance Ouest (MSY901 - Ouest) et une provenance Est (MSY902 - Est). À défaut d'informations, une troisième région de provenance est créée pour la Corse (MSY800 - Corse). La région méditerranéenne est considérée comme une zone sans récolte du fait d'une ressource en pommier sauvage faible à nulle.



L'hybridation du pommier sauvage par les variétés cultivées de pommier (*Malus domestica*) est fréquente en milieu naturel, rendant difficile l'identification des deux espèces (Cornille *et al.*, 2014). Ce phénomène est favorisé par un mécanisme d'auto-incompatibilité partielle du pommier sauvage, limitant les possibilités d'autofécondation chez les individus isolés. En vallée du Rhin, il semblerait cependant qu'un décalage phénologique entre le pommier sauvage et le pommier cultivé permette de limiter les flux de gènes entre les deux espèces (Schnitzler *et al.*, 2014).
Deux massifs forestiers riches en pommiers sauvages[1] pourront être utilisés prioritairement comme source de graines : la forêt domaniale de Dourdan (Essonne) pour la région de provenance MSY901 – Ouest, et les forêts communales de Mackenheim et Marckolsheim (Bas-Rhin) pour la région de provenance MSY902 – Est.

[1] La pureté spécifique des pommiers présumés sauvages de ces massifs a été caractérisée par l'Université d'Orsay. Pour plus d'informations, se renseigner auprès d'Irstea – GeeDAAF (anne.pierangelo@irstea.fr)

Afin de préserver les ressources génétiques de cette espèce disséminée, et en accord avec les engagements de la charte de qualité et de diversité génétiques des lots de graines et plants forestiers, les récolteurs s'assureront de proposer des lots de graines issus d'au moins 10 semenciers et éviteront les récoltes à proximité de vergers de pommiers de culture.

NB : un verger à graines de pommier sauvage, potentiellement adapté aux conditions pédoclimatiques du nord de la France, a été installé entre 2006 et 2008 en Belgique. Cependant, aucune récolte n'a pu y être réalisée à ce jour faute de fructification.

Conseils d'utilisation des MFR

Le tableau ci-dessous présente les conseils d'utilisation par grandes régions écologiques (GRECO) et sylvoécorégions (SER). Leur description complète est consultable sur ***http://inventaire-forestier.ign.fr/spip/spip.php?article686****.*
*Dans ce tableau, la colonne « **Matériels conseillés** » indique les MFR les plus appropriés dans les SER considérées. La colonne « **Autres matériels utilisables** » liste les MFR utilisables en cas de pénurie du matériel conseillé, et ceux utilisables en second choix, selon le diagnostic local de la station, qu'il s'agisse de MFR utilisés dans une région où la plantation de cette essence n'est globalement pas conseillée ou de MFR introduits à des fins de diversification génétique.*

Le pommier sauvage étant une espèce autochtone, et aucun effet particulier du changement climatique n'étant attendu sur les populations, les provenances locales sont recommandées.

Tableau des conseils d'utilisation du pommier sauvage

Zones d'utilisation				Matériels conseillés		Autres matériels utilisables	
GRECO		SER					
code	Nom	code	Nom	Nom	Cat.	Nom	Cat.
A	Grand Ouest cristallin et océanique	-	Toutes	MSY901	I		
B	Centre-Nord semi-océanique	-	Toutes				
C	Grand Est semi-continental	-	Toutes	MSY902	I		
D	Vosges	-	Toutes				
E	Jura	-	Toutes				
F	Sud-Ouest océanique	-	Toutes	MSY901	I		
G	Massif central	-	Toutes				
H	Alpes	-	Toutes	MSY902	I		
I	Pyrénées	-	Toutes	MSY901	I		
J	Méditerranée	-	Toutes	-			
K	Corse	-	Toutes	MSY800	I		

Carte des conseils d'utilisation pour des projets de plantation du pommier sauvage

Zones géographiques dans lesquelles :

- des MFR de pommier sauvage sont conseillés,
- aucun MFR de pommier sauvage n'est conseillé.

Attention, les conseils d'utilisation sont également soumis à l'autécologie du pommier sauvage, décrite en deuxième page.

Carte des conseils d'utilisation du pommier sauvage

Références

Cornille A., Giraud T., 2014. Pommier sauvage : l'impact du climat sur sa génétique. Dossier Le climat change, que faire au jardin ? Jardins de France n°632.

Cornille A., Giraud T., Collin E., 2014. Conserver les ressources génétiques du pommier sauvage en France (Malus sylvestris). Ministère de l'Agriculture, Paris. 3 pages.

Gonin P., Larrieu L., Coello J., Marty P., Lestrade M., Becquey J., Claessens H., 2013. Autécologie des feuillus précieux. Institut pour le développement forestier. 64 pages.

Lamant T., Lévêque L., 2005. Pommier et poiriers sauvages : comment les reconnaître ? Les rendez-vous techniques n°8, Office national des forêts.

Lévêque L., Valadon A., Lamant T., 2005. Pommier et poiriers sauvages : réhabilitons les arbres à pépins en forêt ! Les rendez-vous techniques n°8, Office national des forêts.

Rameau J.C., Mansion D., Dumé G., 1989. Flore forestière française, guide écologique illustré. Tome 1 : plaines et collines. Institut pour le développement forestier.

Schnitzler A., Arnold C., Cornille A., Bachmann O., Schnitzler C., 2014. Wild European apple (Malus sylvestris (L.) Mill.) population dynamics: insight from genetics and ecology in the Rhine valley. Priorities for a future conservation programme. PLoS ONE, vol. 9, n°5.

Stephan B. R., Wagner I., Kleinschmit J., 2003. EUFORGEN Technical Guidelines for genetic conservation and use for wild apple and pear (Malus sylvestris and Pyrus pyraster). International Plant Genetic Resources Institute, Rome, Italie. 6 pages. Disponible sur internet : http://www.euforgen.org/publications/publication/wild-apple-and-pear-emmalus-sylvestris-and-pyrus-pyrasterem/

Annexe 3 : Article sur l'invasion de Microbotryum

The population biology of fungal invasions

Gladieux, P (Gladieux, P.)[1,2] ; Feurtey, A (Feurtey, A.)[1,2] ; Hood, ME (Hood, M. E.)[3] ; Snirc, A (Snirc, A.)[1,2] ; Clavel, J (Clavel, Joanne)[4] ; Dutech, C (Dutech, C.)[5] ; Roy, M (Roy, M.)[6] ; Giraud, T (Giraud, T.)[1,2]

[1] Univ Paris 11, Ecol Systemat & Evolut, F-91405 Orsay, France
[2] CNRS, F-91405 Orsay, France
[3] Amherst Coll, Dept Biol, Amherst, MA 01002 USA
[4] Univ Paris 06, Museum Natl Hist Nat, CNRS, Conservat Especes Restaurat & Suivi Populat CRBPO, F-75005 Paris, France
[5] Univ Bordeaux 1, INRA, Biodivers Genes & Communautes, F-33610 Cestas, France
[6] Univ Toulouse 3, Ecole Natl Format Agron, CNRS, Evolut & Divers Biol, F-31062 Toulouse, France

Introduction

Biological invasions have become a major focus of ecologists and evolutionary biologists, due to their severe negative consequences and their potential usefulness as models for studying basic processes in population biology (Sakai *et al.* 2001). However, invasions by fungi have only recently come into the limelight (Desprez-Loustau *et al.* 2007), even though most biological invasions are probably caused by micro-organisms, and fungal invasions may actually outnumber invasions caused by plants and animals. Furthermore, fungi, in the broad sense of the term (moulds, yeasts, mushrooms, lichens, rusts, smuts, mildews and including the phylogenetically distant oomycetes), display remarkably diverse life histories. This group includes everything from pathogens to mutualists and many free-living saprobes. They have a major impact on many different human activities and control essential ecosystem processes, such as nutrient cycling and carbon storage (Stajich *et al.* 2009). Fungal invasions thus constitute a key component of global change that is often overlooked.

When we think about fungal invasions, the species that come to mind are generally those responsible for emerging fungal diseases that jeopardize food security or drive natural populations into decline or even extinction (Palm 2001; Strange & Scott 2005; Fisher *et al.* 2012).

Jarrah dieback in Australia, caused by the oomycete *Phytophthora cinnamomi* (Hansen *et al.* 2012), and chestnut blight in North America, caused by the ascomycete *Cryphonectria parasitica* (Dutech *et al.* 2012), are textbook examples of ecological disasters caused by the establishment of pathogenic fungi outside their native range. However, it is important to recognize that nonpathogenic fungi are just as likely to have been moved across continents, even though their impact on ecosystem structure and function has rarely been quantified (Rizzo 2005). A particularly relevant example is that of mycorrhizal fungi (Schwartz *et al.* 2006), such as the infamous death-cap mushroom, *Amanita phalloides*, which was introduced into eastern North America, where it is now common (Pringle & Vellinga 2006). Besides the effects of pathogenic fungi on their hosts, little is known about the ecological impact of invasive fungi, in particular on resident microbial communities at the same trophic level. Further community-level or competition studies are therefore required, similar to those typically carried out in the invasion ecology of plants and animals (Díez 2005; Rizzo 2005; Schwartz *et al.* 2006). Fungi play fundamental roles in terrestrial ecosystem function and composition, and there is no reason to assume that the effects of nonpathogenic invasive fungi would necessarily be negligible or positive (Desprez-Loustau *et al.* 2007).

One of the key challenges in fungal invasion ecology is understanding and predicting introduction and establishment pathways, but baseline data concerning the composition of endemic fungal communities are limited, even for well-studied environments in temperate regions, hampering efforts to monitor invasive species and to determine their origin (Pringle & Vellinga 2006). Many invasive species were actually introduced well before they were recognized as such or before their potential impact was understood (Palm 2001; Rizzo 2005). The amphibian-killing chytrid fungus, *Batrachochytrium dendrobatidis*, illustrates this point clearly. It was only after *B. dendrobatidis,* which belongs to an understudied basally branching lineage of fungi, was first described in 1998 (Berger *et al.* 1998) that efforts were made to characterize the diversity of animal-associated chytrids and their environmental reservoirs (Morgan *et al.* 2007; Fisher *et al.* 2009; Farrer *et al.* 2011; Martel *et al.* 2013; McMahon *et al.* 2013). Genetic analyses pointed to a single, recent origin of an asexual worldwide lineage of *B. dendrobatidis*, and revealed additional cryptic global lineages of the fungus (Schloegel *et al.* 2012; Farrer *et al.* 2013). However, early studies were unable to identify the origin of the disease. With the exception of a few well-known examples (Table S1, Supporting information), the sources of invasive fungi and the timing of their introduction remain a mystery. Even the invasive status of some fungi can be unclear, as problematic fungi may have been there already and remained unidentified until their negative effects have begun to become apparent and of great concern. This uncertainty, in turn, impedes

our capacity to manage possible invasion pathways in a proactive manner and to implement the regulatory mechanisms required to reduce the risks of further disasters (Sakalidis *et al.* 2013). Current efforts to catalogue fungal diversity will allow increasingly precise and scientifically sound risk assessments, with the aim of preventing new introductions (Palm 2001).

Another important challenge in the field of invasion biology is determining whether adaptive evolution occurs in the invaded range or whether the successful invaders were already well suited for establishment and spread before their introduction (Facon *et al.* 2006), and which traits (*e.g.* asexuality) or situations (*e.g.* presence of naive hosts) facilitate invasions. Many invasive fungal populations are under sustained pressure to adapt to environments distinct from their endemic ranges. The response to these novel environments can lead to invasive fungi emerging on new hosts (Giraud *et al.* 2010), colonizing new varieties of their hosts (Brown 1994; Guérin *et al.* 2007), or re-emerging with greater pathogenicity (Hovmøller *et al.* 2008).

Here, we summarize and discuss recent studies on the ecology and population genetics of fungal invasions, with the dual aims of highlighting the features of these invasions and the ways in which they can help us to find out more about biological invasions in general. We begin by considering the reasons why fungal invasions have long been overlooked. We then examine the extrinsic factors and biological features underlying invasion success of certain fungal species. We present current knowledge about the origins and dispersal paths of invasive fungi, and documented ecological and evolutionary changes in regions where they have been introduced. We discuss how adaptation may occur so rapidly in bottlenecked, clonal fungal invaders. We also use a case study to illustrate the usefulness of baseline biodiversity data for the study of fungal invasions. For this, we present an original analysis of the invasion of the fungus causing anther-smut disease on white campion in the United States; the interaction between *Microbotryum lychnidis-dioicae* and its host *Silene latifolia* is one of the best-studied nonagricultural plant–microbe pathosystems (Bernasconi *et al.* 2009). Overall, we aim at showing that fungi provide excellent and important models for studying the patterns and processes of biological invasions, for elucidating the factors favouring successful introductions, and for understanding their evolutionary and ecological mechanisms.

