Université des sciences et techniques de Lille 1

Thèse de doctorat

Spécialité : Biologie évolutive

Mise en évidence des forces évolutives agissant au locus d'auto-incompatibilité chez *Arabidopsis halleri*

par

Violaine Llaurens

Thèse soutenue le 14 décembre 2007, devant le jury composé de :

Deborah Charlesworth, Professeur, Université d'Edimbourg	Rapporteur
Philippe Jarne, DR1-CNRS, CEFE, Montpellier	Rapporteur
Sylvain Glémin, MCF, ISEM, Montpellier	Examinateur
Olivier Hardy, CR, Université libre de Bruxelles	Examinateur
Xavier Vekemans, Professeur, Université de Lille1	Directeur de thèse

Remerciements :

Comme beaucoup de gens ont été présent et m'ont soutenu pendant ma thèse, et que je ne sais pas vraiment comment les remercier assez, je vais me contenter des dresser un petit bilan des faits marquants de ma thèse en quelques chiffres. Donc, pendant ma thèse, j'ai notamment :

- harcelé mon directeur de 2892 questions : merci à Xavier pour ses nombreux conseils scientifiques et sa patience
- adopté 2 autres directeurs : merci à Vincent et Sylvain (qui a abusé de sa position pour me traiter 108 fois de morue) pour leur soutien et leur aide précieuse
- trouvé 1 véritable coach sportif : merci à Laurent pour son humour de merde (dont je suis une fan inconditionnelle) et sa compréhension
- eu 3 fidèles compagnons de galère : merci à Kristen, Martin et Sandy pour les innombrables remontages de moral mutuels
- exploité 4 techniciennes de laboratoire : merci à Adeline (pour ses conseils psychologique et BMistique, et ses innombrables plaquettes de chocolats), Cécile, Anne-Cat et Pauline
- utilisé 3000 pots à la serre : merci à Eric, Claire, Nathalie, Angélique et Robert
- bu 2000 tasses de thé
- appris 4 nouveaux gros mots en italien : merda secca ! merci Valeria
- prononcé 1007 fois la phrase « j'y arriverai jamais » et trouvé 1 mec capable de supporter ces plaintes quotidiennes : merci Jérôme pour ta patience
- été aidé par 2 mathématiciens hors pair : merci à Fred et Etienne
- soulé 6 potes avec ma thèse : merci à Luc, Anne-Hélène, Eve, Juju, Pauline et Vincent
- trainé 6 autres potes sans défense dans le ch'nord : merci à Maud, Granmou, Laura, Amélie, Louise, Mathieu d'être venu me voir
- assisté à 1 chorégraphie exceptionnelle du parking des anges : merci Fafa
- appris à me frayer un chemin dans la jungle administrative : merci à Sandrine et Michèle, nos 2 secrétaires de choc
- voyagé dans 6 pays différents (Pologne, Angleterre, Pays-bas, Ecosse, Nouvelle-Zélande, Suède)
- déménagé 4 fois
- été soutenu par 8 membres de ma famille : merci à Gilles, Dominique, Blanche, Florence, Benjamin, Nathalie, Marine et Arthur
- encadré 7 stages : merci à JB, Lucie, Thomas, Edouard et Amélie
- été entouré par les 24 autres membres du laboratoire : merci à Jeff, Pascal, Sarah, Isabelle, Camille, Benjamin, Yves, Stéphane, Sabrina, Fenouil, Patrick, Nina, Henk, Anne, Marie-Hélène, Mathilde, Pierre, Philippe, Joel, Claire-Lise, Hélène, Maxime, Solenn, Nancy.

Sommaire

Introduction

La **sélection naturelle** est un des moteurs de l'évolution des êtres vivants. Cette sélection peut être **directionnelle**, c'est-à-dire conduire à la fixation ou l'élimination d'un gène ou d'un caractère, ou **balancée**, c'est-à-dire conduire au maintien du polymorphisme d'un gène ou d'un caractère. Les caractères soumis à sélection balancée sont divers, et nous nous sommes intéressés à la sélection balancée agissant sur les **types sexuels**.

Chez les espèces à reproduction sexuée, les types sexuels sont des groupes d'individus ne pouvant pas se reproduire entre eux ; la reproduction sexuée n'est alors possible qu'avec des individus de types sexuels différents. Cette contrainte exercée par les types sexuels provoque une sélection balancée de type **sélection fréquence-dépendante**: au sein d'une population, les types sexuels les plus rares sont favorisés parce que capables de se reproduire avec un plus grand nombre de partenaires que les types les plus fréquents. Cette sélection conduit au maintien d'un polymorphisme des types sexuels au sein des populations et des espèces concernées.

Chez les animaux, on observe généralement deux types sexuels, mâle et femelle, présentant un *sex-ratio* équilibré. Fisher (1930) expliquait cet équilibre par une sélection balancée de type fréquence-dépendante favorisant le sexe rare. Chez les champignons Basidiomycètes, par exemple chez *Schizophyllum commune*, on observe jusqu'à **vingt mille types sexuels différents** générés par recombinaison entre gènes responsables du phénotype sexuel (Kronstad & Staben, 1997) ; chez les Myxocètes, **675 types sexuels** sont possibles chez *Physarum polycephalum* (Iwanaga & Sasaki, 2004).

Chez les plantes, la sélection balancée sur les types sexuels générés par **l'auto-incompatibilité** a fait l'objet de nombreuses **études théoriques**, montrant notamment les conséquences de cette sélection sur la dynamique des allèles présents aux locus concernés. Cependant, les **études empiriques** confirmant ces prédictions restent rares. Nous nous sommes attachés à l'étude empirique de la sélection balancée exercée sur les types sexuels en prenant comme modèle l'auto-incompatibilité sporophytique chez les Brassicaceae. L'intérêt du modèle s'explique (1) la disponibilité d'un grand nombre attendus théoriques développés sur ce système, et (2) par l'étendue des connaissances moléculaires sur les gènes impliqués dans le mécanisme d'auto-incompatibilité.

1. Les systèmes d'auto-incompatibilité

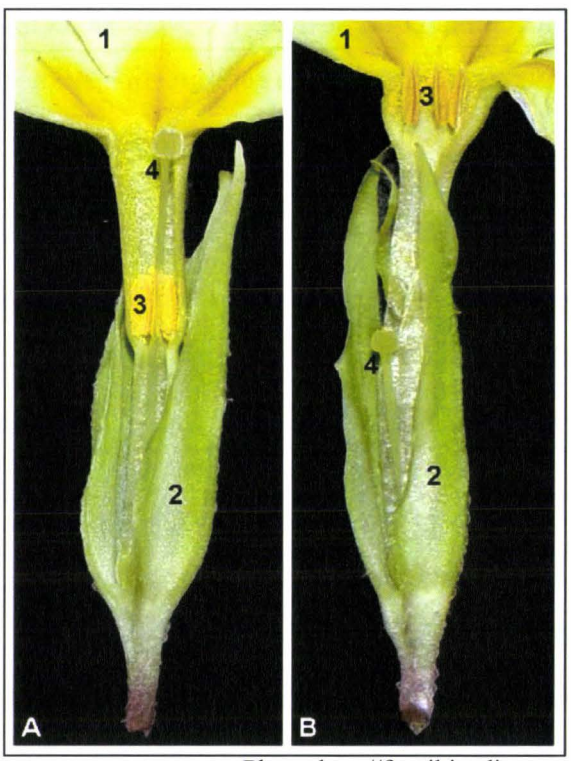


Photo: http://fr.wikipedia.org

Figure 1 : Un exemple d'auto-incompatibilité hétéromorphe, la hétérostylie chez la Primevère. Photo d'une Primevère avec distylie : sur la photo A, fleur de type longistyle et photo B, fleur de type brevistyle. 1 : pétales de la corolle ; 2 : sépales ; 3 : anthère ;4 : style

a. Définition

Chez les Angiospermes, une large majorité d'espèces sont **hermaphrodites**, elles possèdent un organe reproducteur mâle, les anthères qui produisent les gamètes mâles, et un organe reproducteur femelle, le pistil, contenant les gamètes femelles. Ces plantes ont donc la capacité de pratiquer l'**autofécondation**, qui présente des avantages évolutifs importants : elle permet par exemple une **transmission des gènes** deux fois plus importante à la descendance que l'allofécondation ou bien encore une **assurance reproductive** : quelque soit la disponibilité en pollen produit par d'autres individus, la plante a la possibilité de produire des graines. En revanche, l'autofécondation peut également provoquer des phénomènes de **dépression de consanguinité** : l'autofécondation génère plus homozygotie que l'allofécondation, et permet ainsi l'expression de mutations délétères récessives à l'origine **d'une baisse de valeur sélective** (Jarne & Charlesworth, 1993).

Dans certains genres chez les végétaux, sont apparus des systèmes permettant l'évitement de l'autofécondation, on parle alors de système d'**auto-incompatibilité**. L'auto-incompatibilité est un évitement de la reproduction sexuée entre individus de même type sexuel. Le mécanisme est contrôlé par une région génomique que l'on appelle le **locus S** (pour *Self-incompatibility*). Le locus S peut contenir plusieurs gènes qui s'expriment soit au niveau du pistil, soit au niveau du pollen. Il peut se présenter sous des formes très différentes selon le système d'auto-incompatibilité considéré.

b. Diversité des systèmes d'auto-incompatibilité

Les systèmes d'auto-incompatibilité sont apparus de nombreuses fois au cours de l'évolution des Angiospermes, sous différentes formes et de manières indépendantes (de Nettancourt, 2001).

- L'auto-incompatibilité hétéromorphe

Dans les systèmes d'**auto-incompatibilité hétéromorphe**, les types sexuels sont morphologiquement différents. Le dimorphisme sexuel est alors directement responsable de l'auto-incompatibilité, car il favorise les croisements entre types sexuels différents. C'est le

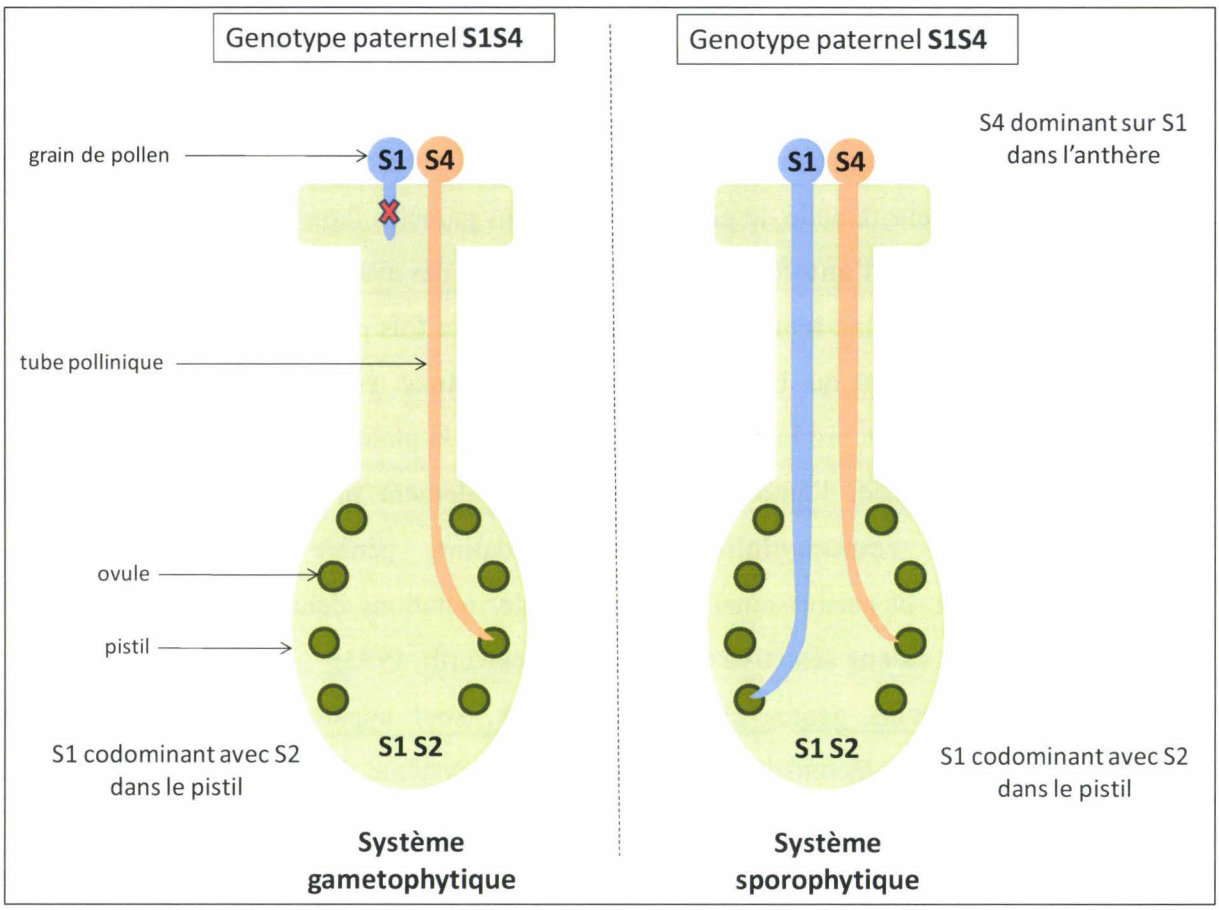


Figure 2 Schéma fonctionnel des systèmes d'auto-incompatibilité gamétophytique (GSI) et sporophytique (SSI).
Dans l'exemple présenté ici, on considère un croisement entre un pistil de génotype S1S2 avec du pollen provenant d'un individu de génotype S1S4. Dans les deux systèmes (GSI et SSI), S1 et S2 sont codominants dans le pistil : le pistil a donc pour phénotype S1S2.
Dans le système GSI, le phénotype du pollen est codé par son propre génotype : ainsi le pollen de génotype S1 a pour phénotype S1 et est donc rejeté tandis que le pollen de génotype S4 a pour phénotype S4 et est donc compatible avec le pistil de phénotype S1S2.
Dans le système SSI, le phénotype du pollen est codé par le génotype du parent S1S4. Dans cet exemple, l'allèle S4 est dominant sur l'allèle S1, les grains de pollen de génotype S1 et S4 ont donc tous deux pour phénotype S4. Ils sont donc compatibles tous les deux avec le pistil de phénotype S1S2.

cas de l'**hétérostylie** que l'on peut observer par exemple chez *Primula vulgaris* (Primulacée) (Richards, 1997). Au sein de cette espèce, il existe deux types de fleurs portés par des individus différents : des fleurs à style long et étamines insérées à mi-hauteur de la corolle chez les uns, des fleurs à style court et étamines insérées à l'extrémité de la corolle chez les autres (Fig. 1). Le pollen déposé et transporté sur le corps d'un insecte atteindra les styles de même longueur que l'étamine productrice.

- **L'auto-incompatibilité homomorphe**

Dans les systèmes d'**auto-incompatibilité homomorphe**, l'auto-incompatibilité est due à des **mécanismes biochimiques** inhibant la **germination** ou la **croissance des tubes polliniques**. Par ces mécanismes, les grains de pollen exprimant un allèle du locus d'auto-incompatibilité sont incapables de féconder les ovaires portés par des pistils exprimant le même allèle. Le phénotype d'auto-incompatibilité exprimé dans le pollen ou le pistil est également appelé **spécificité**. Ces systèmes sont responsables non seulement de l'évitement de l'autofécondation, mais aussi plus généralement de la reproduction entre individus apparentés. Il existe plusieurs types de systèmes d'auto-incompatibilité homomorphe qui semblent être issus d'histoires évolutives indépendantes. On distingue deux groupes principaux : les systèmes « gamétophytiques » et les systèmes « sporophytiques » (Fig.2).

- Les systèmes d'auto-incompatibilité gamétophytiques (GSI)

Ces systèmes sont les plus communs : ils concerneraient environ la moitié des familles d'Angiospermes. Ils restent aujourd'hui encore assez mal connus dans de nombreuses familles et n'ont été caractérisés sur le plan moléculaire que chez les Solanaceae, Rosaceae, Plantaginaceae, et Papaveraceae (Charlesworth *et al.*, 2005).

Ces systèmes sont dits « gamétophytique » car c'est le génome du **gamétophyte** mâle (le pollen) qui détermine son phénotype d'auto-incompatibilité. En effet, dans les GSI, la spécificité exprimée par le pistil est déterminée par les deux allèles codominants portés par la plante mère, tandis que la spécificité exprimée par le pollen est déterminée par le génome **haploïde** du grain de pollen lui-même. Par conséquent, lorsque le pollen porte un allèle identique à l'un des deux allèles exprimés dans le pistil, le tube pollinique voit sa croissance stoppée et la fécondation ne peut donc pas avoir lieu.

- Les systèmes d'auto-incompatibilité sporophytiques (SSI)