Fungal biogeography: ‘nothing is generally everywhere’

The first obvious step in any study of fungal invasions is the characterization of the geographic limits of endemic species of fungi. A failure to do this correctly can severely limit the ability to detect invasions and, thus, to infer their causal factors. Like other organisms of less

than a few millimetres in length, fungi have long been thought to have global distributions, contrasting the highly restricted geographic ranges of larger organisms (Bisby 1943; Martiny *et al.* 2006; Taylor *et al.* 2006). This hypothesis was based on the observation that almost all fungi have small, powder-like propagules (typically spores <10 µm in diameter) and that many have structures favouring their dissemination (Pringle *et al.* 2005), suggesting that dispersal ability *per se* is unlikely to limit the geographic distribution of these organisms (Bisby 1943; Peay *et al.* 2010; Sato *et al.* 2012). The misconception that many fungi had global distributions also resulted from the use of morphological species recognition criteria that are largely inappropriate for fungi, providing an erroneous picture of fungal diversity, distributions and ecologies (Swann *et al.* 1999; Taylor *et al.* 2006). Obviously, if all fungi were ubiquitous, there would be no reason for writing this article, as no fungal species would be exotic anywhere. As startling as it may sound to researchers studying plants and animals, it is only recently that most fungi have been shown to have distributions similar to those of their distant macrobial relatives (Green & Bohannan 2006; Thorsten Lumbsch *et al.* 2008). They are, for example, rarely present across entire continents (Tedersoo *et al.* 2012), and they satisfy the expectations of island biogeography theory (MacArthur & Wilson 2001; Peay *et al.* 2007).

The use of DNA markers have made it possible to identify new species that were previously cryptic, as these methods are more discriminating in fungi than traditional approaches based on morphological phenotypes or mating success (Giraud *et al.* 2008a). Modern evolutionary concepts of species are based on the lack of gene flow, with species being lineages that evolve independently (De Queiroz 2007; Giraud *et al.* 2008a), and DNA markers allow direct assessment of allelic exchange. In particular, genealogical concordance phylogenetic species recognition (GCPSR) is rapidly becoming the gold-standard approach for delimiting species (Henk *et al.* 2011; Hibbett & Taylor 2013). GCPSR uses breaks in the concordance of phylogenies for multiple genes to determine the limits of recombining entities (Avise & Wollenberg 1997; Taylor *et al.* 2000). It is one of the most objective, reproducible, achievable and widely applicable criteria for species identification (Giraud *et al.* 2008a). This approach appears to be the most powerful and accurate in fungi, leading to the identification of larger numbers of species than previously estimated (Giraud *et al.* 2008a; Stewart *et al.* 2014).

The cryptic species identified using DNA markers usually have narrower distributions in terms of geography range or host diversity than had been suggested in the prior literature (Taylor & Fisher 2003). The use of DNA markers has thus revealed that major geographic features (*e.g.* continental breaks, mountains) can act as effective barriers to fungal dispersal and that limited dispersal, combined with competitive advantage due to the colonization of resources

ahead of competitors (Kennedy 2010), is a major driver of fungal endemicity (Peay *et al.* 2010). In fungal symbionts, strong specialization on different hosts and host characteristics (*e.g.* life history traits and distribution) appears to be key elements underlying biogeographic patterns (Vellinga *et al.* 2009; Hood *et al.* 2010; Schulze-Lefert & Panstruga 2011).

Which biological features underlie the success of invasive fungal species?

Most fungi thus did not have cosmopolitan distributions, but an increasing number of fungi are spreading around the world, some with dramatic effects on plant or animal populations (Fisher *et al.* 2012). It is therefore of paramount importance to examine the extrinsic factors and biological features underlying invasion success of certain fungal species. The success of invasive species is dependent on their ability to overcome a series of barriers, from introduction to establishment and invasion in a novel environment (Kolar & Lodge 2001), which involve stochastic and selection events in addition to initial transport by humans (Fig. 1). Many fungal species are highly prolific, producing large numbers of spores (Peay *et al.* 2012; Prussin *et al.* 2014), thereby increasing propagule pressure, and potentially facilitating dispersal and persistence in the new environment, even if the initial degree of adaptation is low (Giraud *et al.* 2010).

Unlike many invasive plant and animal species that were initially deliberately introduced into new areas for economic or other purposes (*e.g.* recreational), fungal introductions are mostly an unintended consequence of human-mediated movement and trade (Desprez-Loustau *et al.* 2007). Such 'hitchhiking' dispersal is favoured by some typical fungal features, including their inconspicuousness and the production of numerous, small propagules. The trade of animal and plant goods is a major pathway of fungal introductions, due to the close association of many fungi with plant or animal hosts (in mutualistic, pathogenic or saprotrophic interactions; Desprez-Loustau *et al.* 2010; Liebhold *et al.* 2012). Even though some fungi provide striking examples of long-distance aerial dispersal (Brown & Hovmøller 2002) and of global distributions (Pringle *et al.* 2005; Henk *et al.* 2011), it is generally very difficult to discount a role for human-mediated long-range movement in current distributions or migration pathways. Dispersal rates of several kilometres or even up to tens of kilometres per day over a season have nevertheless been reported for some species (Aylor 2003; Mundt *et al.* 2013). Dispersal capacity, including fecundity and spore traits in particular, is critical in determining colonization success following the introduction of a species into a new environment and has been shown to affect variation in

invasion success between species in a large data set of fungal tree pathogens (Philibert *et al.* 2011). Mode of dispersal may also account for the overrepresentation of aerial pathogens among invasive plant pathogenic fungi (Desprez-Loustau *et al.* 2010).

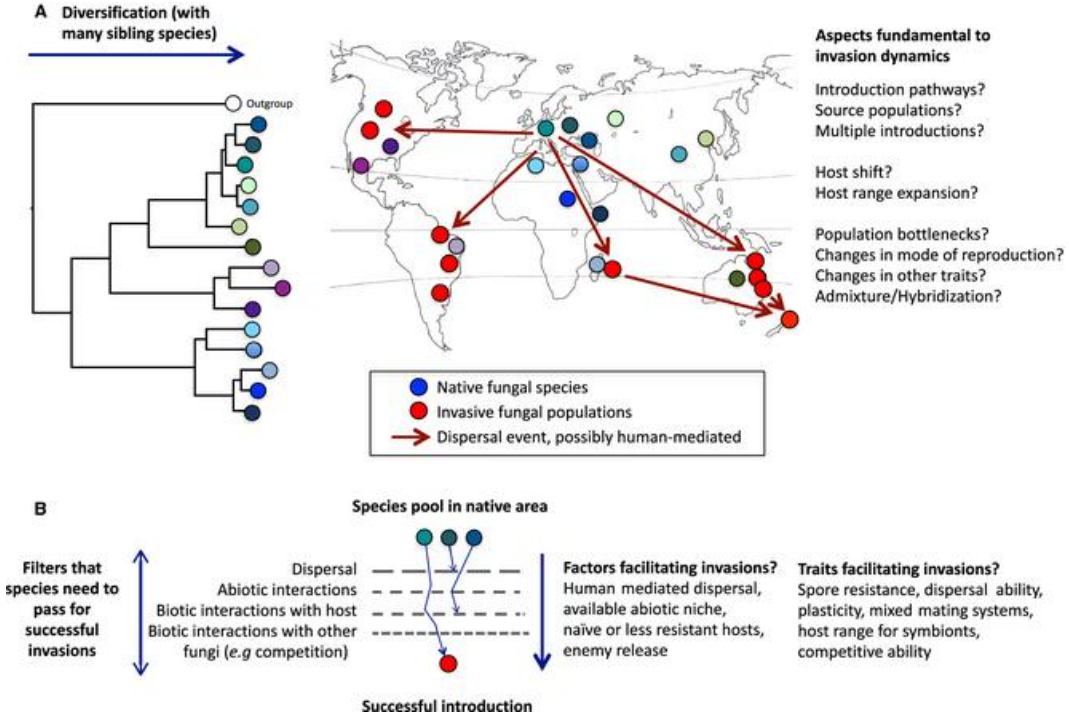


Figure 1 : An invasion scenario typical for fungi and essential questions to illuminate the process. (A) The figure shows the phylogenetic diversity and geographic distribution of a fungal group from which a species has invaded other regions and lists the questions fundamental to invasion dynamics: recognition of local species diversity in the native range, with the delimitation of sibling species, identification of the traits facilitating invasion, source populations, introduction pathways, occurrence of bottlenecks or of multiple introductions, possible evolutionary or ecological changes, such as host shift, changes in mode of reproduction or in other traits. The red circles represent introduced populations while the nonred circles represent native fungal species. (B) Successive filters that species must pass for invasions to be successful (dispersal, abiotic and biotic interactions) and possible mechanisms facilitating invasions (Human-mediated dispersal, available abiotic and biotic niche, enemy release, spore resistance, dispersal ability, plasticity, mixed mating systems, host range for symbionts or competitive ability).

As with other organisms, the successful establishment of fungal species outside their native range can often be accounted for by some level of pre-adaptation: the presence of attributes allowing these species to survive elsewhere, in novel environments (Mayr 1965; Ianzen 1985), possibly due to traits displaying relatively high phenotypic plasticity or conferring a broad ecological niche, *for example,* in terms of temperature or possible host range (Parker & Gilbert 2004; Agosta & Klemens 2008; Hulme & Barrett 2013). Pre-adaptation has been proposed to explain the ability of fungal species to colonize extreme environments (Robert & Casadevall 2009; Gostinčar *et al.* 2010). An example of phenotypic plasticity that has been experimentally evaluated for the potential to facilitate invasion is the breadth of potential host range in pathogens. Pathogenicity tests have indeed shown that fungal pathogens of plants are often able

to infect hosts closely related to their original host, although initially with low aggressiveness (Table 1; de Vienne *et al.* 2009; Gilbert & Webb 2007).

Table 1. Glossary

Term	Definition
Host range expansion	Colonization of a new host species, while retaining the ability to infect the ancestral host
Host shift	Colonization of a new host species associated with a loss of the ability to infect the ancestral host
Host tracking	Codivergence of the host and the pathogen during crop domestication, with a lineage of fungal pathogens differentiating from that infecting wild crop progenitors, specialization on the domesticated species and the subsequent invasion of crops worldwide
Effectors	Secreted proteins that manipulate host innate immunity, enabling infection to occur (Dodds & Rathjen 2010)
Virulence	Ability of a pathogen to cause disease on a host (qualitative measure of pathogenicity)
Aggressiveness	Degree of damage caused by a pathogen on a host (quantitative measure of pathogenicity)

Alternatively, successful invaders may adapt to the new environment after their introduction, and rapid and drastic evolutionary changes may be facilitated by the huge census population size and mixed mating systems of fungi. Many fungal species can undergo both sexual and asexual reproduction, which could facilitate establishment from few founding propagules and accelerate adaptation in new environments. Theoretical studies have shown that species with mixed mating systems (both asexual reproduction and rare sex events) have the highest potential for invasion success in terms of probability and time to establishment (Bazin *et al.* 2014). Mitotic spores and vegetative reproduction promote rapid increases in population size and the spread of the fittest genotypes, and several cycles of asexual reproduction may be completed in a single season (*e.g.* polycyclic plant pathogens). Meiotic spores promote the creation of novel genotypes through recombination, and they are mostly used as dispersal and survival structures (Taylor *et al.* 1999).

Several traits have been shown to be predictive of invasion success for forest pathogenic fungi, including traits relating to long-distance dispersal, sexual reproduction combined with

asexual reproduction, spore shape and size, number of cells in spores, optimal temperature for growth and broad host range for pathogens (host range and infected organs; Hayes & Barry 2008; Philibert *et al.* 2011; Pyšek *et al.* 2010). The traits cited above are intuitively appealing as facilitators of introduction and establishment, but other, more complex factors, such as spore resistance, competitive ability, host range for symbionts and evolutionary or plasticity potential, are also probably important. Indeed, fungal invaders successfully establish themselves in new environments, to which they probably must adapt despite their low genetic diversity. This expectation to evolve despite the genetic resources is often described as the 'genetic paradox of invasions'. Below, we call into question the reality of this paradox in fungal invasions, and we review some of the solutions adopted by fungi.