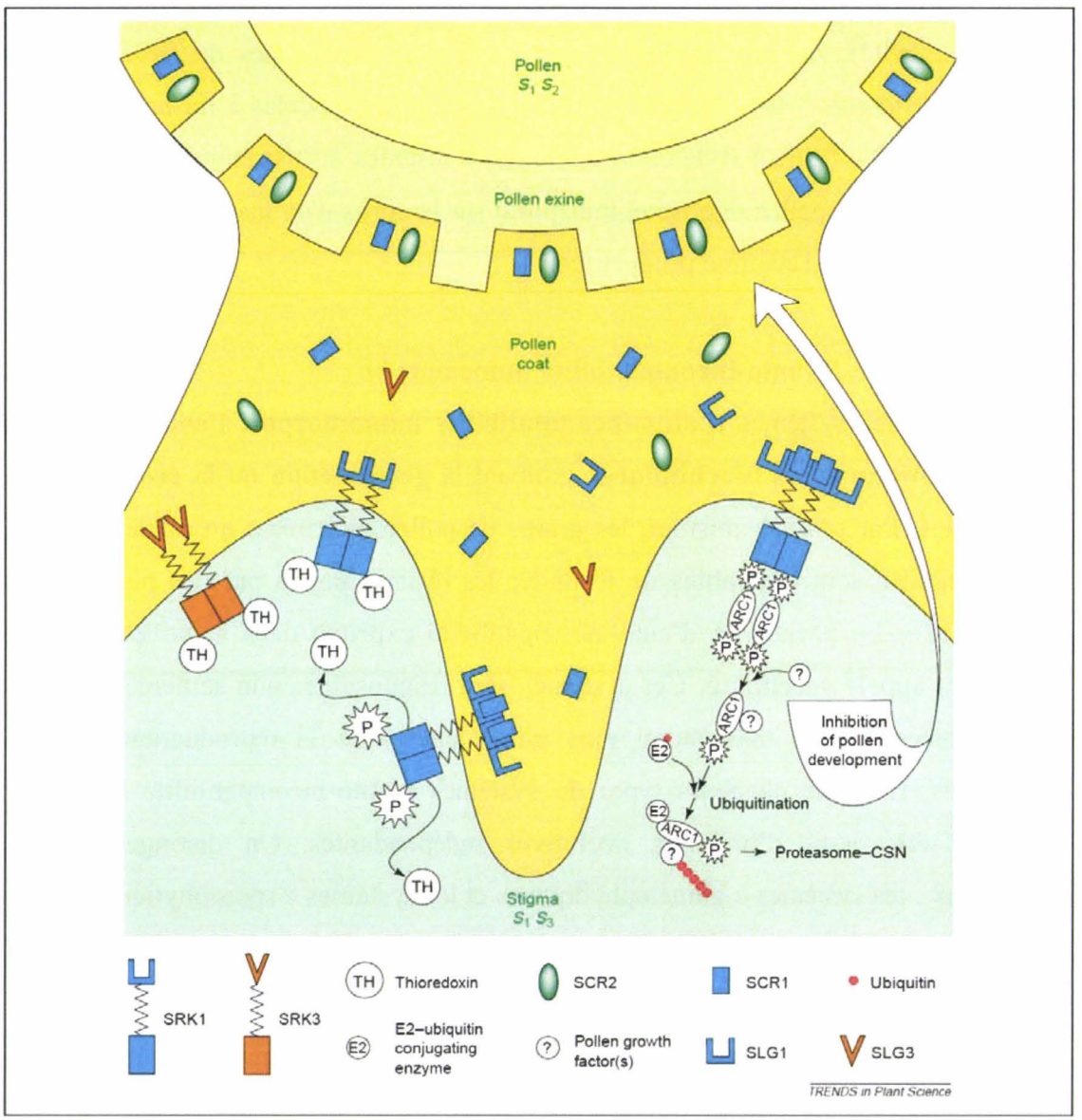


Figure 1. Model for the mechanism of sporophytic self-incompatibility (SSI) in *Brassica*. *S*-locus receptor kinase (SRK), the female determinant of SSI is a serine–threonine receptor kinase that spans the plasma membrane of stigmatic papilla cells. *S*-locus cysteine-rich protein (SCR), the male determinant of SSI is the cognate ligand of SRK. SLG is a secreted glycoprotein with high sequence similarity to the receptor region of SRK; evidence indicates that in most cases *S*-locus glycoprotein (SLG) is not essential for SSI although it has been shown to form part of the SCR–SRK interaction complex. *S*-gene products from the same *S*-haplotype are depicted in the same colour. In this example, pollen and stigma share gene products of the S_1 haplotype and therefore are incompatible. Pollen recognition involves a haplotype-specific interaction between SCR1 and the receptor domain of SRK1, which 'activates' the kinase domain. In an 'inactive' state, SRK molecules appear to associate as dimers or oligomers with thioredoxins bound to their kinase domains. Association with the thioredoxin is thought to prevent autophosphorylation. Activation of SRK following SCR binding is accompanied by dissociation of the thioredoxins and autophosphorylation on serine and threonine residues in the kinase domain. A stigma-specific arm-repeat motif-containing protein, ARC1, interacts in a phosphorylation-dependent manner with the kinase domain of SRK and functions as an E3 ubiquitin ligase. Therefore, ARC1 is thought to be responsible for directing localized inhibition of self-pollen development by initiating the degradation of pollen growth promoting factors via the ubiquitin pathway, which targets ubiquitinated proteins to the proteasome–COP9 signalasome (CSN) for degradation. Figure adapted from Ref. [39].

Figure 3 Mécanismes moléculaires de la réaction d'auto-incompatibilité chez *Brassica* montrant la localisation transmembranaire de la protéine SRK et la localisation à la périphérie de l'exine du grain de pollen de la protéine SCR, d'après Hiscock et McInnis (2003)

Ces systèmes, présents chez les *Brassicaceae* et les *Asteraceae*, sont dits **sporophytiques** car c'est le génome diploïde du sporophyte producteur du pollen qui détermine le phénotype d'auto-incompatibilité du pollen ou du pistil. Le phénotype du pollen est exprimé au niveau de l'enveloppe pollinique fabriquée dans les anthères du parent, et tient compte des deux allèles parentaux. Ainsi, pour un parent de génotype diploïde S1/S2, tous les grains de pollen produits, pourtant haploïdes et ne portant qu'un allèle S1 ou S2, expriment un phénotype codé par le génotype **diploïde** S1/S2.

Des **relations de dominance** complexes entre allèles du locus S ont été observées dans les systèmes d'auto-incompatibilité sporophytique (Bateman, 1952). Ces relations de dominance modifient l'expression des phénotypes du pollen et du pistil : ainsi dans certains cas, un pollen peut féconder un pistil portant un allèle en commun si cet allèle est récessif dans le pollen ou le pistil.

c. Le système d'auto-incompatibilité des Brassicaceae

Dans le cadre de cette thèse, nous nous sommes intéressés plus particulièrement au cas de l'auto-incompatibilité sporophytique des *Brassicaceae* dont les mécanismes moléculaires ont été bien caractérisés. Dans cette famille, le locus d'auto-incompatibilité possède deux gènes essentiels (Hiscock & McInnis, 2003) :

- Le gène ***SRK*** (**S**-**R**eceptor **K**inase) est exprimé dans le pistil dans les cellules papillaires du stigmate. Il code pour une protéine trans-membranaire présentant notamment un domaine extracellulaire récepteur (domaine S), un domaine trans-membranaire et un domaine intra-cellulaire à activité kinase (domaine K) responsable de l'activation de la cascade de réaction induisant le rejet du pollen.

- Le gène ***SCR*** (**S**-locus **C**ysteine-**R**ich protein), exprimé dans les cellules tapétales de l'anthère, et qui code pour une protéine localisée sur l'enveloppe du grain de pollen capable de se lier spécifiquement au récepteur de la cellule du stigmate.

Lorsqu'un grain de pollen se dépose sur un stigmate dont il partage la spécificité allélique (auto-pollen par exemple), la protéine SCR portée par son enveloppe se lie au récepteur SRK spécifique. Le domaine kinase de la protéine transmembranaire SRK est alors activé, ce qui déclenche une cascade de réactions aboutissant à l'inhibition des aquaporines membranaires. Ceci empêche la sortie des molécules d'eau nécessaires à la germination du grain de pollen, et par conséquent, la fécondation (Fig. 3).

Les gènes *SCR* et *SRK* du locus S semblent étroitement liés génétiquement. En effet, plusieurs études ont suggéré que le **taux de recombinaison** dans la région du locus S était très faible. A l'extrême, Casselman *et al.* (2000) n'ont pas trouvé de recombinaison au locus S chez *Brassica sp.*. L'étude de Kamau *et al.* (2005) sur la diversité nucléotidique dans la région du locus S chez *Arabidopsis lyrata* a aussi suggéré un taux de recombinaison bas. Kawabe *et al.* (2006) ont également montré un taux de recombinaison faible au locus S à partir d'analyse de descendances chez *A. lyrata*. Les gènes *SCR* et *SRK* forment un **système « clef-serrure »** avec une reconnaissance spécifique des allèles *SCR* par les allèles *SRK* ce qui suggère une **coévolution** forte entre les deux gènes. En cas de recombinaison entre les gènes *SRK* et *SCR*, on observerait une rupture de l'auto-incompatibilité pouvant permettre notamment l'autofécondation aboutissant à l'expression de la **dépression de consanguinité**. La coévolution des gènes *SRK* et *SCR* permet donc une conservation du fonctionnement de l'auto-incompatibilité et aurait pour moteur la sélection négative exercée par la dépression de consanguinité.

2. Mécanismes évolutifs agissant au locus d'auto-incompatibilité

La question des forces évolutives agissant au locus d'auto-incompatibilité a été bien étudiée sur le plan théorique (Castric & Vekemans, 2004). La principale force évolutive impliquée semble être la **sélection fréquence-dépendante**, dont les mécanismes ont été analysés à l'aide de différents modèles.

a. Sélection fréquence-dépendante symétrique

La première force évolutive supposée agir sur le locus S a été suggérée par **Wright (1939)**. Wright proposait d'expliquer l'évolution des fréquences alléliques au locus S par l'existence d'une **sélection fréquence- dépendante négative agissant sur la valeur sélective paternelle**. En effet, un pollen portant une spécificité allélique rare au locus S, aurait plus de partenaires compatibles possibles au sein de la population. L'individu produisant ce pollen aurait donc potentiellement un plus grand nombre de descendants. Cet allèle « rare » augmenterait donc en fréquence grâce à la sélection fréquence–dépendante.

Cette théorie a fait l'objet d'un modèle qui peut s'appliquer à un système gamétophytique (GSI) ou à un système sporophytique (SSI), où tous les allèles seraient codominants (modèle SSI-COD). Ce modèle prend en compte la mutation, la dérive, et la sélection fréquence-dépendante. Les n allèles présents dans la population sont considérés

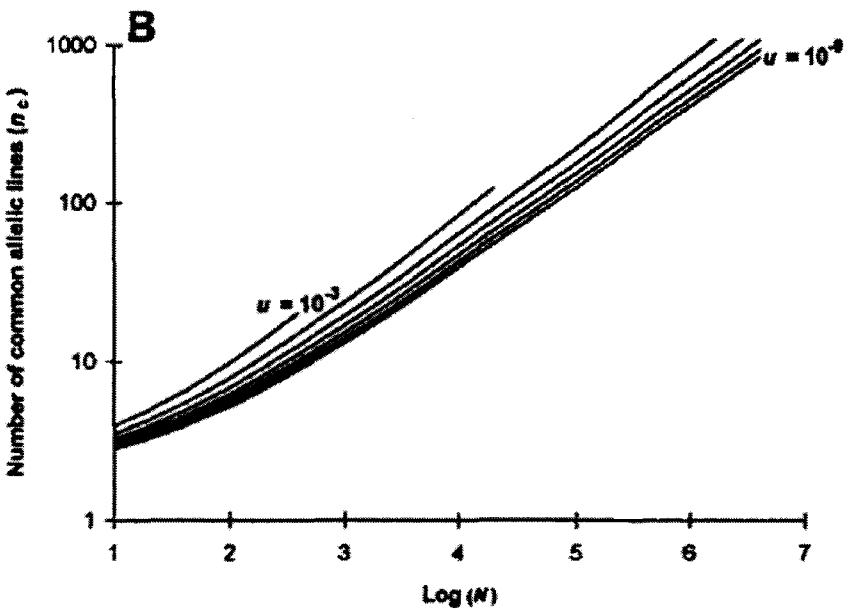


Figure 4 : Nombre de lignées allèliques (n_c) en fonction de la taille de la population (N) pour une gamme de taux de mutation par locus (u), d'après Vekemans & Slatkin (1994)

comme ayant toute la même valeur sélective intrinsèque. Ainsi leur transmission relative à la génération suivante ne dépend que de leur fréquence dans la population. En population infinie, le modèle prédit qu'à l'équilibre, tous les allèles du locus S se retrouvent à la même fréquence, qui s'élève donc à l'inverse du nombre d'allèles présent dans la population ($1/n$). La sélection fréquence-dépendante est donc qualifiée de **symétrique** puisqu'elle agit en diminuant la fréquence des allèles communs symétriquement avec l'augmentation de la fréquence des allèles rares. En population finie, le modèle démontre que les fréquences alléliques au locus S fluctuent au cours des générations, tout en restant centrées autour de la fréquence observée à l'équilibre en population infinie. La variance autour de ce pic de fréquence allélique serait proportionnelle à l'intensité de la dérive génétique.

Ce modèle a permis de faire des prédictions théoriques sur le nombre et la fréquence des allèles présents chez une espèce, ainsi que sur leur temps de divergence. Ainsi, le modèle de Wright (1939) a démontré qu'un grand nombre d'allèles pouvait être maintenu au sein d'une population. De plus, ce nombre d'allèles augmente lorsque la taille de la population augmente (Voir Fig. 4). Vekemans *et al.* (1994) ont également mis en évidence que le temps de divergence entre allèle du locus S est supérieur aux barrières d'espèces.

Ce **fort polymorphisme** allélique attendu en sélection balancée semble en effet présent dans les espèces auto-incompatibles. Ainsi, une grande diversité allélique au locus S a été démontrée chez plusieurs *Brassicaceae*. Par exemple, 30 allèles ont été détectés chez *Brassica campestris* (Nou *et al.*, 1993), 18 chez *Raphanus sativus* (Sakamoto *et al.*, 1998), 50 allèles chez *Brassica oleracea* (Ockendon, 2000), 20 chez *Brassica insularis* (Glémin *et al.*, 2005), et 30 chez *Arabidopsis lyrata* (Prigoda *et al.*, 2005); ce qui semble confirmer l'action de la sélection balancée sur le locus d'auto-incompatibilité.

Le modèle initial de sélection fréquence-dépendante proposé par Wright (1939) faisait l'hypothèse d'une sélection agissant sur la voie mâle uniquement. Dans ce modèle, toutes les valeurs sélectives femelles étaient égales, chaque pistil ayant tous ses ovules fécondés par du pollen compatible, quel que soit son génotype au locus S. Vekemans *et al.* (1998) ont par la suite introduit la notion de « **fecundity selection** » c'est-à-dire un processus de sélection fréquence-dépendante agissant sur le locus S à travers la voie femelle : dans ce modèle, les pistils possédant des spécificités alléliques rares auraient une meilleure valeur sélective que les pistils portant une spécificité commune dans la population. Cette « fecundity selection » pourrait notamment avoir lieu dans des populations où la disponibilité en partenaires serait limitée, et où par conséquent certains pistils pourraient ne pas recevoir en quantité suffisante du pollen compatible. L'ajout au modèle de cette sélection fréquence-dépendante par la voie

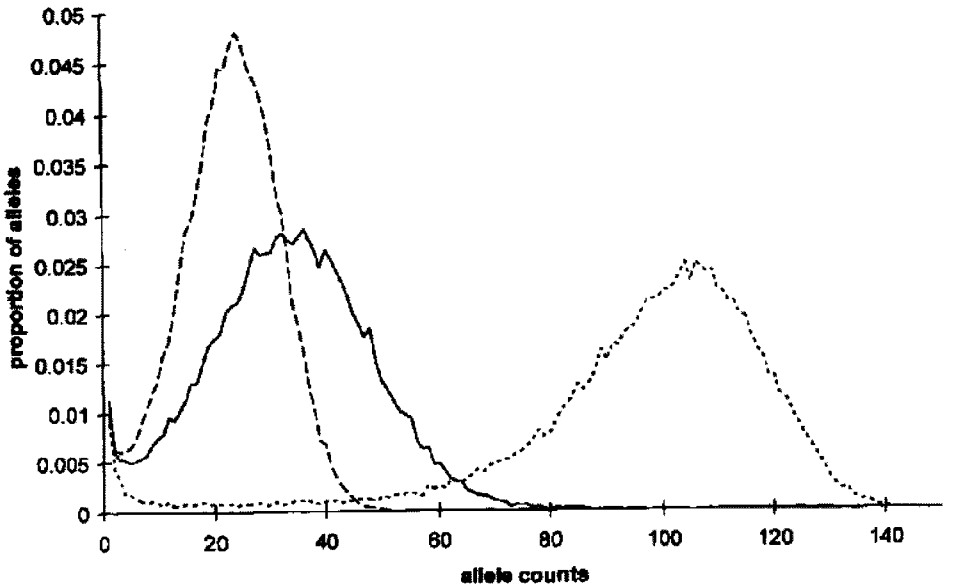


Figure 5 : Distribution des fréquences à l'équilibre d'allèles appartenant à différentes classes de dominance dans un modèle d'auto-incompatibilité sporophytique DOM avec taille de population N : 100 et taux de mutation $u=10^{-4}$ d'après Schierup *et al.* (1997)

--- dominant ··· intermédiaire —— récessif

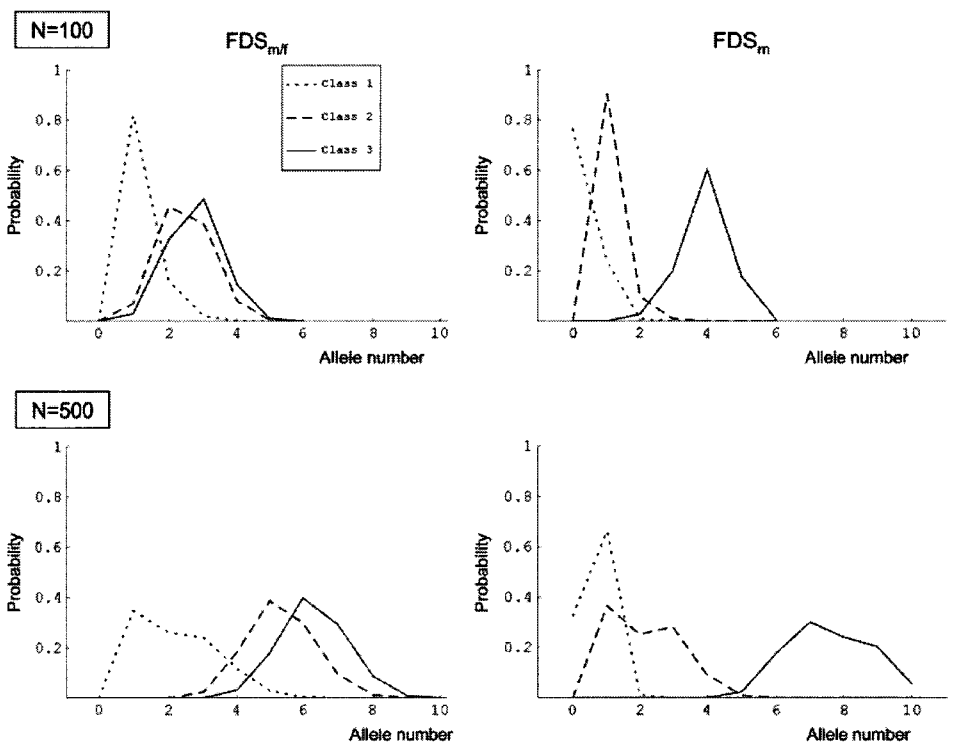


Figure 6 : Distribution du nombre d'allèles d'auto-incompatibilité par classe de dominance dans un modèle SSI avec relation de dominance de type DOMCOD (d'après Billiard *et al.* 2007)
Class 1 : Classe des allèles dominants ; Class 2 : Classe des allèles de dominance intermédiaire ; Class 3 : Classe des allèles récessifs ; N : taille de la population ; $FDS_{m/f}$: modèle de sélection fréquence dépendante par les voies mâle et femelle ; FDS_m : modèle de sélection fréquence dépendante par la voie mâle uniquement

femelle modifie la dynamique des fréquences alléliques au locus S. Par la suite, la plupart des auteurs travaillant sur des systèmes d'auto-incompatibilité ont conservé la comparaison entre les deux types d'approches, puisqu'elles correspondent à des situations écologiques différentes rencontrées dans la nature. L'existence de différences dans la production de graines en fonction de la disponibilité en pollen compatible a ainsi été observée en populations naturelles (Wagenius *et al.*, 2007).

b. Sélection fréquence-dépendante asymétrique : effet de la dominance des allèles au locus S

Le modèle de Wright ne s'applique pas aux systèmes **d'auto-incompatibilité sporophytique** (SSI) présentant des **relations de dominance** entre allèles. En effet, dans ce cas et contrairement aux modèles précédents, la sélection fréquence-dépendante agit de manière différente sur les allèles d'auto-incompatibilité en fonction de leur niveau d'expression. Les valeurs sélectives des différents allèles sont donc **asymétriques** et dépendent de leur niveau de dominance. A l'aide de simulations numériques, Schierup *et al.* (1997) ont étudié plusieurs modèles d'auto-incompatibilité sporophytique (SSI). Le **modèle DOM** présente une hiérarchie de **dominance stricte** dans le pollen et le pistil : la population considérée possède par exemple trois allèles : un récessif par rapport au deuxième, lui-même récessif par rapport au troisième. Le **modèle DOMCOD** présente une hiérarchie de **dominance** dans le pollen et une **codominance** entre les trois allèles dans le pistil. Les relations de dominance entre allèles impliquent que l'expression des allèles récessifs dépend de l'autre allèle présent au locus S. Lorsqu'ils ne sont pas exprimés, les allèles récessifs ne sont donc pas soumis à la sélection fréquence-dépendante et ont donc un comportement neutre. Par conséquent, la pression de sélection fréquence-dépendance exercée sur les allèles récessifs est donc plus faible que sur les allèles dominants. La fréquence à l'équilibre d'un allèle dépend donc de son niveau de dominance dans ces modèles. Ainsi, dans les deux types de modèle (DOM ou DOMCOD), plus un allèle est récessif, plus sa fréquence dans la population sera élevée, ce phénomène a donc été nommé l'**effet récessif**. (Fig 5).