The fungal solution to the genetic paradox of invasions

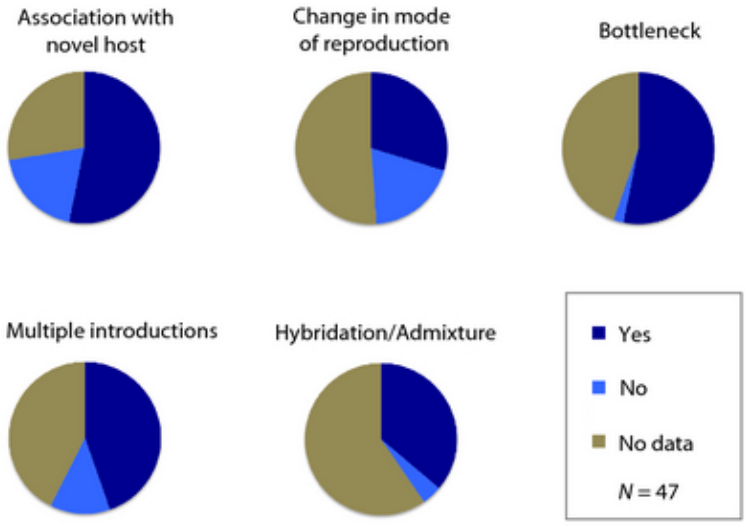
Is reduced genetic diversity commonplace in invasive fungi?

Colonization bottlenecks are generally thought to decrease adaptive potential. Nevertheless, founder events have been shown to contribute to invasion success in some cases, by increasing the genetic variance through the conversion of epistatic to additive variance (Naciri-Graven & Goudet 2003) or by reducing the genetic load (Facon *et al.* 2006). Admixture between populations from multiple introduction sources can also promote adaptation by rapidly creating novel allelic combinations (Prentis *et al.* 2008).

To assess whether bottlenecks or admixture occurred, one-first needs to determine the origin of the invasive species and the genetic make-up of the source populations. For fungal symbionts of domesticated agricultural and forest plants, it is tempting to speculate that the origin of an invasive population is the geographic origin of the host. This has been shown to be the case for many crop pathogens (Table S1), through the demonstration that the centre of diversity of the fungus coincides with the centre of origin of its host plant (Gladieux *et al.* 2008; Munkacsi *et al.* 2008). However, some fungal pathogens have been found to have a higher level of diversity outside the centre of origin of their current host, and this has been taken as indirect evidence for diversification following host-shift/range expansion (Zaffarano *et al.* 2009; Ali *et al.* 2014). Advances in clustering methods and approximate Bayesian computation have made it possible to carry out rigorous comparisons of complex scenarios modelling genealogical relationships between populations and their geographic and demographic dynamics (*e.g.* Barrès *et al.* 2012; Dilmaghani *et al.* 2012; Dutech *et al.* 2012; Goss *et al.* 2014). Despite these

methodological advances, it remains difficult to identify the origin of invasive symbiotic fungi because they are inconspicuous, hindering delimitation of the possible regions, substrates and hosts of origin. This uncertainty raises possibilities for endless disputes between experts (Goss *et al.* 2014) and hampers our understanding of the initial steps leading to invasion, such as the prevalence of major niche shifts.

In most cases in which genetic diversity has been compared between introduced and native fungal populations, a bottleneck signal has been detected, but such footprints have rarely been quantified in terms of effective population size ratios (Fig. 2; Table S1). Highly contrasting situations have been reported, from the introduction of a single genetic lineage (*e.g.* Raboin *et al.* 2007; Hovmøller *et al.* 2008) to greater genetic diversity in the invaded area than in the native range due to multiple introductions from diverse sources (Dilmaghani *et al.* 2012; Fig. 2; Table S1). Admixture between heterogeneous genetic pools has been detected in several studies on invasive fungi (*e.g.* Ahmed *et al.* 2012; Dilmaghani *et al.* 2012; Dutech *et al.* 2012; Jezic *et al.* 2012), but the source populations have rarely been identified (Fig. 2; Table S1). The detection of older admixture events requires the use of coalescent theory-based methods if several rounds of reproduction have erased allele frequency differences between the founding populations.



Pie charts representing the frequency of documented cases of associations with novel hosts, bottlenecks, multiple introductions, hybridization/admixture and changes in mode of reproduction (shift to clonality), in a sample of N = 47 *invasive fungi or oomycetes. These are the 47 cases, found in an exhaustive literature search, for which data were available on the population genetics of invasion or on the occurrence of genetic changes in introduced regions. Details and references are presented in Table S1. For each of the 47 cases, the following information was recorded when available: species name, taxonomic affiliation (*e.g. *Basidiomycota, Oomycota), lifestyle (*e.g. *tree mutualist, animal pathogen, crop pathogen, tree pathogen), native region, introduced area, host in native range, host in introduced range, change in mode of reproduction (shift to clonality), change in other life history traits (*e.g. *increase in aggressiveness, new virulence traits, adaptation to higher temperatures), putative introduction pathway (*e.g. *trade of host plants), evidence for the occurrence of a bottleneck and its strength, evidence for the occurrence of multiple introductions in at least one introduced area, evidence for the occurrence of hybridization between species or admixture between gene pools and the corresponding references. Admixture refers to cases where the mixing of two distinct lineages has been shown in clustering analyses or when it has been explicitly tested using simulations (*i.e. *the occurrence of multiple introduction* per se *is not considered as sufficient to infer 'admixture'). When the native region has not been identified, an emergence without invasion (*e.g. *by evolution of higher virulence or host shift) can often not be excluded.*

Interspecific hybridization: increased genetic diversity promoting rapid evolutionary changes

Interspecific hybridization can also promote rapid adaptation (Prentis *et al*. 2008) and has been reported between native and invasive species in fungi (*e.g*. Bovers *et al*. 2008; Brasier 2001; Gonthier *et al*. 2007; Lin *et al*. 2007; Montarry *et al*. 2010; Stukenbrock *et al*. 2012; Fig. 2; Table S1). In some cases, hybridization has been shown to be a powerful mechanism of fungal adaptation to new hosts and environments, allowing the acquisition of a missing mating type, for example, in *Ophiostoma novo-ulmi* (Paoletti *et al*. 2006). Interspecific hybridization, as a process allowing the adaptation of organism to new environments, may be facilitated in fungi by the large number of sibling species with incomplete reproductive isolation (Le Gac & Giraud 2008) and by the possibility of non-Mendelian forms of exchange, such as hyphal anastomosis that allows, for instance, the transfer of supplementary chromosomes (Roper *et al*. 2011; Gladieux *et al*. 2014).

Changes in life history traits during fungal invasions: mode of reproduction, virulence, host range and adaptation to abiotic factors

Investigators are generally interested in the questions discussed above, regarding levels of genetic diversity in invasive populations, because they implicitly assume a direct relationship between genetic variation and adaptive potential and a need for evolutionary change in introduced ranges. Many changes in life history traits during fungal invasions have actually been documented, with most studies focusing on ecological interactions of symbiotic fungi and particularly on plant pathogens. These changes have involved the infection of a new host population or species, adaptation to abiotic factors (*e.g*. temperature) and changes in reproductive mode (Fig. 2; Table S1). New ecological associations may result from fungi exhibiting wide ecological niches breadth, plasticity (*i.e*. phenotypic changes without evolution) or, alternatively, may have involved adaptation or trade-offs, but these possibilities are rarely tested experimentally. For instance, the infection of a new host by a fungal population may involve a pre-existing broad host range, a host shift, a host-range expansion or plasticity, but the relative importance of these processes is difficult to establish unequivocally, as the pathogenicity of fungal populations infecting ancestral and new hosts has rarely been compared directly. For simplicity, we will use the term ‘host shift’ below, but it should be borne in mind that it has not been clearly demonstrated whether these phenomena correspond to host shifts in the narrow

sense (*i.e.* with evolution, allowing infection of a new host and a decrease in the ability to infect the original host), host range expansion (*i.e.* with evolution increasing host range), plasticity (*i.e.* expansion of observed host range without genetic change) or an initially broad host range (*i.e.* with no expansion of observed host range).

Several well-documented cases of emerging plant diseases have been attributed to such infection of a new host by introduced fungal pathogens (Desprez-Loustau *et al.* 2007; Stukenbrock & McDonald 2008; Giraud *et al.* 2010). Examples include wild tree or shrub pathogens, with the devastating shifts of *Cryphonectria parasitica* from the Asian to European and North American chestnuts, *Cronartium ribicola* from Asian to American and European pines, and *Discula destructiva* from Asian to American *Cornus* spp. (Palm & Rossman 2003; and references therein). Other examples, for crop pathogens, include the rice blast fungus *Magnaporthe oryzae*, which originated on *Setaria* millet (Couch *et al.* 2005), or the wheat pathogen *Zymoseptoria tritici*, which is derived from wild species infecting other grasses (Stukenbrock *et al.* 2007). The Irish potato famine pathogen was also recently shown to have emerged from populations infecting wild relatives of potato in Mexico (Goss *et al.* 2014). These host shifts may be due to recent contact with naïve host or may have involved evolutionary changes to permit infection of the new host. Such shifts to a new host would be facilitated by certain life history traits of fungal pathogens favouring the colonization of new hosts, such as mating within hosts and strong selective pressure on a limited number of genes (Giraud *et al.* 2010; Gladieux *et al.* 2011). Host shifts have actually been shown to be the main mode of diversification in fungal pathogens (Giraud *et al.* 2008a, 2010; de Vienne *et al.* 2013b).

Associations with new hosts after introductions into new geographic regions have been less frequently documented for mutualists. There are nonetheless several examples in ectomycorrhizal fungi, such as *Laccaria* spp. in Costa Rica (Mueller & Strack 1992) and the emblematic *Amanita muscaria*, where introductions of these fungi appear to have occurred with their host trees species, followed by the formation of new associations with native hosts (Bagley & Orlovich 2004; Díez 2005; Rizzo 2005; Vellinga *et al.* 2009; Jairus *et al.* 2011).

Aside from host shifts or host range expansions occurring in introduced ranges due to contact with new hosts, several invasive crop pathogens have also emerged through host tracking (Table 1), a process by which host and pathogen codiverge during crop domestication and the development of agro-ecosystems (Stukenbrock & McDonald 2008). Movement of the pathogen from the wild progenitor of the crop to incipient domesticated populations can be followed by invasion at a worldwide scale associated with agricultural trade and product distributions. Examples include corn smut, caused by *Ustilago maydis* (Munkacsi *et al.* 2008),

and ascochyta blight of chickpea, caused by *Didymella rabiei* (Frenkel *et al.* 2010). An interesting case is *Venturia inaequalis*, causing apple scab, which tracked its host tree during domestication in Asia and along the Silk Route to Europe, where host range expansion to the European wild apple occurred, with a certain degree of pathogen specialization on apple varieties (Gladieux *et al.* 2008, 2010; Le Van *et al.* 2012). In this case, cross-inoculation experiments indicated the occurrence of evolutionary change, in particular the increased aggressiveness and the acquisition of new virulence traits (Le Van *et al.* 2012).

Only a few studies have investigated whether changes in life history traits have occurred in introduced fungal populations, and whether such changes were due to adaptation in particular. However, some evidence has been obtained for adaptation in the form of ecological traits that function to enhance fitness in the new environments. Such changes include an increase in aggressiveness and/or larger numbers of virulence traits in introduced regions (Table 1; Fig. 2; Table S1; Enjalbert *et al.* 2005; Gilbert & Parker 2010; Le Van *et al.* 2012; Markell & Milus 2008; Milus *et al.* 2009; Rouxel *et al.* 2013). The development of greater aggressiveness or a broader virulence spectrum may be a general feature of crop pathogens, connected to the agricultural practice of very dense, homogeneous populations of host plants growing on a vast scale (Stukenbrock & McDonald 2008; McDonald 2010). Indeed, the presence in the environment of numerous individual hosts with the same genotype strengthens the selection on the pathogen for the ability to infect this genotype. Furthermore, the availability of numerous, regularly replanted host individuals should select for higher aggressiveness, because the pathogen would not benefit from a prudent host exploitation strategy when there are constantly new host individuals to infect in its environment (Frank 1996).