Ce modèle a été complexifié par Uyenoyama (2000) qui a présenté un modèle DOMCOD où les allèles S sont regroupés en deux classes de dominance dans le pollen : les allèles de la classe I étant tous dominants par rapport à ceux de la classe II et les allèles d'une même classe étant codominants. Ce modèle a montré que seul un allèle était maintenu dans la

Genotypes and number of S-alleles detected and estimated in the five populations

	Populations					
	Teghime	Caporalino	Inzecca	Punta Gorbaghiola	Punta Calcina	Total
No. of plants:	50	55	74	61	33	273
Alleles detected						
Class I	8	11	5	6	2	18
Class II	2	2	2	2	1	2
Ambiguous	3	3	2	0	0	4

Figure 7 : Nombre d'allèles d'auto-incompatibilité détectés dans cinq populations naturelles de *Brassica insularis*. Les allèles de la classe I sont dominants par rapport aux allèles de la classe II dans le pollen, d'après Glémin *et al.* (2005)

Assumed dominance hierarchy of AlSRK haplotypes, and the corresponding type designations used in this article

Dominance class	Haplotypes	Designation
Most dominant	9, 11, 12, 15, 22, 38	S_1, S_2, S_3, S_4, S_5, S_6
Dominant	16, 25	T_1, T_2
Least dominant	6, 14	U_1, U_2
Recessive	1	V

Codominance within each class and the same dominance hierarchy in males and females are assumed.

Figure 8 : Relations de dominance supposées entre allèles du locus S chez *A. lyrata,* d'après Schierup *et al.* (2006)

classe récessive tandis que plusieurs pouvaient être conservés dans la classe plus dominante, et ce à fréquence égale. Il existerait donc un phénomène de **limitation du nombre d'allèles récessifs** conservés au cours de l'évolution.

Un modèle général a été développé par Billiard *et al.* (2007) qui ont modélisé un système SSI dans lequel les catégories de dominance et le nombre d'allèles par catégories peuvent varier. Ce modèle, appliqué en population finie avec mutations aléatoires indépendantes du niveau de dominance, confirme le maintien d'un seul allèle dans la classe la plus récessive mais suggère également que les allèles d'une catégorie de dominance sont d'autant plus nombreux que cette catégorie est dominante par rapport aux autres (Fig. 6).
Ces conclusions ont été confirmées par des études empiriques. L'étude de Glémin *et al.* (2005) a mis en évidence que les allèles d'incompatibilité chez *Brassica insularis* étaient beaucoup plus nombreux dans la classe I des allèles dominants que dans la classe II des allèles récessif (Fig. 7). De la même manière, Schierup *et al.* (2006) ont montré un patron de dominance hiérarchique chez *Arabidopsis lyrata* avec un nombre d'allèle par classe augmentant avec le niveau de dominance de la classe: un seul allèle a été détecté dans la classe la plus récessive, deux allèles sont présents dans les deux classes intermédiaires et six allèles sont observés dans la classe la plus dominante (Fig. 8). Cette étude a aussi mis en évidence l'effet récessif sur les fréquences alléliques au locus S : par exemple l'allèle le plus récessif AlSRK01 présente la fréquence la plus élevée (38%) dans la population considérée. Cette fréquence de 38% est très largement supérieure à la fréquence attendue sous l'hypothèse d'une sélection fréquence-dépendante symétrique (1/11 allèles = 9%).

D'autre part, peu d'éléments sont connus sur les **mécanismes responsables des patrons de dominance** : il n'est pas encore clair si la dominance est due à des contraintes d'ordre biochimiques et mutationnelles, ou si elle est le fruit de pression de sélection. La question de la détermination de la dominance au locus S pourrait avoir des conséquences évolutives importantes. Les modèles décrits ci-dessus considèrent la dominance comme intrinsèque à l'allèle, mais on pourrait se demander si ces prédictions seraient identiques dans le cas où la dominance serait codée par un locus *modifieur* d'expression, et pourrait donc évoluer indépendamment.

c. Sélection fréquence-dépendante asymétrique et fardeau génétique associé au locus S

Un autre phénomène pourrait également modifier l'évolution des allèles du locus S : la présence d'un **fardeau génétique lié**. En effet, le taux de recombinaison étant très bas dans la région du locus S, des mutations peuvent donc être associée aux allèles S sur de longues périodes évolutives. Etant donné que le taux d'hétérozygotie au locus S est très élevé et que la divergence entre allèle S est forte, ces mutations seraient rarement purgées si elles étaient récessives. C'est pourquoi, Uyenoyama (1997) a suggéré qu'il pourrait y avoir une **accumulation de mutations délétères récessives liées au locus S**. Ce cortège de mutations délétères serait fortement associé à chaque allèle S et formerait un fardeau génétique protégé différent associé à chaque allèle S. L'efficacité de la purge de ces mutations délétères dépendrait ainsi du niveau de dominance de l'allèle S associé. Les allèles dominants se retrouvant le plus rarement à l'état homozygote, leurs mutations délétères associées sont moins purgées et s'accumulent donc d'autant plus. L'importance de ce fardeau serait donc reliée au **niveau de dominance** de l'allèle auquel il est associé, et conditionnerait ainsi l'évolution des allèles S. La modélisation de ce fardeau suggère que l'apparition de nouveaux allèles au locus S serait ralentie (Uyenoyama, 2005) : en effet les nouveaux allèles créés partageraient le même cortège de mutations délétères que l'allèle ancestral. Une compétition entre le nouvel allèle et son ancêtre aurait donc lieu, diminuant la vitesse d'apparition de nouveaux allèles.

d. Dérive génétique

La **dérive génétique** a également une grande importance sur la dynamique des allèles S. En effet, les allèles récessifs, lorsqu'ils ne sont pas exprimés, ne bénéficient pas de l'avantage du rare conféré par la sélection fréquence-dépendante. Ainsi les **allèles récessifs** ont une probabilité plus grande d'être éliminés par la dérive génétique que les allèles dominants lorsqu'ils sont en faible fréquence. Un nouvel allèle S aura ainsi une plus grande probabilité d'envahir une population finie s'il est dominant que s'il est récessif (Schierup *et al.*, 1997).

La dérive génétique a une influence sur le **nombre d'allèles S** qu'il est possible de maintenir dans une population finie mais peut aussi modifier les **fréquences alléliques** au locus S. Elle modifie donc la gamme de fréquence possible pour les allèles S. Il est donc

primordial de la prendre en compte lorsque l'on veut tester l'effet de la sélection fréquence-dépendante sur le polymorphisme au locus S.

e. Migration

L'avantage conféré aux allèles rares par la sélection fréquence-dépendante a aussi des conséquences sur la dispersion des allèles S : quand la structuration des populations est importantes, certains **allèles migrants** se retrouvent en faibles fréquences dans les populations dans lesquelles ils arrivent. Ils sont donc très fortement sélectionnés par la sélection fréquence-dépendante. On suppose ainsi que la **migration efficace** des allèles S serait supérieure à la migration des allèles neutres.

Des études théoriques montrent que la différenciation génétique entre populations est attendue plus faible au locus S qu'à des marqueurs neutres (Charlesworth *et al.*, 2003a). Cette tendance a été confirmée par des études empiriques chez *Brassica insularis* (Glémin *et al.*, 2005) ou chez *Senecio squalidus* (Brennan *et al.*, 2006). Cette différence de migration efficace entre les allèles du locus S et marqueurs neutres pourrait également être observée à l'intérieur même d'une population.

f. Mutation

Au locus d'auto-incompatibilité, la fixation des mutations est conditionnée par la sélection fréquence-dépendante. Au sein du gène *SRK*, dans la région codant pour la reconnaissance du pollen (appelée domaine hyper-variable), on observe un rapport de **mutations non synonymes** sur mutations synonymes supérieur à l'attendu neutre (Castric & Vekemans, 2007). Ces mutations non-synonymes sont responsables des différentes spécificités d'auto-incompatibilité et ont été soumises à une **sélection positive** de la part de la sélection fréquence-dépendante. Cependant, l'apparition de nouvelles spécificités est conditionnée par la co-évolution entre les gènes exprimés pour les spécificités du pollen et du pistil. Il existe donc également une **sélection purifiante** au sein des spécificités alléliques, qui favorise les mutations synonymes par rapport aux non-synonymes.

Plusieurs modèles d'apparition de nouvelles spécificités ont été proposés : soit par mutations simples sur les gènes pollen et pistil supposant une phase de perte de spécificité, soit par sélection à partir de **nuage d'allèles polymorphes** au sein de chaque spécificité. Dans ce cas, il y aurait une sélection positive des variants présentant les meilleures capacités de

Figure 9 : Photo *d'Arabidopsis halleri* (Brassicaceae) en culture à la serre. Cette espèce présente une croissance végétative sous forme de rosette et un développement de hampes florales portant de petites fleurs à pétales blancs.

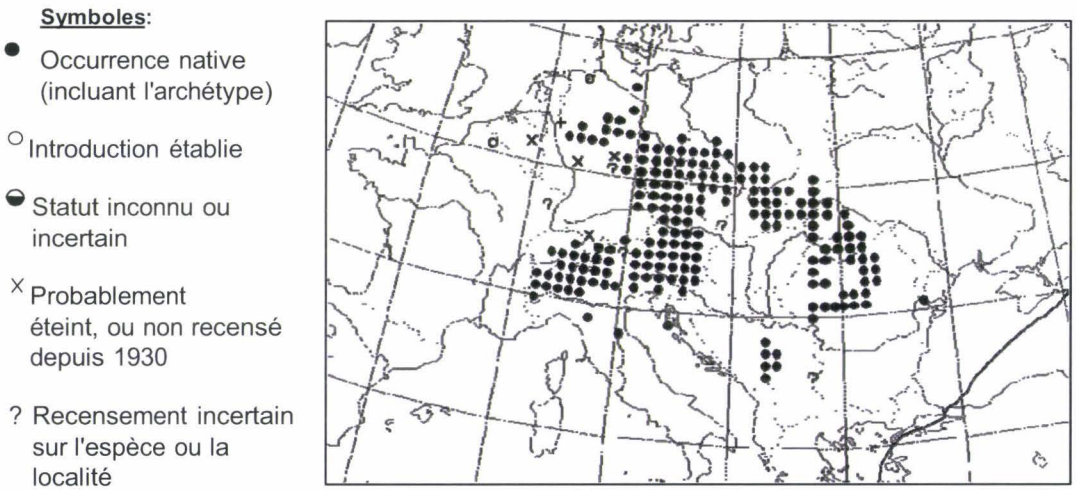


Figure 10 : Aire de répartition d'*Arabidopsis halleri* en Europe. (d'après l'Atlas *flora europea*)

discriminations des spécificités. C'est cette sélection qui pourrait être à l'origine de la divergence au sein du nuage, et donc de l'apparition de nouvelle spécificité. Par des manipulations expérimentales de séquences de gènes d'auto-incompatibilité polliniques, Chookajorn *et al.* (2004) ont montré que les spécificités d'auto-incompatibilité chez *Brassica* étaient résistantes à de nombreux changements. Cependant, ce type de phénomène ne peut avoir lieu que s'il existe une variabilité importante au sein des spécificités alléliques. Bien que peu d'études se soient penchées sur cette question, Charlesworth *et al.* (2005) estiment que cette variabilité doit être faible au vu de la taille efficace faible des population d'allèles S.