Another frequent change in life history traits of introduced fungi is the loss of sexual reproduction. Many studies have documented clonal population structures throughout the entire introduced range (Table S1; Fig. 2; Ali *et al.* 2010; Couch *et al.* 2005; Dobrowolski *et al.* 2003; Duan *et al.* 2010; Enjalbert *et al.* 2005; Farrer *et al.* 2011; Fisher *et al.* 2009; Goodwin *et al.* 1994; Goss *et al.* 2009; Ivors *et al.* 2006; Mboup *et al.* 2007, 2009; Saleh *et al.* 2012a,b; Singh *et al.* 2011; Smart & Fry 2001; Stajich *et al.* 2009) or over parts of that range (Gladieux *et al.* 2008; Dutech *et al.* 2010; Dilmaghani *et al.* 2012). Evidence has also been provided for an evolution towards a lower capacity for sexual reproduction in introduced regions (Couch *et al.* 2005; Enjalbert *et al.* 2005; Mboup *et al.* 2007, 2009; Ali *et al.* 2010; Duan *et al.* 2010; Saleh *et al.* 2012a,b). F_{ST} versus Q_{ST} comparisons have suggested that this pattern is indeed an adaptation, rather than merely the consequence of genetic drift (Ali *et al.* 2010).

There may initially be several proximal causes for a loss of sexual reproduction: (i) a lack of appropriate mates where asexual propagation is possible, as in the case of *Phytophthora infestans, P. cinnamomi, P. ramorum* and *O. novo-ulmi,* for which a single mating type was present for long periods in introduced ranges (Goodwin *et al.* 1994; Brasier 2001; Dobrowolski *et al.* 2003; Hardham 2005); (ii) a hybridization event, *for example,* in the human pathogen *Cryptococcus neoformans* (Lin *et al.* 2007); and (iii) a lack of the alternate host on which sex occurs in the case of complex life cycles, *for example*, in the rusts *Puccinia striiformis f. sp.tritici* and *Melampsora larici-populina* (Mboup *et al.* 2007, 2009; Barrès *et al.* 2008).

There may also be diverse evolutionary causes for the selection against sexual reproduction (Bazin *et al.* 2014): (i) reduced advantages of sex due to the release of an enemy, such as viral pathogens of fungi, or the lower host resistance variation in the case of pathogens introduced into large, genetically homogeneous crops; (ii) in heterokaryotic fungi and diploid oomycetes, preservation of the initial heterozygosity, which might otherwise rapidly convert to homozygosity in small founder populations, leading to inbreeding depression; (iii) a release of constraints not related to recombination, maintaining sex in the short term in native ranges, such as the need for sexual spores for winter survival or long-distance dispersal (de Vienne *et al.* 2013a), which is particularly likely for introduced pathogens on crops, due to the availability of large, homogeneous host populations; and (iv) clonality may allow the locally adapted genotypes to be propagated without breaking up adaptive allelic combinations; again, this is advantageous on genetically homogeneous crops planted over vast areas.

Finally, abiotic factors can indeed limit the success of introductions or require adaptation, as a tolerance to high temperatures has been shown to have evolved in the introduced ranges of some fungi. In models of distribution range, temperature has been shown to be a critical factor for successful establishment of invasive fungi (Wolfe *et al.* 2010). Some populations of wheat yellow rust in the Mediterranean basin and south-eastern regions of the USA have evolved temperature-specific adaptations (Mboup *et al.* 2012), as did the *D. rabiei* population infecting domesticated chickpea (Frenkel *et al.* 2010). Furthermore, the successful establishment of exotic fungi has been shown to be sensitive to moisture conditions in attempts to deliberately introduce edible fungi such as truffles (Vellinga *et al.* 2009).

Invasive fungi, particularly pathogens of agricultural and forest plants, thus provide many examples of environmental change or colonization of a new niche acting as the primary drivers of post-colonization adaptive evolution. It is clearly important to understand the biological parameters driving selection on pathogens and disease outbreaks, given their impact on ecosystem functioning, human health and agricultural production.

Fungal pathogens evolve rapidly in agricultural and forest ecosystems

The potential for a fast pace of evolution in invasive fungal pathogens is a major impediment to maintaining ecosystem health and sustainable agricultural production. Measurements of genetic diversity and population structure are commonly used to draw inferences regarding the evolutionary potential of these pathogens, based on the assumption that adaptation can proceed quickly only from standing variation (McDonald & Linde 2002; McDonald 2014). Life history strategies (*e.g.* transmission and dispersal, virulence, mode of reproduction) are considered to be important determinants of diversity and, thus, for the potential of pathogens to colonize and spread on new host genotypes (Barrett *et al.* 2008). While standing genetic variation in a population is clearly a key influence on its ability to adapt in response to an environmental challenge, the general validity of predictive frameworks based on measured diversity patterns is questionable, given that evolution occurs rapidly in many invasive plant pathogens despite relatively modest levels of genetic diversity. For instance, some of the textbook examples of rapidly adapting fungal pathogens, such as lettuce downy mildew and wheat yellow rust, are also notoriously clonal and harbour little genetic diversity (Valade 2012; Ali *et al.* 2014).

The classical model of positive selection assumes that adaptation proceeds principally from *de novo* mutations, in a mutation-limited regime (Maynard-Smith *et al.* 1974). Under this regime, the lag time to adaptation depends on the product of the rate of mutation to the adaptive variant at the relevant locus and effective population size (*Ne;* Charlesworth 2009; Garud *et al.* 2013). Many plant pathogens have a small *Ne* as estimated from levels of standing neutral variation (*e.g.* Zaffarano *et al.* 2009; Dilmaghani *et al.* 2012; Ali *et al.* 2014). This apparent barrier to adaptation is because the 'diversity *Ne*' values are primarily determined by long-term population dynamics, which are dominated by extreme episodes of population crashes (bottlenecks or founder events; Lewontin 1974), hallmarks coincidentally of both disease cycles and invasions dynamics. If we consider the mutation rate per site to be between 10^{-8} and 10^{-10} (Kasuga *et al.* 2002), the adaptation of invasive fungi should, therefore, be strongly mutation limited*, that is* relatively slow and involving a single adaptive variant (a 'hard selective sweep').

At least two factors may nevertheless allow the rapid adaptation of invasive fungi, despite their relatively small 'diversity *Ne*'. First, recent studies have shown that predicting the pattern of adaptation (speed of adaptation and number of adaptive alleles) is more accurate if *Ne* is estimated over timescales relevant to adaptation (Karasov *et al.* 2010). For an invasive fungal pathogen, this may be the timescale over which a new fungicide is applied or a new

resistant variety is released. Despite the regular crashes of fungal populations (*e.g.* during winters), their short-term *Ne* can be expected to be much higher than their long-term *Ne*, particularly as they generally have periods of huge census size. In such very large populations, the distinction between standing variation and *de novo* mutations becomes blurred, because almost all the possible mutations of the target gene are presented in the population (Barton 2010; Karasov *et al.* 2010).

Second, adaptation may be faster if the mutational targets are larger than a single nucleotide (Pritchard *et al.* 2010). The mutational target size is the number of base pairs within an allele for which mutations would alter a particular phenotype and increase fitness in the new environment. For fungal secreted virulence determinants (effectors, Table 1), for instance, the mutational target size may be of the order of hundreds of base pairs within the gene, rather than one: changing any one among hundreds of nucleotide, or even a large deletion in fungal effectors, may allow escaping host recognition (Rafiqi *et al.* 2012). Adaptive alleles would thus be expected to evolve repeatedly and relatively quickly.

Recent case studies in invasive fungal pathogens have revealed several examples of adaptive mutations of multiple, independent *de novo* origins sweeping through populations (so-called soft selective sweeps), providing evidence that adaptation is not necessarily mutation-limited in these organisms. For instance, the molecular characterization of virulence alleles of effector genes in *Leptosphaeria maculans* (stem canker of oilseed rape) and *Rhynchosporium secalis* (barley leaf scald) revealed a diversity of mutational events consistent with a model of adaptation via soft sweeps (Schürch *et al.* 2004; Gout *et al.* 2007; Fudal *et al.* 2009; Daverdin *et al.* 2012). Another example is provided by fungicide resistance in *Plasmopara viticola* (grapevine downy mildew), which has arisen from several independent mutations (Chen *et al.* 2007).

Case study of anther-smut fungus on white campion

The insights into fungal invasions using comprehensive analyses of species boundaries, population structure and life history traits can be illustrated using the anther-smut disease of Caryophyllaceae, caused by the basidiomycete *Microbotryum violaceum sensu lato*. The anther-smut fungi produce spores in anthers of diseased plants and are mostly transmitted by pollinators. A sex event is required before infection of a new plant, and the mating system is highly selfing (Giraud *et al.* 2008b). *Microbotryum* pathogens on the Caryophyllaceae are relatively uniform morphologically and were long thought to be a single, widely distributed

species. However, GCPSR has shown that the old epithet actually consists of a large number of different, host-specialized species (Lutz *et al.* 2005; Le Gac *et al.* 2007) that diversified following host shifts (Refregier *et al.* 2008). One of the defined species, *Microbotryum lychnidis-dioicae,* and its particular host plant, *Silene latifolia,* are native to Europe. Based on a collection of samples of unprecedented density and geographic scale for a pathogen, cluster analyses revealed the existence of strong subdivisions within the geographic range of this fungal species, reflecting recolonization from southern glacial refugia (Vercken *et al.* 2010; Gladieux *et al.* 2013). The genetic structure of the pathogen populations was found to be generally congruent with the genetic structure of the host, *S. latifolia* (Taylor & Keller 2007), both exhibiting three genetic clusters distributed, respectively, across western Europe, eastern Europe and Italy. This congruence of population structures suggests that the migration pathway of the fungus had been dependent on its host over a time scale of hundreds or thousands of years. However, *M. lychnidis-dioicae* exhibited a much stronger structure than observed for its host, each of the three main fungal clusters being subdivided into further genetic clusters, with more restricted geographic distributions (Vercken *et al.* 2010). This suggested that *M. lychnidis-dioicae* has recolonized central and northern Europe from more numerous, smaller refugia than sheltered its plant host and that it has experienced fewer large-scale dispersal events.

Both the pathogen and its host were introduced relatively recently in North America, and the invasion history of *S. latifolia* has been extensively studied (Keller *et al.* 2009, 2012). The diversity of the host plant, *S. latifolia,* was found to be high in the introduced range, with the introduction of multiple lineages from several European genetic clusters that are now spread across North America (Keller *et al.* 2009, 2012). The anther-smut disease appeared much more recently on *S. latifolia* in North America, and with a more eastern restricted distribution, than observed for the host plant (Antonovics *et al.* 2003; Fontaine *et al.* 2013). The delimitation of anther-smut species confirms that the disease on the introduced *S. latifolia* was the result of a later introduction by its specialist pathogen, *M. lychnidis-dioicae,* instead of the host shift of a *Microbotryum* species endemic to North American on other *Silene* species or from another European host species (Fontaine *et al.* 2013). The very strong genetic structure of *M. lychnidis-dioicae* in Europe made it possible to identify the United Kingdom as the most probable source of the limited set of genotypes introduced into the central East Coast of North America. Given the assignment results and the limited number of genotypes in North America, a single introduction event appeared the most likely scenario (Fontaine *et al.* 2013). However, only a few samples from the United Kingdom were available for study at that time, and we have since identified an additional diseased population in the state of Massachusetts.

We therefore sampled further populations in the United Kingdom and from the newly identified diseased regions in North America, in an effort to ascertain the origin of the invasion more precisely and to check whether independent introductions may be responsible for the invasions of central East Coast and more northern populations. The samples analysed here represent an exhaustive sampling of the extant North American populations of *M. lychnidis-dioicae,* and a representative subset of the European populations previously analysed (Vercken *et al.* 2010; Fontaine *et al.* 2013). Fungal isolates were genotyped and analysed as previously described (Fontaine *et al.* 2013), with the inclusion here of more microsatellite markers (fourteen, see data on dryad). Fourteen multilocus genotypes were found in North America, representing a strong bottleneck compared with the diversity of *M. lychnidis-dioicae* across Europe, where 174 genotypes were detected in the present sample. The values of allelic richness, a standardized measure of diversity corrected for sample size differences, also pointed to a strong bottleneck: allelic richness was 3.78 in North America versus 7.46 in Europe and 5.55 in the UK alone. Genotypes were highly homozygous in all geographic regions, and no admixture was detected, indicating that the fungus has retained its highly selfing mating system in the introduced range (Giraud *et al.* 2008b). Sexual reproduction has also remained obligate for plant infection in the introduced range (MEH, personal observation)

We used structure assignment analyses to investigate the origin of invasive fungal populations. This provided evidence for an additional, independent introduction of *M. lychnidis-dioicae* in North America (Fig. 3A; Fig. S1, Supporting information). The previously sampled populations in Virginia and one of the populations from Massachusetts still appeared, as in our previous study (Fontaine *et al.* 2013), to have originated in the Lothian (Edinburgh) region in southern Scotland (red cluster in Fig. 3A). However, the second Massachusetts population was assigned to another genetic cluster (yellow cluster in Fig. 3A), which included populations spread across the UK and could not be separated by further increasing the *K* parameter in structure analyses (number of clusters). The examination of individual genotypes, however, allowed narrowing down the possible sources of this second introduction. Genotypes that were identical to the ones found in the North American yellow cluster (Fig. 3A) were also found in the north of Scotland, in the Moray council area and in the Highland close to cities of Inverness and Thurso (Fig. 3B). The finding of two distinct genetic clusters in North American populations of *M. lychnidis-dioicae*, found in different locations and assigned to clusters originating from different regions in Scotland, altogether strongly suggests two or more independent introductions of the anther-smut fungus.

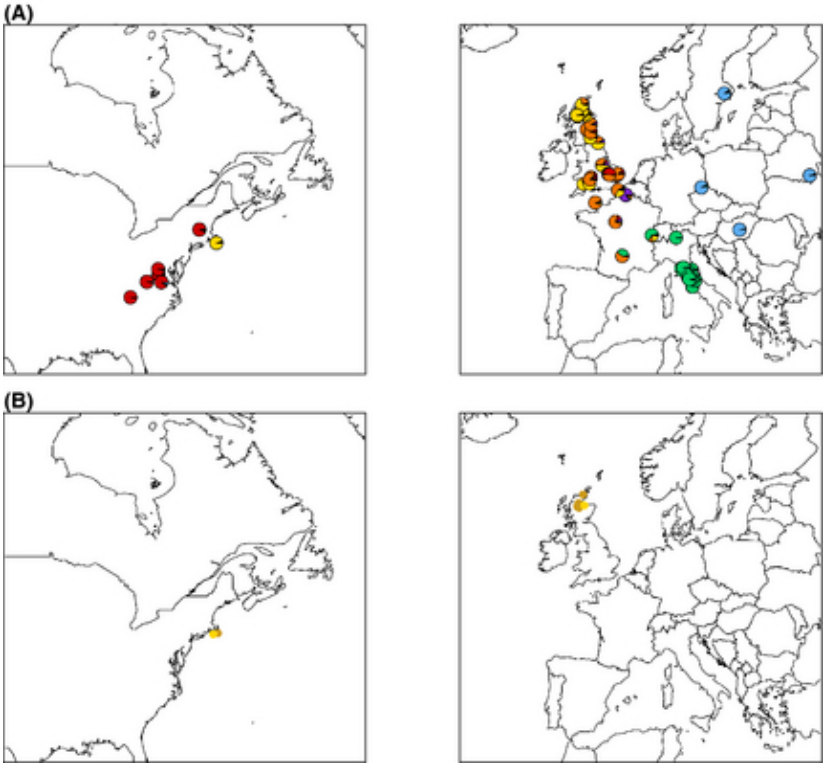


Figure 3: Assignment of North American fungal Microbotryum lychnidis-dioicae *samples, collected on* Silene latifolia, *to European populations. (A) Map of the samples collected in the United States and Europe (*N *= 328 individuals), with colours representing the mean membership proportions inferred by structure for* K *= 6; higher* K *did not allow assigning more precisely American isolates to European populations, see Fig. S1. (B) Map of the samples assigned to the yellow cluster in (A) and with identical genotypes in the United States and in Scotland; a specific colour (yellow or brown) has been given for each of the two multilocus genotypes of the yellow cluster in the USA. Colours represent the inferred ancestry from* K *ancestral populations. The percentages indicated are the proportions of runs that found the main solution shown here.*

The fungal invasion history still represents a much smaller number of introductions than for its host plant species (Keller *et al.* 2009, 2012), probably reflecting a restriction by the pathogen's life history traits. Indeed, this obligate, specialized pathogen is neither transmitted by the seeds of its host nor persistent in the environment. It can be transmitted only from a living plant to another living plant. This life history trait limits the probability of introduction with its host plants, as weed introduction often occurs through seeds. The question of which traits prevent or facilitate invasions, during the several steps of dispersal and establishment, is of particular importance for fungal invasions in general. The *Microbotryum* case nevertheless shows that even the specialized, obligate fungus unable to disperse by wind (being mostly insect-borne) or seeds, or to persist in contaminated soil, can successfully invade a new continent, and multiple times independently. This case study further illustrates the importance of delimiting cryptic species in the native range and the usefulness of comprehensive sampling for performing

relevant population structure analyses and identifying the locations of invasion source populations. It also highlights the possibility of successful invasion without change in mode of reproduction and despite a strong bottleneck.

Conclusion and future prospects

In conclusion, fungal invasions are valuable yet often overlooked models to contribute to our view of biological invasions. In particular, studies on large data sets including many fungal tree pathogens have been successful at identifying the traits favouring fungal invasions, *for example,* fecundity and spore traits favouring dispersal, as well as mixed mating systems, with both clonal and sexual reproduction. Furthermore, the ecological and evolutionary processes favouring successful introductions could be revealed in several case studies, such as adaptation to higher temperatures, increased aggressiveness, acquisition of novel virulences, phenotypic plasticity, novel host associations, and admixture under multiple introductions or hybridization. Sophisticated methods of population genetic inference have elucidated the dispersal pathways and native sources of several fungal invasions, pinpointing in particular host tracking for pathogens and the possibility of invasions even with low genetic variability*, that is,* following strong bottlenecks or loss of sex. The rapid adaptation of invasive fungi, in particular of crop pathogens, often despite limited standing variation, is likely facilitated by their huge population sizes and the large genomic targets of selection, where any change within an effector gene for instance may allow escaping host recognition. Altogether, these insights broaden our understanding of biological invasions in eukaryotes.

Fungal invasion biology is a dynamic, recently emergent area of research. Many fungal invasions, possibly even the majority that exist, remain to be discovered, and there is much yet to be learned. The use of high-throughput sequencing technologies and the availability of universal primers for the sequencing of a single fungal barcode, the ITS (White *et al.* 1990), are greatly advancing our understanding of fungal diversity (Peay *et al.* 2008; Talbot *et al.* 2014) and increasing the power to detect invasive fungi, which are often concealed within the soil or in plant tissues. The use of sequencing technologies and substantial sampling efforts should shed light on the processes underlying invasions by nonpathogenic fungi, about which we know very little, despite their tremendous importance in ecosystem functioning, as decomposers for example. More generally, technological advances in the characterization of fungal populations and progress in inference methods (Stukenbrock & McDonald 2008; Neher 2013) are creating new, unprecedented opportunities to study fungal ecology and evolution, and fungi are proving

to be excellent organisms for challenging the universality of expectations derived from plants and animals (Gladieux *et al.* 2014).

Progress has recently been made towards determining the origins and pathways of fungal invasions, but a more comprehensive understanding of the processes underlying fungal invasions will require integrating population and comparative genomics within an experimental frameworks as has been applied to plant models. Phenotypic differences between introduced and native populations occur under nonequilibrium demographic conditions and may be affected by chance events, evolutionary history and adaptive responses to divergent selection (Taylor & Keller 2007). Experimental comparisons of species from their native and introduced ranges, testing hypotheses of adaptation against null models of neutral phenotypic evolution, should make it possible to determine whether the causal mechanisms of phenotypic change are adaptive. It is now increasingly feasible to push such studies further with population genetics to identify potentially adaptive genomic features and then to make use of bioinformatics and reverse genetics tools to connect selected genomic features to important phenotypes. Full-genome sequencing of multiple strains is now routinely carried out in fungi. Improvements in the functional annotation of fungal genomes, and their amenability to experimental manipulation, are making it possible to implement a genuine 'reverse ecology' approach involving the identification of genes with unusual patterns of evolution and environmental parameters consistent with the function of the genes and then testing hypotheses experimentally (Ellison *et al.* 2011, 2014).

Adaptive evolution in invasive fungi may pose a major threat to plant and animal health, but it also provides us with interesting opportunities to improve our understanding of the evolution of biotic interactions (McDonald 2010). We must not waste this opportunity to learn from the natural experiments represented by fungal invasions if we are to develop sustainable regulatory mechanisms based on evolutionary models, using predictive frameworks to control the pace of pathogen evolution in natural and agro-ecosystems (Parker & Gilbert 2004; Thrall *et al.* 2007, 2011).

Acknowledgements

This work was funded by ANR grants 07-BDIV-003-Emerfundis and 12-ADAP-0009-02-GANDALF, the National Science Foundation grant DEB-1115765, a Marie Curie postdoctoral fellowship awarded to PG (FP7-PEOPLE-2010-IOF-No.273086) and the ERC starting grant GenomeFun 309403 awarded to TG. We acknowledge the REID for funding attendance at

meetings and workshops. We thank Marie-Laure Desprez-Loustau for her considerable assistance with this manuscript and for discussions. We thank all the people who have helped with completing our *Microbotryum* collection from the UK: Deborah Charlesworth, Jacky Muscott, Clare Gachon and Damien Hicks, and we thank Britta Büker and Thomas Meagher for identifying locations in Massachusetts.

P.G., C.D. and T.G. wrote the manuscript, with contribution from M.E.H. and M.R. All authors read and approved the manuscript. C.D. and T.G. compiled data presented in Table S1. M.R. prepared Fig. 1. J.C. organized field trip in the UK and collected samples assisted by P.G. M.E.H. and T.G. provided the remaining samples. A.S. carried out molecular genotyping. A.F. analysed data, supervised by T.G. and P.G.

Data accessibility

Information on sampling locations as well as the microsatellite genotypic data has been deposited at Dryad: doi:10.5061/dryad.k5786.

Supplementary Information

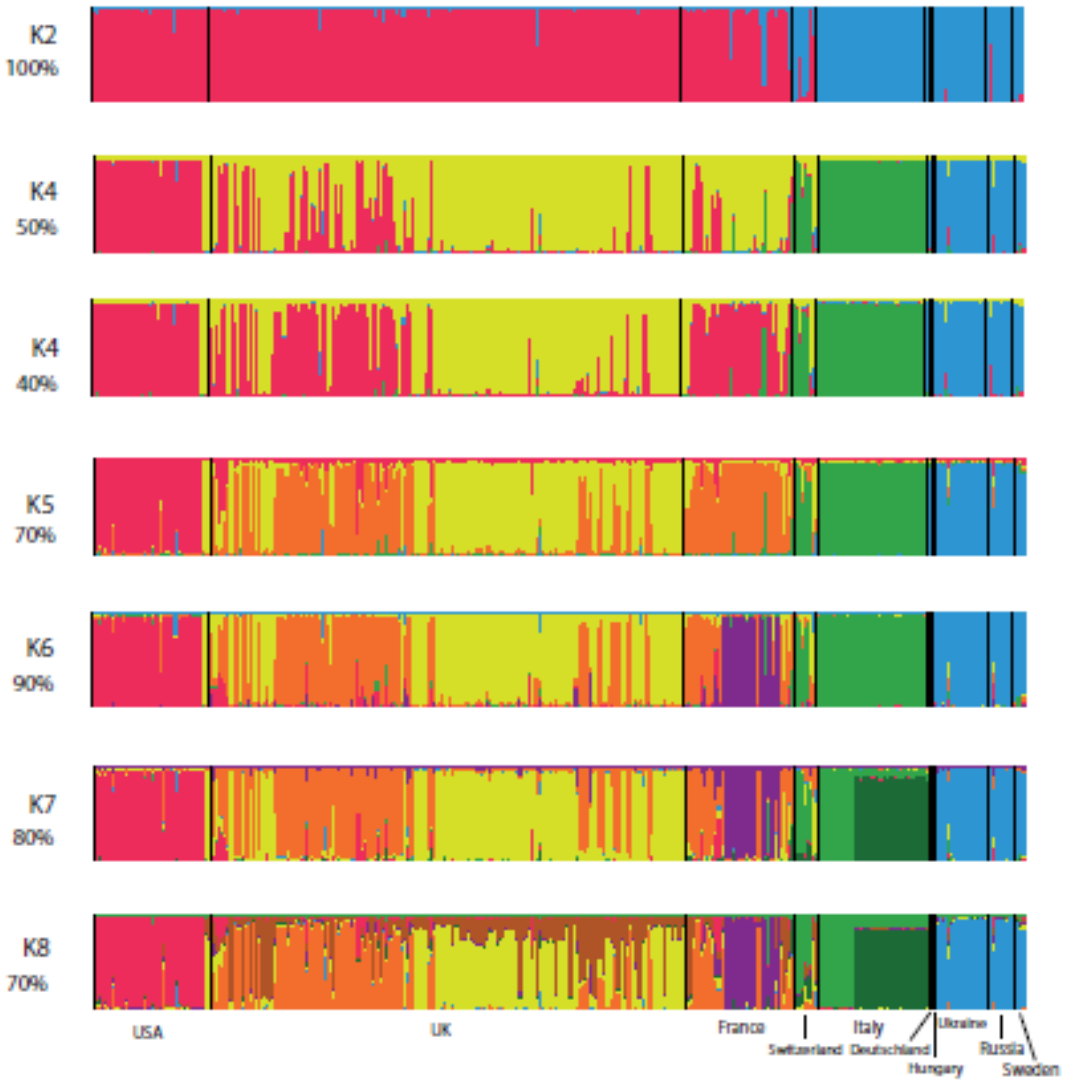


Fig. S1 Population structure of fungus Microbotryum lychnidis-dioicae collected on Silene latifolia (N = 328 individuals), inferred with Structure, for K = 2 to K = 8. Each vertical line represents an individual. Individuals are grouped on the horizontal

References

Agosta SJ, Klemens JA (2008) Ecological fitting by phenotypically flexible genotypes: implications for species associations, community assembly and evolution. *Ecology Letters*, 11, 1123–1134.