Le mécanisme d'apparition de nouvelles spécificités modulerait donc l'hypothèse de la sélection fréquence-dépendante classique et par conséquent la dynamique des allèles au locus d'auto-incompatibilité.

3. Espèce modèle : *Arabidopsis halleri*

L'objectif de ma thèse est d'éclaircir l'importance des différents facteurs impliqués dans l'évolution des allèles S présentés ci-dessus en utilisant l'espèce modèle ***Arabidopsis halleri*** (*Brassicaceae*). C'est une **espèce herbacée**, formant des rosettes (Fig. 9), **pérenne et auto-incompatible**. Elle fleurit une à deux fois par an : au printemps et parfois en automne. Les individus peuvent s'étendre sur des surfaces assez larges en formant des stolons. La reproduction sexuée est **entomophile**. Les plantes produisent un grand nombre de graines.

Cette espèce pousse sur milieu acide, oligotrophe et humide, et possède la particularité d'être **tolérante et hyperaccumulatrice du zinc et du cadmium**. Elle est présente dans toute l'Europe Occidentale et Centrale, dans des zones continentales ou montagneuses, sur des sols acides et frais (Fig. 10). En France, on la trouve exclusivement sur des sols contaminés par les métaux lourds, et il semble que son apparition corresponde au développement des premières industries d'extraction minière qui ont conduit à une forte pollution par ces métaux. Le genre *Arabidopsis* comprend également les espèces *A. thaliana* et *A. lyrata* (Fig. 11). La grande proximité phylogénétique entre *A. halleri* et ***A. thaliana*** permet une transposition facile des connaissances moléculaires obtenues sur cette espèce modèle dont le génome est séquencé depuis 2000 et progressivement annoté (The Arabidopsis Genome Initiative, 2000). *Arabidopsis halleri* est également proche d'*A. lyrata* (Fig. 11), qui est un modèle bien étudié pour l'auto-incompatibilité sporophytique, et dont un projet de séquençage du génome complet est en cours. *Arabidopsis halleri* est décrite comme une espèce à **système d'auto-**

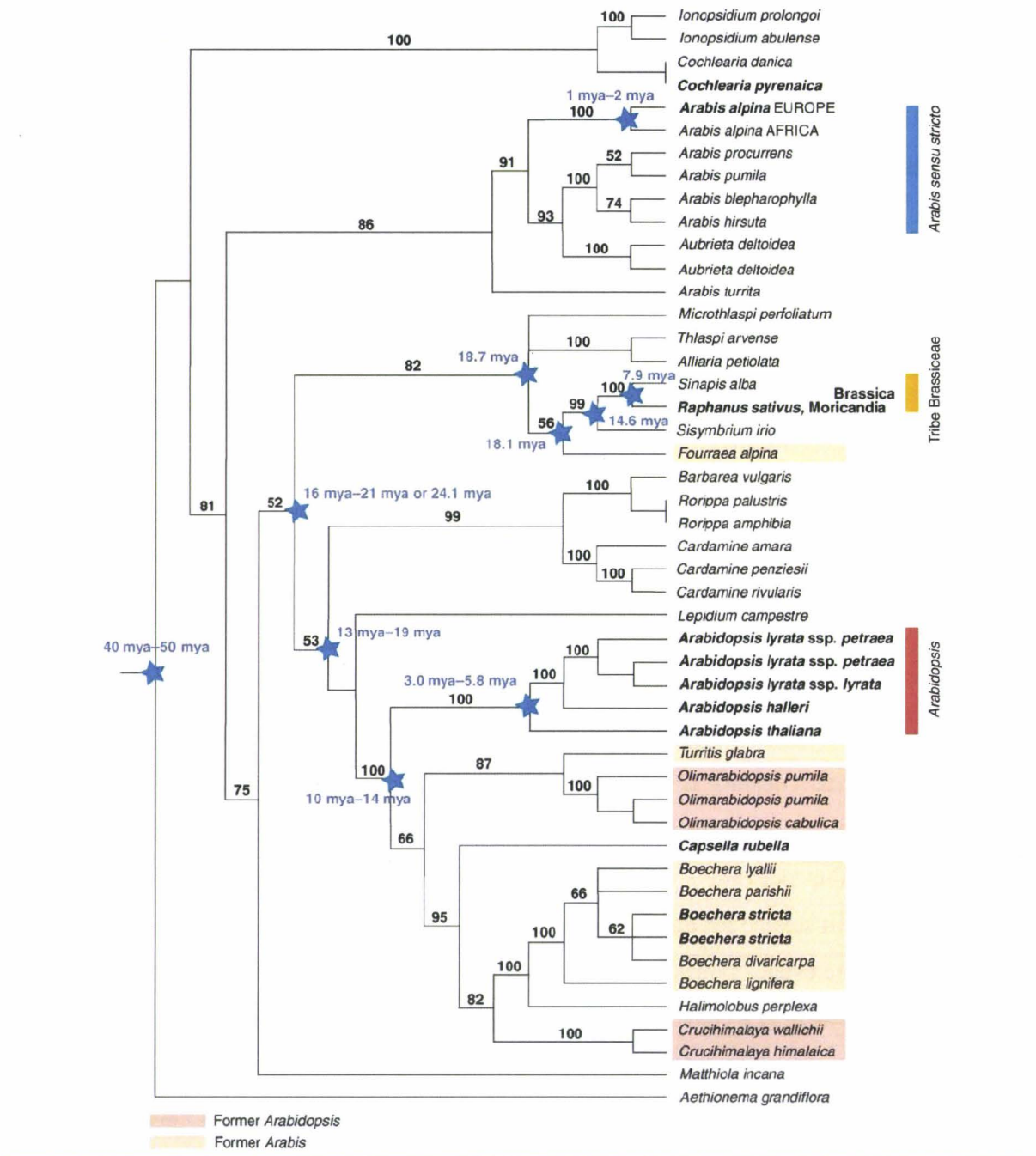


Figure 11 : Phylogénie des Brassicaceae montrant la position du genre Arabidopsis, d'après Clauss & Koch (2006). Arbre consensus de neuf arbres les plus parsimonieux obtenus à partir de séquences de matK et Chs (abbreviation : mya = millions d'années)

incompatibilité sporophytique (Clauss & Koch, 2006), mais jusqu'à présent cependant, aucune vérification expérimentale n'est venue confirmer ces observations.

Egalement, *A. halleri* est un modèle bien étudié au sein du laboratoire notamment pour ses capacités de tolérance et d'accumulation des métaux lourds. Cette espèce pourrait jouer un rôle dans la phyto-extraction de métaux lourds de sols contaminés. Du point de vue écologique, cette particularité en fait également un modèle pertinent : cette espèce subit des contraintes écologiques fortes, en survivant sur des milieux hétérogènes et ayant une adaptation locale récente.

L'objectif de ma thèse était de caractériser les forces évolutives agissant au locus d'auto-incompatibilité chez *A. halleri* au niveau populationnel. Fortement guidée par les attendus théoriques développés dans la littérature, j'ai choisi d'orienter ma thèse autour de quatre thèmes principaux.

Chapitre 1 : Régime de reproduction et dispersion en population naturelle

Tout d'abord, nous avons voulu étudier les **patrons de reproduction** chez une espèce auto-incompatible, et tester ainsi l'effet de ce système de reproduction sur la structure génétique spatiale. Nous avons intensivement étudié avec des marqueurs neutres sur deux générations une population isolée d'*A. halleri*. Tout d'abord, nous avons testé la **fonctionnalité du système d'auto-incompatibilité**, caractérisé les patrons de reproduction intra-population, et tenter de déterminer quelle part de ces patrons est dirigée par l'auto-incompatibilité. Puis nous nous sommes intéressés à **l'influence de l'auto-incompatibilité sur la structuration de variabilité génétique au locus S et pour des locus neutres**, afin de tester l'hypothèse d'une migration efficace supérieure pour les allèles du locus S par rapport aux allèles neutres.

Chapitre 2 : Influence de la sélection fréquence- dépendante et de la dominance sur le polymorphisme au locus S

De nombreux modèles de sélection fréquence-dépendante ont été développés pour les systèmes d'auto-incompatibilité sporophytiques. Nous avons voulu vérifier les attendus théoriques sur le **polymorphisme allélique au locus S en population naturelle**. Nous nous sommes ainsi posés les questions suivantes : Dans la même petite population naturelle d'*A. halleri*, combien y a-t-il d'allèles au locus S ? Quels sont leurs fréquences ? Quel modèle de sélection permet d'expliquer les fréquences observées ? Quel est le patron de **relation de dominance** entre allèles observés ? A-t-il une influence sur les fréquences observées au locus S ? Quels sont les effets de la **dérive génétique** sur ces attendus ?

Chapitre 3 : Existence et importance du fardeau génétique lié au locus S

L'existence d'un **fardeau génétique** lié au locus S ayant été suggéré à plusieurs reprises, nous avons voulu caractériser l'importance de ce phénomène. Nous avons **modélisé l'accumulation de mutations délétères associées** au locus d'auto-incompatibilité. Grâce à des expériences de pollinisations forcées, nous avons tenté de mettre en évidence expérimentalement **l'existence de ce fardeau**. Nous avons mesuré l'importance relative de ce fardeau, ainsi que de la **dépression de consanguinité** génomique pour une espèce auto-incompatible. D'autre part, nous avons testé la corrélation entre l'intensité du fardeau génétique lié et le **niveau de dominance** de l'allèle au locus S associé.