Ahmed S, de Labrouhe DT, Delmotte F (2012) Emerging virulence arising from hybridisation facilitated by multiple introductions of the sunflower downy mildew pathogen Plasmopara halstedii. *Fungal Genetics and Biology*, 49, 847–855.

Ali S, Leconte M, Walker A-S, Enjalbert J, de Vallavieille-Pope C (2010) Reduction in the sex ability of worldwide clonal populations of Puccinia striiformis f.sp tritici. *Fungal Genetics and Biology*, 47, 828–838.

Ali S, Gladieux P, Leconte M et al. (2014) Origin, migration routes and worldwide population genetic structure of the wheat yellow rust Puccinia striiformis f. sp tritici. *PLoS Pathogens*, 10, e1003903.

Antonovics J, Hood ME, Thrall PH, Abrams JY, Duthie GM (2003) Herbarium studies on the distribution of anther-smut fungus (Microbotryum violaceum) and Silene species (Caryophyllaceae) in the eastern United States. *American Journal of Botany*, 90, 1522–1531.

Avise JC, Wollenberg K (1997) Phylogenetics and the origin of species. *Proceedings of the National Academy of Sciences*, USA, 94, 7748–7755.

Aylor DE (2003) Spread of plant disease on a continental scale: role of aerial dispersal of pathogens. *Ecology*, 84, 1989–1997.

Bagley SJ, Orlovich DA (2004) Genet size and distribution of Amanita muscaria in a suburban park, Dunedin, New Zealand. *New Zealand Journal of Botany*, 42, 939–947.

Barrès B, Halkett F, Dutech C et al. (2008) Genetic structure of the poplar rust fungus Melampsora larici-populina: evidence for isolation by distance in Europe and recent founder effects overseas. *Infection, Genetics and Evolution*, 8, 577–587.

Barrès B, Carlier J, Seguin M et al. (2012) Understanding the recent colonization history of a plant pathogenic fungus using population genetic tools and Approximate Bayesian Computation. i, 109, 269–279.

Barrett LG, Thrall PH, Burdon JJ, Linde CC (2008) Life history determines genetic structure and evolutionary potential of host-parasite interactions. *Trends in Ecology & Evolution*, 23, 678–685.

Barton N (2010) Understanding adaptation in large populations. *PLoS Genetics*, 6, e1000987.

Bazin E, Mathe-Hubert H, Facon B, Carlier J, Ravigne V (2014) The effect of mating system on invasiveness: some genetic load may be advantageous when invading new environments. *Biological Invasions*, 16, 875–886.

Berger L, Speare R, Daszak P et al. (1998) Chytridiomycosis causes amphibian mortality associated with population declines in the rain forests of Australia and Central America. *Proceedings of the National Academy of Sciences*, USA, 95, 9031–9036.

Bernasconi G, Antonovics J, Biere A et al. (2009) Silene as a model system in ecology and evolution. *Heredity*, 103, 5–14.

Bovers M, Hagen F, Kuramae EE, Boekhout T (2008) Six monophyletic lineages identified within Cryptococcus neoformans and Cryptococcus gattii by multi-locus sequence typing. *Fungal Genetics and Biology*, 45, 400–421.

Brasier CM (2001) Rapid evolution of introduced plant pathogens via interspecific hybridization. *BioScience*, 51, 123–133.

Brown JKM (1994) Chance and selection in the evolution of barley mildew. *Trends in Microbiology*, 2, 470–475.

Brown JKM, Hovmøller MS (2002) Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science*, 297, 537–541.

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10, 195–205.

Chen W-J, Delmotte F, Cervera SR et al. (2007) At least two origins of fungicide resistance in grapevine downy mildew populations. *Applied and Environmental Microbiology*, 73, 5162–5172.

Couch BC, Fudal I, Lebrun MH et al. (2005) Origins of host-specific populations of the blast pathogen Magnaporthe oryzae in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics*, 170, 613–630.

Daverdin G, Rouxel T, Gout L et al. (2012) Genome structure and reproductive behaviour influence the evolutionary potential of a fungal phytopathogen. *PLoS Pathogens*, 8, e1003020.

De Queiroz K (2007) Toward an integrated system of clade names. *Systematic Biology*, 56, 956–974.

Desprez-Loustau M-L, Robin C, Buee M et al. (2007) The fungal dimension of biological invasions. *Trends in Ecology & Evolution*, 22, 472–480.

Desprez-Loustau M-L, Courtecuisse R, Robin C et al. (2010) Species diversity and drivers of spread of alien fungi (sensu lato) in Europe with a particular focus on France. *Biological Invasions*, 12, 157–172.

Díez J (2005) Invasion biology of Australian ectomycorrhizal fungi introduced with eucalypt plantations into the Iberian Peninsula. *Biological Invasions*, 7, 3–15.

Dilmaghani A, Gladieux P, Gout L et al. (2012) Migration patterns and changes in population biology associated with the worldwide spread of the oilseed rape pathogen Leptosphaeria maculans. *Molecular Ecology*, 21, 2519–2533.

Dobrowolski MP, Tommerup IC, Shearer BL, O'Brien PA (2003) Three clonal lineages of Phytophthora cinnamomi in Australia revealed by microsatellites. *Phytopathology*, 93, 695–704.

Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics*, 11, 539–548.

Duan X, Tellier A, Wan A et al. (2010) Puccinia striiformis f.sp tritici presents high diversity and recombination in the over-summering zone of Gansu, China. *Mycologia*, 102, 44–53.

Dutech C, Fabreguettes O, Capdevielle X, Robin C (2010) Multiple introductions of divergent genetic lineages in an invasive fungal pathogen, Cryphonectria parasitica, in France. *Heredity*, 105, 220–228.

Dutech C, Barres B, Bridier J et al. (2012) The chestnut blight fungus world tour: successive introduction events from diverse origins in an invasive plant fungal pathogen. *Molecular Ecology*, 21, 3931–3946.

Ellison CE, Hall C, Kowbel D et al. (2011) Population genomics and local adaptation in wild isolates of a model microbial eukaryote. *Proceedings of the National Academy of Sciences*, USA, 108, 2831–2836.

Ellison CE, Kowbel D, Glass NL, Taylor JW, Brem RB (2014) Discovering functions of unannotated genes from a transcriptome survey of wild fungal isolates. *MBio* 5, e01046–13.

Enjalbert J, Duan X, Leconte M, Hovmøller MS, De Vallavieille-Pope C (2005) Genetic evidence of local adaptation of wheat yellow rust (Puccinia striiformis f. sp tritici) within France. *Molecular Ecology*, 14, 2065–2073.

Facon B, Genton BJ, Shykoff J et al. (2006) A general eco-evolutionary framework for understanding bioinvasions. *Trends in Ecology & Evolution*, 21, 130–135.

Farrer RA, Weinert LA, Bielby J et al. (2011) Multiple emergences of genetically diverse amphibian-infecting chytrids include a globalized hypervirulent recombinant lineage. *Proceedings of the National Academy of Sciences*, USA, 108, 18732–18736.

Farrer RA, Henk DA, Garner TWJ et al. (2013) Chromosomal copy number variation, selection and uneven rates of recombination reveal cryptic genome diversity linked to pathogenicity. *PLoS Genetics*, 9, e1003703.

Fisher MC, Garner TWJ, Walker SF (2009) Global emergence of Batrachochytrium dendrobatidis and amphibian chytridiomycosis in space, time, and host. *Annual Review of Microbiology*, 63, 291–310.

Fisher MC, Henk DA, Briggs CJ et al. (2012) Emerging fungal threats to animal, plant and ecosystem health. *Nature*, 484, 186–194.

Fontaine MC, Gladieux P, Hood ME, Giraud T (2013) History of the invasion of the anther smut pathogen on Silene latifolia in North America. *New Phytologist*, 198, 946–956.

Frenkel O, Peever TL, Chilvers MI et al. (2010) Ecological genetic divergence of the fungal pathogen Didymella rabiei on sympatric wild and domesticated Cicer spp. (Chickpea). *Applied and Environmental Microbiology*, 76, 30–39.

Fudal I, Ross S, Brun H et al. (2009) Repeat-Induced Point Mutation (RIP) as an alternative mechanism of evolution toward virulence in Leptosphaeria maculans. *Molecular Plant-Microbe Interactions*, 22, 932–941.

Garud NR, Messer PW, Buzbas EO, Petrov DA (2013) Soft selective sweeps are the primary mode of recent adaptation in Drosophila melanogaster. arXiv preprint arXiv:1303.0906.

Gilbert GS, Parker IM (2010) Rapid evolution in a plant-pathogen interaction and the consequences for introduced host species. *Evolutionary Applications*, 3, 144–156.

Gilbert GS, Webb CO (2007) Phylogenetic signal in plant pathogen-host range. *Proceedings of the National Academy of Sciences*, USA, 104, 4979–4983.

Giraud T, Refregier G, Le Gac M, de Vienne DM, Hood ME (2008a) Speciation in fungi. *Fungal Genetics and Biology*, 45, 791–802.

Giraud T, Yockteng R, Lopez-Villavicencio M, Refregier G, Hood ME (2008b) Mating system of the anther smut fungus Microbotryum violaceum: selfing under heterothallism. *Eukaryotic Cell*, 7, 765–775.

Giraud T, Gladieux P, Gavrilets S (2010) Linking the emergence of fungal plant diseases with ecological speciation. *Trends in Ecology & Evolution*, 25, 387–395.

Gladieux P, Zhang XG, Afoufa-Bastien D et al. (2008) On the origin and spread of the scab disease of apple: out of Central Asia. *PLoS ONE*, 3, e1455.

Gladieux P, Zhang XG, Roldan-Ruiz I et al. (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.

Gladieux P, Guerin F, Giraud T et al. (2011) Emergence of novel fungal pathogens by ecological speciation: importance of the reduced viability of immigrants. *Molecular Ecology*, 20, 4521–4532.

Gladieux P, Devier B, Aguileta G, Cruaud C, Giraud T (2013) Purifying selection after episodes of recurrent adaptive diversification in fungal pathogens. *Infection Genetics and Evolution*, 17, 123–131.

Gladieux P, Ropars J, Badouin H et al. (2014) Fungal evolutionary genomics provides insight into the mechanisms of adaptive divergence in eukaryotes. *Molecular Ecology*, 23, 753–773.

Gonthier P, Nicolotti G, Linzer R, Guglielmo F, Garbelotto M (2007) Invasion of European pine stands by a North American forest pathogen and its hybridization with a native interfertile taxon. *Molecular Ecology*, 16, 1389–1400.

Goodwin SB, Cohen BA, Fry WE (1994) Panglobal distribution of a single clonal lineage of the Irish potato famine fungus. *Proceedings of the National Academy of Sciences*, USA, 91, 11591–11595.