Chapitre 4 : Evolution de la dominance au locus S

Les modèles précédents qui traitent de la sélection fréquence-dépendante au locus S présentent la dynamique d'allèles d'auto-incompatibilité à niveau de dominance fixé. Cependant, nous avons pu observer des variations de la dominance de certains allèles d'auto-incompatibilité. L'évolution possible de la dominance avait été évoquée par Brennan (2006), comme une réponse adaptative possible à la limitation en partenaire compatible imposé par un nombre réduits d'allèles au locus S. Nous avons donc voulu tester théoriquement si **l'évolution de la dominance** au locus S pouvait être sélectionnée, et si oui dans quelles conditions. Nous avons ainsi cherché à comprendre les patrons complexes de dominance observés, et tenter d'identifier si ces patrons sont dus à des contraintes biochimiques et mutationnelles exercées sur les allèles S ou à des pressions de sélection.

Chapitre 1 - Auto-incompatibilité et structuration de la variabilité génétique dans l'espace

L'auto-incompatibilité conditionne les possibilités de reproduction entre individus issus de population naturelle. Elle implique une contrainte sur la dynamique de ces populations.

J'ai souhaité traiter deux aspects de cette question : (1) Comment l'auto-incompatibilité conditionne-t-elle les patrons de reproduction entre individus au sein d'une population naturelle ? (2) Quelle est l'influence de la sélection fréquence-dépendante sur la structuration génétique spatiale neutre et au locus S en population naturelle ?

1. Auto-incompatibilité et patrons de reproduction en population naturelle

L'auto-incompatibilité est un mode de reproduction qui empêche l'autofécondation et certains croisements allogames. Les patrons de reproduction d'espèces auto-incompatibles doivent donc être influencés par cette limitation en partenaires. Des études sur des populations d'espèces menacées suggèrent ainsi que la limitation en allèles au locus S pourrait empêcher la formation de graines, et conduire à une disparition des populations à faibles densités (Wagenius *et al.*, 2007).

Cependant, les patrons de reproduction en population naturelle sont aussi largement conditionnés par des paramètres écologiques, tels que la capacité de dispersion du pollen et des graines ou la densité de la population, c'est pourquoi il nous a paru important de réaliser une caractérisation fine des patrons de reproduction dans une population naturelle d'une espèce auto-incompatible, *A. halleri*.

Cette étude a fait l'objet de la publication ci-jointe : *High paternal diversity in the herbaceous perennial herb Arabidopsis halleri despite functional self-incompatibility, clonal reproduction and spatially-restricted pollen dispersal.* **Llaurens V.**, Castric, V., Austerlitz, F., Vekemans X., soumis à Molecular Ecology.

High paternal diversity in the herbaceous perennial herb *Arabidopsis halleri* despite functional self-incompatibility, clonal reproduction and spatially-restricted pollen dispersal

Llaurens V.*, Castric V.*, Austerlitz F. +, Vekemans X.*

*Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8016

Université des Sciences et Technologies de Lille - Lille1

F-59655 Villeneuve d'Ascq cedex

France

+ CNRS, Laboratoire Ecologie Systématique et Evolution, UMR 8079, Orsay, F-91405; Univ Paris-Sud, Orsay, F-91405; AgroParisTech, Paris, F-75231

Keywords: multiple paternity, mating patterns, paternity analysis

Corresponding author: Xavier Vekemans

Laboratoire de Genetique et Evolution des Populations Vegetales, UMR CNRS 8016

Universite des Sciences et Technologies de Lille - Lille1

F-59655 Villeneuve d'Ascq cedex

France

Email : xavier.vekemans@univ-lille1.fr

Running title: Multiple paternity in a SI species

Abstract

The number of sires fertilizing a given dam is a key parameter of the mating system in species with spatially restricted offspring dispersal, since genetic relatedness among maternal sibs determines the intensity of sib competition. In flowering plants, the extent of multiple paternity is determined by factors such as floral biology, properties of the pollen vector, selfing rate, spatial organization of the population, and genetic compatibility between neighbors. To assess the extent of multiple paternity and identify ecological factors involved, we performed a detailed study of mating patterns in a small population of a self-incompatible clonal herbaceous species, *Arabidopsis halleri*. We sampled and genotyped 364 individuals and 256 of their offspring at 12 microsatellite loci and jointly analyzed the level of multiple paternity, pollen and seed dispersal, and spatial genetic structure. We found very low levels of correlated paternity among sibs (r_p=0.076) indicating high multiple paternity. Our estimate of the outcrossing rate was 98.7%, suggesting functional self-incompatibility. Patterns of pollen dispersal indicated that the pollen dispersal distribution was significantly restricted (mean effective pollen dispersal distance: 4.42 m) but long-distance pollination occurred and immigrating pollen was rare (10% of all pollination events). Patterns of genetic structure indicated little extent of clonal reproduction, and a low but significant spatial genetic structure typical for a self-incompatible species. Overall, in spite of a functional self-incompatibility system and in spite of restricted pollen dispersal, the multiple paternity was very high, a result that we interpret as a consequence of high plant density in this population.

Introduction

In species with spatially restricted offspring dispersal, genetic relatedness among maternal sibs determines the intensity of sib competition (Yasui 1998). Hence, the extent of multiple paternity, i.e. the number of sires fertilizing a given dam, is a key parameter of the mating system in these species (Ellstrand 1984). In plants, the extent of multiple paternity is mainly determined by several factors, including spatial and temporal patterns of flower display and the characteristics of the pollination vector, modulated by the spatio-temporal structure of flowering individuals in the populations and the ability of pistils to filter out individual pollen grains, including self-pollen, according to their origin (Smouse and Sork 2004, Hardy et al. 2004).

Correlated paternity among outcrossed maternal sibs can arise when fertilization events are not independent. This may arise for instance when pollen carried by an animal pollinator leads to fertilization of all seeds of a fruit, as shown for instance in species where pollen grains are carried together in pollinia (Broyles & Wyatt 1991). However, a pollinator may also deposit a mixed pollen load simultaneously in a single visit due to the mechanism of pollen carryover (Marshall and Ellstrand 1985; Campbell 1998). As pointed out by Hardy *et al.* (2004), correlated paternity also arises when the same pollen sources are used in independent pollination events, because the number of pollen donors for a given maternal plant is limited.

Despite the availability of several methodological approaches to estimate the extent of multiple paternity, the available data are still limited. A few trends in relation to life history characteristics have been deduced, however, from the limited data available. Correlated paternity is generally higher in herbaceous than in tree species; and among tree species, correlated paternity is overall lower in wind-pollinated than in animal-pollinated species, and decreases with increasing conspecific density (Smouse and Sork 2004, Hardy et al. 2004).

Overall however, little is known on the factors influencing multiple paternity in flowering plants, particularly for herbaceous species.

In particular the impact of self-incompatibility (SI) in this context has received only limited attention. Self-incompatibility systems are genetic systems that prevent selfing of hermaphrodite plants as well as cross-fertilization between individuals sharing identical self-incompatibility phenotypes. The effect of self-incompatibility is twofold: on the one hand, self-incompatibility prevents self-fertilization and thus enforces outcrossing pollen sources. On the other hand, self-incompatibility limits the mate availability and thus also constrains multiple paternity. Hence the net effect of SI on patterns of multiple paternity is difficult to predict. Several studies have provided evidence that in self-incompatible species, limited availability of compatible pollen sources may result in overall pollen limitation (Wagenius et al. 2007). Theoretical investigations have shown that quantitatively this process will depend on (1) the type of SI; (2) the number of self-incompatibility alleles maintained within populations; and (3) the diversity of pollen sources landing on a given pistil, which is influenced notably by the extent of geitonogamous pollination and of population isolation restricting the availability of external pollen sources (Byers and Meagher 1992, Vekemans et al. 1998).

Because clones typically occur close to each other, vegetative reproduction is another potentially important factor determining paternal diversity when pollen dispersal is restricted. In several species with SI systems, high level of vegetative propagation had been observed, causing the occurrence of patches of identical genotypes. In *Prunus avium* populations, Stoeckel *et al.* (2006) found several clone groups, with more than half of the trees belonging to one of these clone groups, with clone size up to 34 individuals on an overall sample of 247 adult trees. Because most pollination events occur over short spatial distance, this may severely impact multiple paternity of plants within these patches. In species with reduced

number of SI alleles, this reduction in the number of mates may at the extreme cause a substantial decrease in their seed-set. Such low seed-set in clonal species with SI has been reported in *Illicium floridanum* (Thien *et al.* 1983), in *Filipendula rubra* (Aspinwall & Christian 1992), *Hymenoxyx acaulis* (Demauro 1993), and the cactus *Stenocereus eruca* (Ricardo *et al.* 2006).

A third factor influencing multiple paternity is the composition of the pollen deposited on the pistils. This composition is determined by spatial genetic structure, including spatial arrangement of individuals across the population, patterns of pollen and seed dispersal, population density and population isolation (Robledo-Arnuncio & Austerlitz, 2006).

In this paper we performed a full mating system assessment and investigated the extent of multiple paternity in a small isolated population of *Arabidopsis halleri* (Brassicaceae), a species with sporophytic self-incompatibility. To evaluate the relative importance of factors determining multiple paternity, we assessed the role of (1) the self-incompatibility system, (2) the extent of clonality and (3) spatial genetic structure, including the distribution of pollen dispersal distances within the population.