Goss EM, Carbone I, Grunwald NJ (2009) Ancient isolation and independent evolution of the three clonal lineages of the exotic sudden oak death pathogen Phytophthora ramorum. *Molecular Ecology*, 18, 1161–1174.

Goss EM, Tabima JF, Cooke DEL et al. (2014) The Irish potato famine pathogen Phytophthora infestans originated in central Mexico rather than the Andes. *Proceedings of the National Academy of Sciences*, USA, 201401884.

Gostinčar C, Grube M, De Hoog S, Zalar P, Gunde-Cimerman N (2010) Extremotolerance in fungi: evolution on the edge. *FEMS Microbiology Ecology*, 71, 2–11.

Gout L, Kuhn ML, Vincenot L et al. (2007) Genome structure impacts molecular evolution at the AvrLm1 avirulence locus of the plant pathogen Leptosphaeria maculans. *Environmental Microbiology*, 9, 2978–2992.

Green J, Bohannan BJM (2006) Spatial scaling of microbial biodiversity. *Trends in Ecology & Evolution*, 21, 501–507.

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.

Hansen EM, Reeser PW, Sutton W (2012) Phytophthora beyond agriculture. *Annual Review of Phytopathology*, 50, 359–378.

Hayes KR, Barry SC (2008) Are there any consistent predictors of invasion success? *Biological Invasions*, 10, 483–506.

Henk DA, Eagle CE, Brown K et al. (2011) Speciation despite globally overlapping distributions in Penicillium chrysogenum: the population genetics of Alexander Fleming's lucky fungus. *Molecular Ecology*, 20, 4288–4301.

Hibbett DS, Taylor JW (2013) Fungal systematics: is a new age of enlightenment at hand? *Nature Reviews Microbiology*, 11, 129–133.

Hood ME, Mena-Ali JI, Gibson AK et al. (2010) Distribution of the anther-smut pathogen Microbotryum on species of the Caryophyllaceae. *New Phytologist*, 187, 217–229.

Hovmøller MS, Yahyaoui AH, Milus EA, Justesen AF (2008) Rapid global spread of two aggressive strains of a wheat rust fungus. *Molecular Ecology*, 17, 3818–3826.

Hulme PE, Barrett SCH (2013) Integrating trait-and niche-based approaches to assess contemporary evolution in alien plant species. *Journal of Ecology*, 101, 68–77.

Ivors K, Garbelotto M, Vries IDE et al. (2006) Microsatellite markers identify three lineages of Phytophthora ramorum in US nurseries, yet single lineages in US forest and European nursery populations. *Molecular Ecology*, 15, 1493–1505.

Jairus T, Mpumba R, Chinoya S, Tedersoo L (2011) Invasion potential and host shifts of Australian and African ectomycorrhizal fungi in mixed eucalypt plantations. *New Phytologist*, 192, 179–187.

Jezic M, Krstin L, Rigling D, Curkovic-Perica M (2012) High diversity in populations of the introduced plant pathogen, Cryphonectria parasitica, due to encounters between genetically divergent genotypes. *Molecular Ecology*, 21, 87–99.

Karasov T, Messer PW, Petrov DA (2010) Evidence that adaptation in Drosophila is not limited by mutation at single sites. *PLoS Genetics*, 6, e1000924.

Kasuga T, White TJ, Taylor JW (2002) Estimation of nucleotide substitution rates in eurotiomycete fungi. *Molecular Biology and Evolution*, 19, 2318–2324.

Keller SR, Sowell DR, Neiman M, Wolfe LM, Taylor DR (2009) Adaptation and colonization history affect the evolution of clines in two introduced species. *New Phytologist*, 183, 678–690.

Keller SR, Gilbert KJ, Fields PD, Taylor DR (2012) Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, Silene latifolia. *Molecular Ecology*, 21, 4721–4734.

Kennedy P (2010) Ectomycorrhizal fungi and interspecific competition: species interactions, community structure, coexistence mechanisms, and future research directions. *New Phytologist*, 187, 895–910.

Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution*, 16, 199–204.

Le Gac M, Giraud T (2008) Existence of a pattern of reproductive character displacement in Homobasidiomycota but not in Ascomycota. *Journal of Evolutionary Biology*, 21, 761–772.

Le Gac M, Hood ME, Fournier E, Giraud T (2007) Phylogenetic evidence of host-specific cryptic species in the anther smut fungus. *Evolution*, 61, 15–26.

Le Van A, Gladieux P, Lemaire C et al. (2012) Evolution of pathogenicity traits in the apple scab fungal pathogen in response to the domestication of its host. *Evolutionary Applications*, 5, 694–704.

Lewontin RC (1974) *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

Liebhold AM, Brockerhoff EG, Garrett LJ, Parke JL, Britton KO (2012) Live plant imports: the major pathway for forest insect and pathogen invasions of the US. *Frontiers in Ecology and the Environment*, 10, 135–143.

Lin X, Litvintseva AP, Nielsen K et al. (2007) αADα hybrids of Cryptococcus neoformans: evidence of same-sex mating in nature and hybrid fitness. *PLoS Genetics*, 3, e186.

Lutz M, Göker M, Piatek M et al. (2005) Anther smuts of Caryophyllaceae: molecular characters indicate host-dependent species delimitation. *Mycological Progress*, 4, 225–238.

MacArthur RH, Wilson EO (2001) The theory of island biogeography. *Princeton Landmarks in Biology*, 224.

Markell SG, Milus EA (2008) Emergence of a novel population of Puccinia striiformis f. sp. tritici in eastern United States. *Phytopathology*, 98, 632–639.

Martel A, Spitzen-van der Sluijs A, Blooi M et al. (2013) Batrachochytrium salamandrivorans sp. nov. causes lethal chytridiomycosis in amphibians. *Proceedings of the National Academy of Sciences*, USA, 110, 15325–15329.

Martiny JBH, Bohannan BJM, Brown JH et al. (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4, 102–112.

Maynard-Smith J, Smith NH, O'Rourke M, Spratt BG (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, 23, 23–35.

Mayr E (1965) The nature of colonization of birds. In: *The Genetics of Colonizing Species* (eds Baker HG, Stebbins GL), pp. 29–44. Academic Press, New York, London.

Mboup M, Leconte M, De Vallavieille Pope C, Enjalbert J (2007) Evidence of genetic recombination in Chinese wheat yellow rust populations. *Phytopathology*, 97, S72–S73.

Mboup M, Leconte M, Gautier A et al. (2009) Evidence of genetic recombination in wheat yellow rust populations of a Chinese oversummering area. *Fungal Genetics and Biology*, 46, 299–307.

Mboup M, Bahri B, Leconte M et al. (2012) Genetic structure and local adaptation of European wheat yellow rust populations: the role of temperature-specific adaptation. *Evolutionary Applications*, 5, 341–352.

McDonald B (2010) How can we achieve durable disease resistance in agricultural ecosystems? *New Phytologist*, 185, 3–5.

McDonald BA (2014) Using dynamic diversity to achieve durable disease resistance in agricultural ecosystems. *Tropical Plant Pathology*, 39, 191–196.

McDonald BA, Linde C (2002) Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology*, 40, 349–379.

McMahon TA, Brannelly LA, Chatfield MWH et al. (2013) Chytrid fungus Batrachochytrium dendrobatidis has nonamphibian hosts and releases chemicals that cause pathology in the absence of infection. *Proceedings of the National Academy of Sciences*, USA, 110, 210–215.

Milus EA, Kristensen K, Hovmøller MS (2009) Evidence for increased aggressiveness in a recent widespread strain of Puccinia striiformis f. sp. tritici causing stripe rust of wheat. *Phytopathology*, 99, 89–94.

Montarry J, Andrivon D, Glais I et al. (2010) Microsatellite markers reveal two admixed genetic groups and an ongoing displacement within the French population of the invasive plant pathogen Phytophthora infestans. *Molecular Ecology*, 19, 1965–1977.

Morgan JAT, Vredenburg VT, Rachowicz LJ et al. (2007) Population genetics of the frog-killing fungus Batrachochytrium dendrobatidis. *Proceedings of the National Academy of Sciences*, USA, 104, 13845–13850.

Mueller GM, Strack BA (1992) Evidence for a mycorrhizal host shift during migration of Laccaria trichodermophora and other agarics into neotropical oak forests. *Mycotaxon*, 45, 249–256.

Mundt CC, Wallace LD, Allen TW et al. (2013) Initial epidemic area is strongly associated with the yearly extent of soybean rust spread in North America. *Biological Invasions*, 15, 1431–1438.

Munkacsi AB, Stoxen S, May G (2008) Ustilago maydis populations tracked maize through domestication and cultivation in the Americas. *Proceedings of the Royal Society B-Biological Sciences*, 275, 1037–1046.

Naciri-Graven Y, Goudet J (2003) The additive genetic variance after bottlenecks is affected by the number of loci involved in epistatic interactions. *Evolution*, 57, 706–716.

Neher RA (2013) Genetic draft, selective interference, and population genetics of rapid adaptation. arXiv preprint arXiv:1302.1148.

Palm ME (2001) Systematics and the impact of invasive fungi on agriculture in the United States. *BioScience*, 51, 141–147.

Palm ME, Rossman AY (2003) Invasion pathways of terrestrial plant-inhabiting fungi. In: *Bioinvasions: Pathways, Vectors, and Management Strategies* (eds Ruiz GM, Carlton JT), pp. 31–43. Island Press, New York, New York.

Paoletti M, Buck KW, Brasier CM (2006) Selective acquisition of novel mating type and vegetative incompatibility genes via interspecies gene transfer in the globally invading eukaryote Ophiostoma novo-ulmi. *Molecular Ecology*, 15, 249–262.

Parker IM, Gilbert GS (2004) The evolutionary ecology of novel plant-pathogen interactions. Annual Review of Ecology, *Evolution and Systematics*, 35, 675–700.

Peay KG, Bruns TD, Kennedy PG, Bergemann SE, Garbelotto M (2007) A strong species–area relationship for eukaryotic soil microbes: island size matters for ectomycorrhizal fungi. *Ecology Letters*, 10, 470–480.

Peay KG, Kennedy PG, Bruns TD (2008) Fungal community ecology: a hybrid beast with a molecular master. *BioScience*, 58, 799–810.

Peay KG, Bidartondo MI, Arnold AE (2010) Not every fungus is everywhere: scaling to the biogeography of fungal-plant interactions across roots, shoots and ecosystems. *New Phytologist*, 185, 878–882.

Peay KG, Schubert MG, Nguyen NH, Bruns TD (2012) Measuring ectomycorrhizal fungal dispersal: macroecological patterns driven by microscopic propagules. *Molecular Ecology*, 21, 4122–4136.

Philibert A, Desprez-Loustau M-L, Fabre B et al. (2011) Predicting invasion success of forest pathogenic fungi from species traits. *Journal of Applied Ecology*, 48, 1381–1390.

Prentis PJ, Wilson JRU, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in invasive species. *Trends in Plant Science*, 13, 288–294.

Pringle A, Vellinga EC (2006) Last chance to know? Using literature to explore the biogeography and invasion biology of the death cap mushroom Amanita phalloides (Vaill. ex Fr.: Fr.) Link. *Biological Invasions*, 8, 1131–1144.

Pringle A, Baker D, Platt J et al. (2005) Cryptic speciation in the cosmopolitan and clonal human pathogenic fungus Aspergillus fumigatus. *Evolution*, 59, 1886–1899.

Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, 20, R208–R215.

Prussin AJ, Szanyi NA, Welling PI, Ross SD, Schmale III DG (2014) Estimating the production and release of ascospores from a field-scale source of Fusarium graminearum inoculum. *Plant Disease*, 98, 497–503.

Pyšek P, Jarošík V, Hulme PE et al. (2010) Disentangling the role of environmental and human pressures on biological invasions across Europe. *Proceedings of the National Academy of Sciences*, USA, 107, 12157–12162.

Raboin L-M, Selvi A, Oliveira KM et al. (2007) Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen Ustilago scitaminea. *Fungal Genetics and Biology*, 44, 64–76.

Rafiqi M, Ellis JG, Ludowici VA, Hardham AR, Dodds PN (2012) Challenges and progress towards understanding the role of effectors in plant–fungal interactions. *Current Opinion in Plant Biology*, 15, 477–482.

Refregier G, Le Gac M, Jabbour F et al. (2008) Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. *BMC Evolutionary Biology*, 8, 100.

Rizzo DM (2005) Exotic species and fungi: interactions with fungal, plant and animal communities. *The Fungal Community: Its Organization and Role in the Ecosystem*, pp. 857–880. CRC Press, Boca Raton, Florida.

Robert VA, Casadevall A (2009) Vertebrate endothermy restricts most fungi as potential pathogens. *Journal of Infectious Diseases*, 200, 1623–1626.

Roper M, Ellison C, Taylor JW, Glass NL (2011) Nuclear and genome dynamics in multinucleate ascomycete fungi. *Current Biology*, 21, R786–R793.

Rouxel M, Mestre P, Comont G et al. (2013) Phylogenetic and experimental evidence for host-specialized cryptic species in a biotrophic oomycete. *New Phytologist*, 197, 251–263.

Sakai AK, Allendorf FW, Holt JS et al. (2001) The population biology of invasive species. *Annual Review of Ecology & Systematics*, 32, 305–332.

Sakalidis ML, Slippers B, Wingfield BD, Hardy GESJ, Burgess TI (2013) The challenge of understanding the origin, pathways and extent of fungal invasions: global populations of the Neofusicoccum parvum–N. ribis species complex. *Diversity and Distributions*, 19, 873–883.

Saleh D, Milazzo J, Adreit H, Tharreau D, Fournier E (2012a) Asexual reproduction induces a rapid and permanent loss of sexual reproduction capacity in the rice fungal pathogen Magnaporthe oryzae: results of in vitro experimental evolution assays. *BMC Evolutionary Biology*, 12, 42.

Saleh D, Xu P, Shen Y et al. (2012b) Sex at the origin: an Asian population of the rice blast fungus Magnaporthe oryzae reproduces sexually. *Molecular Ecology*, 21, 1330–1344.

Sato H, Tsujino R, Kurita K, Yokoyama K, Agata K (2012) Modelling the global distribution of fungal species: new insights into microbial cosmopolitanism. *Molecular Ecology*, 21, 5599–5612.

Schloegel LM, Toledo LF, Longcore JE et al. (2012) Novel, panzootic and hybrid genotypes of amphibian chytridiomycosis associated with the bullfrog trade. *Molecular Ecology*, 21, 5162–5177.

Schulze-Lefert P, Panstruga R (2011) A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. *Trends in Plant Science*, 16, 117–125.

Schürch S, Linde CC, Knogge W, Jackson LF, McDonald BA (2004) Molecular population genetic analysis differentiates two virulence mechanisms of the fungal avirulence gene NIP1. *Molecular Plant-Microbe Interactions*, 17, 1114–1125.

Schwartz MW, Hoeksema JD, Gehring CA et al. (2006) The promise and the potential consequences of the global transport of mycorrhizal fungal inoculum. *Ecology Letters*, 9, 501–515.

Singh RP, Hodson DP, Huerta-Espino J et al. (2011) The emergence of Ug99 races of the stem rust fungus is a threat to world wheat production. *Annual Review of Phytopathology*, 49, 465–481.

Smart CD, Fry WE (2001) Invasions by the late blight pathogen: renewed sex and enhanced fitness. *Biological Invasions*, 3, 235–243.

Stajich JE, Berbee ML, Blackwell M et al. (2009) The fungi. *Current Biology*, 19, R840–R845.

Stewart JE, Timmer LW, Lawrence CB, Pryor BM, Peever TL (2014) Discord between morphological and phylogenetic species boundaries: incomplete lineage sorting and recombination results in fuzzy species boundaries in an asexual fungal pathogen. *BMC Evolutionary Biology*, 14, 38.

Strange RN, Scott PR (2005) Plant disease: a threat to global food security. *Annual Review of Phytopathology*, 43, 83–116.

Stukenbrock EH, McDonald BA (2008) The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, 46, 75–100.

Stukenbrock EH, Banke S, Javan-Nikkhah M, McDonald BA (2007) Origin and domestication of the fungal wheat pathogen I via sympatric speciation. *Molecular Biology and Evolution*, 24, 398–411.

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. *Proceedings of the National Academy of Sciences*, USA, 109, 10954–10959.

Swann EC, Frieders EM, McLaughlin DJ (1999) Microbotryum, Kriegeria and the changing paradigm in basidiomycete classification. *Mycologia*, 91, 51–66.

Talbot JM, Bruns TD, Taylor JW et al. (2014) Endemism and functional convergence across the North American soil mycobiome. *Proceedings of the National Academy of Sciences*, USA, 111, 6341–6346.

Taylor JW, Fisher MC (2003) Fungal multilocus sequence typing—it's not just for bacteria. *Current Opinion in Microbiology*, 6, 351–356.

Taylor DR, Keller SR (2007) Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution*, 61, 334–345.

Taylor JW, Jacobson D, Fisher M (1999) The evolution of asexual fungi: reproduction, speciation and classification. *Annual Review of Phytopathology*, 37, 197–246.

Taylor JW, Jacobson DJ, Kroken S et al. (2000) Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology*, 31, 21–32.

Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D (2006) Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 361, 1947–1963.

Tedersoo L, Bahram M, Toots M et al. (2012) Towards global patterns in the diversity and community structure of ectomycorrhizal fungi. *Molecular Ecology*, 21, 4160–4170.

Thorsten Lumbsch H, Buchanan PK, May TW, Mueller GM (2008) Phylogeography and biogeography of fungi. *Mycological Research*, 112, 423–424.

Thrall PH, Hochberg ME, Burdon JJ, Bever JD (2007) Coevolution of symbiotic mutualists and parasites in a community context. *Trends in Ecology & Evolution*, 22, 120–126.

Thrall PH, Oakeshott JG, Fitt G et al. (2011) Evolution in agriculture: the application of evolutionary approaches to the management of biotic interactions in agro-ecosystems. *Evolutionary Applications*, 4, 200–215.

Valade R (2012) Potentiel évolutif et adaptation des populations de l'agent du mildiou de la laitue, Bremia lactucae, face aux pressions de sélection de la plante hôte, Lactuca sativa.

Vellinga EC, Wolfe BE, Pringle A (2009) Global patterns of ectomycorrhizal introductions. *New Phytologist*, 181, 960–973.

Vercken E, Fontaine MC, Gladieux P et al. (2010) Glacial refugia in pathogens: European genetic structure of anther smut pathogens on Silene latifolia and Silene dioica. *PLoS Pathogens*, 6, e1001229.

de Vienne DM, Hood ME, Giraud T (2009) Phylogenetic determinants of potential host shifts in fungal pathogens. *Journal of Evolutionary Biology*, 22, 2532–2541.

de Vienne DM, Giraud T, Gouyon P-H (2013a) Lineage selection and the maintenance of sex. *PLoS ONE*, 8, e66906.

de Vienne DM, Refregier G, Lopez-Villavicencio M et al. (2013b) Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytologist*, 198, 347–385.

White TJ, Bruns T, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR Protocols: A Guide to Methods and Applications, 18, 315–322.

Wolfe BE, Richard F, Cross HB, Pringle A (2010) Distribution and abundance of the introduced ectomycorrhizal fungus Amanita phalloides in North America. *New Phytologist*, 185, 803–816.

Zaffarano PL, McDonald BA, Linde CC (2009) Phylogeographical analyses reveal global migration patterns of the barley scald pathogen Rhynchosporium secalis. *Molecular Ecology*, 18, 279–293.

Title: Inter-specific hybridizations in apple trees and host-pathogen coevolution

Key words: population genetics, *Malus*, gene flow, co-evolution, host-pathogen

Abstract: In the first part of my thesis project, I studied the evolutive history of apple tree species in Europe, including interspecific hybridizations. Analyses of whole genome data confirmed that the progenitor species of the cultivated apple tree was *M. sieversii*, a Central Asia wild apple tree, but also that a high proportion of European cultivated varieties form a genetic group distinct from the wild species. These varieties show traces of introgressions from *M. sylvestris* and of population subdivision along an East-West axis. Microsatellite markers also showed that introgressions from the cultivated apple were also quite frequent in wild apple tree populations and thus threaten their genetic integrity. We found that introgression levels were correlated to anthropic activities of apple tree cultivation and gave rise to hybrids with no detectable reduced fitness on the traits measured. Our study of the European wild apple tree phylogeography allowed us to detect differentiated genetic groups resulting from the past climatic history of the planet and which should be considered as different evolutionary significant units in conservation.

In the second part of my thesis project, I studied coevolution between plant species and their pathogenic fungi in two different systems. I first compared the evolutive history of cultivated and wild apple trees in Central Asia and that of their scab pathogen, *Venturia inaequalis*. In the Kazakhstan Mountains, the cultivated apple tree has been reintroduced back from Europe during the last two centuries. This created a secondary contact zone, not only between the two apple tree species, but also between the agricultural and the wild types of *V. inaequalis*. While the invasion of natural populations by the cultivated apple trees still seems geographically limited, the agricultural pathogen is widespread in the forests and on the wild host trees. However, the number of hybrids in the pathogen was limited, probably because of intrinsic reproductive barriers since no ecological barriers were found. I also compared spatial genetic structures at the European scale in the plant *Silene latifolia* and its anther-smut pathogen *Microbotryum lychnidis-dioicae*. Our dataset included genotypes from the pathogens and the plants on which they were collected, and the population structures appeared remarquably congruent. Three phylogenetic groups were identified in these both species, corresponding with the temperate species range contraction-expansion cycles during the past glaciations. A substructure was identified in the pathogen suggesting the possibility of a more complex history.

Titre : Hybridations inter-spécifiques chez le pommier et co-évolution hôte-pathogène

Mots clés : génétique des populations, *Malus*, flux de gènes, co-évolution, hôte-pathogène

Résumé : Dans une première partie de mon projet de thèse, je me suis intéressée aux hybridations interspécifiques et à l'histoire évolutive des espèces de pommiers en Europe. L'analyse de génomes complets m'a permis de confirmer que l'ancêtre du pommier cultivé est bien le pommier sauvage asiatique *M. sieversii*, mais aussi que de nombreuses variétés cultivées originaires de l'Europe forment un groupe génétique distinct de cette espèce sauvage. Ces variétés montrent des traces d'introgressions par le pommier sauvage européen *M. sylvestris* et une structuration de populaton selon un axe est-ouest. Les introgressions depuis le pommier cultivé sont également abondantes dans les populations de pommier sauvage européen *M. sylvestris,* et menacent leur intégrité génétique. Les niveaux des introgressions étaient corrélés aux activités anthropiques d'exploitation des pommiers et les hybrides n'avaient pas de réduction détectable de valeur sélective sur les caractères mesurés. Notre étude de la phylogéographie du pommier sauvage européen dans l'ensemble de son aire de distribution nous a permis de détecter des groupes génétiques différenciés résultant de l'histoire climatique passée de la planète. Ces groupes représentent autant d'unités de conservation différentes.

Dans une seconde partie, j'ai étudié la coévolution entre espèces de plantes et espèces de champignons pathogènes dans deux systèmes différents. Je me suis tout d'abord intéressée à l'histoire évolutive combinée des pommiers cultivés et sauvages d'Asie centrale et du champignon causant la tavelure, *Venturia inaequalis*. Dans les montagnes du Kazakhstan, le pommier cultivé a été réintroduit au cours des deux derniers siècles, créant une zone de contact secondaire, non seulement entre *M. domestica* et *M. sieversii,* mais aussi entre les populations de *V. inaequalis* de types agricole et sauvage. Alors que l'invasion des populations sauvages par le pommier cultivé semble encore limité géographiquement, le pathogène de type agricole est largement répandu dans les forêts et sur l'hôte sauvage. Cependant, le nombre d'hybrides détectés chez le pathogène reste limité, probablement en raison de barrières intrinsèques à la reproduction, puisque les deux types de pathogènes infectent les mêmes arbres hôtes. Dans un second temps, j'ai comparé les structures génétiques et spatiales à l'échelle européenne de la plante *Silene latifolia* et de son pathogène *Microbotryum lychnidis-dioicae*. Notre jeu de données constitué du génotype d'un pathogène et de la plante qu'il infectait nous a permis de comparer la structure à très fine échelle et celle-ci apparait remarquablement congruente. Trois groupes phylogéographiques ont été identifiés chez les deux organismes, correspondant à l'histoire de contraction-expansion des aires de distribution des espèces tempérées lors de la dernière glaciation. Une structure génétique plus fine a également été détectée chez le pathogène, suggérant la possibilité d'une histoire plus complexe.