Material and methods

Population sampling and DNA extraction

A. halleri is a European insect-pollinated herbaceous species growing both in mountainous areas and on heavy-metal polluted sites, close to industrial areas for example. It is a perennial species which survives in winter as a rosette and is able to develop stolons for asexual reproduction. *A. halleri* has been described as self-incompatible (Clauss & Koch 2006), but no study has been performed yet to estimate the functionality of this system in a natural population. We focused on a single isolated population located in Nivelle (France, 50° 28' North, 3° 27' East). With the aim of performing a paternity analysis, we intended to sample

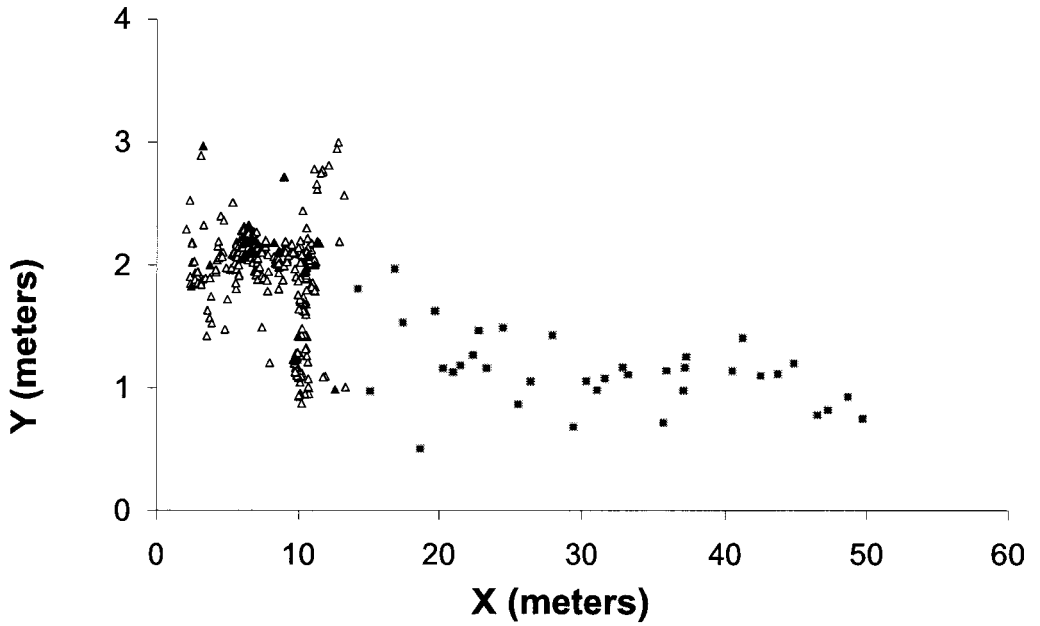


Figure 1: Spatial location of sampled individuals. Triangles represent individuals from patch A (exhaustively sampled) and squares individuals from patch B (non-exhaustively sampled). Solid triangles represent individuals from which seeds were collected for the paternity analysis.

exhaustively all individuals from the population. Because of the very high plant density, an exhaustive sampling could be performed in only about one half of the population (Part A, Fig. 1), where 328 individuals were sampled by collecting leaves from each ramet that visually appeared as a distinct individual. In the other part of the population, we sampled leaves from 36 regularly spaced individuals (Part B, Fig.1) to investigate spatial genetic structure throughout the population. Each sampled individual was mapped according to two spatial axes (x;y). For each of 22 maternal plants from part A (Fig. 1), we sampled five fruits, collected the seeds, and grew the offspring for eight weeks in a greenhouse. All leaf samples (adult + offspring populations) were oven-dried at 55°C, and DNA was extracted using the Nucleospin® Multi-96 Plant extraction kit from Macherey-Nagel®. Overall, we collected DNA from 364 adult and 256 offspring individuals.

Genotyping microsatellite markers

Based on a preliminary polymorphism survey using a small species-wide sample, twelve microsatellite loci were chosen to investigate genetic structure in this population. We amplified separately by PCR five microsatellite loci: ATH, ELF3, GC16, LYR133 and LYR417 (see Table 1 for primer sequences). For these loci, each forward primer contained a 5'-tail of 19 bp homologous to the consensus M13 forward sequence (Oetting *et al.* 1995). The reaction mixture (15µL) contained 20 ng DNA, 1X buffer (Applied biosystem®), 2 mM of $MgCl_2$, 200 µM of Fermentas® dNTP mix, 200 µg/mL of BSA, 0.2 µM of each microsatellite primer, 0.15 µM of M13 primer (fluorescence-labelled with either IRD-700 or IRD-800) and 0.025 U/µL of *Taq* polymerase (Amplitaq DNA polymerase, Applied biosystem®). The amplification was carried out 5 min at 95°C, 8 cycles of 30 s at 95°C, 45 sec. at 50°C, 40 s at 72°C, then 30 cycles of 30 s at 95°C, 20 s at Tm°C (Tab. 1), 40 s at 72°C, and one cycle of 7 min at 72°C and performed in MJ research PTC 200 thermocycler®.

Table 1: Primer sequences and annealing conditions used for amplifying microsatellite loci

Locus		Primer sequence (5'➔ 3')	Annealing temperature (Tm)
ATH	F	TCTATCAACAGAAACGCACCGAG	56
	R	CCACTTGTTTCTCTCTCTAG	
ELF3	F	CGGAAGGACTGATATACAAGC	50
	R	GTTGGGTGTTCTGAAGAT	
GC16	F	TTTTGGAGTTAGACACGGATCTG	50
	R	GTTGATCGCAGCTTGATAAGC	
LYR133	F	GTTGCTGCTGCTGATGGTT	56
	R	CAAGGAAGGCAGCAAAGAAA	
LYR417	F	AATCCCATCTCTTTCCGCTT	56
	R	GGAAGGAGAACCAACGATCA	
ATHZFPG	F	TTGCGTTTCCACATTTGTTT	50
	R	TGGGTCAATTCACATGTAGAGA	
ICE13	F	GATCCTTCACCGGGTCTTG	50
	R	GTGGTGGAGACTCTTCGAGC	
NGA112	F	TAATCACGTGTATGCAGCTGC	50
	R	CTCTCCACCTCCTCCAGTACC	
GC22	F	GGTCTAATTGCCGTTGTTGC	50
	R	GAATTCTGTAACATCCCATTTCC	
H117	F	CCGTCCTTGATCCTTGAGATTCTGAG	50
	R	CAATTCCGAAAATCATATTCATGCACC	

NGA361	F	AGGGTTTTCCCAAAGAGATGA	50
	R	TCTTGTCCTTCGATTTTAGACCA	
MDC16	F	GAGTGGCCTCGTGTAGAGAAAG	50
	R	TGTCACTCTTTTCCTCTGGTTTG	

PCR products were separated on 6% polyacrylamide gels and visualized through fluorescence of M13 primers on a Li-Cor sequencer. Size standards were run in every third lane to allow accurate band sizing.

We also amplified simultaneously by multiplex PCR seven additional loci (ATHZFPG, ICE13, NGA112, MDC16, GC22, H117, NGA361, Table 1) using primers labeled with Applied Biosystems® dyes (VIC for MDC16 and GC22, 6FAM for NGA112 and NGA361, NED for H117 and ICE13, PET for ATHZFPG and LIZ for the sizing standard). The reaction mixture (15µL) contained 20 ng DNA, 1X buffer (Applied biosystem), 2 mM of $MgCl_2$, 0.6 mM of Fermentas® dNTP mix, 200µg/ml of BSA, 0.3 µL DMSO, 0.28 µM of each primer for H117, NGA361 and ATHZFPG, 0.1 µM for ICE13, 0.2 µM for GC22 and NGA112, 0.42 µM for MDC16 and 0.125 U/µL of *Taq* polymerase (Amplitaq DNA polymerase, Applied Biosystems®). Amplifications were performed on a MJ research PTC 200 thermocycler with the following conditions: 5 min at 95°C, 35 cycles of 40 s at 95°C, 40 s at 50°C, 1 min at 72°C, and one cycle of 7 min at 72°C. Samples were loaded on a 16-capillaries ABI 3100 sequencer, and genotypes were determined with the software Genemapper (Applied Biosystems®).

Estimation of outcrossing rate and patterns of correlated paternity

We used the mixed mating model based on progeny arrays analysis (Ritland & Jain 1981) to directly estimate the maximum likelihood outcrossing rate t_m using the software MLTR (Ritland 2002). The MLTR software was also used to compute the multilocus correlation of paternity. Estimates of multiple paternity are sometimes expressed in total number of distinct fathers per fruit or per maternal plant (e.g. Ellstrand 1984, Campbell 1998), irrespective of the relative contribution of each father, or alternatively as an effective number of fathers, N_{ep}, which is the number of fathers necessary to cause the observed level of

correlated paternity assuming equal mating chances and independent mating events (Ritland 1989, Austerlitz and Smouse 2001). The latter estimate is obtained as the inverse of the correlation of outcrossed paternity among maternal sibs, or proportion of full-sib pairs among pairs of maternal sibs (Hardy et al. 2004). We used the statistics r_p, which corresponds to the proportion of full-sibs among outcrossed sibs. Significance of theses statistics were tested by performing 1000 bootstraps across maternal families using the method described in Fernandez-M. & Sork (2005). Bootstrap estimates of t_m were compared to 1, corresponding to strict outcrossing, and r_p values were compared to 0.

To determine whether spatially close dams were sired by genetically similar sets of sires, we assessed the impact of the spatial location of the dams on the paternal composition of the pollen load siring them. This was done with the method developed by Hardy *et al* (2004), which consists in computing Loiselle *et al*'s (1995) kinship coefficient between paternal genes for pairs of maternal family and in correlating these kinship coefficients with the pairwise spatial distance between dams. A Mantel test was then used to check whether spatially close plants had a higher probability of sharing the same sires. The software Spagedi (Hardy & Vekemans 2002) was used to perform these computations.

Analyses of pollen dispersal based on progeny genotypes

We obtained an estimate of the distribution of effective pollen dispersal in one generation (i.e. pollen dispersal having led to a fertilization event) by performing a paternity assignment analysis based on the maximum-likelihood method detailed in Marshall *et al.* (1998). Given the multilocus genotype of an offspring, of its mother, and of a set of potential fathers, the method assigns paternity to the father with the highest likelihood. Confidence in assignment was determined by computer simulations with the software CERVUS v. 3.0 (Marshall *et al.* 1998). Briefly, assignment was considered unambiguous if the likelihood of

the most likely father was significantly higher than that of the second most likely father. Significance was determined by simulating a population that included the genotyped individuals and a fixed proportion of unsampled individuals. These simulations were then used to compute for each offspring the log-likelihood difference between the two most likely potential fathers (Δ value) and determine the threshold value of Δ above which the most likely father was indeed the true father with a given confidence level. We chose the relaxed confidence level (i.e. 80%), which is commonly used in the literature when one of the two parents is already known and allowed for 5% genotyping error in likelihood calculations. The average exclusion probability, given genetic information on the second parent, was calculated using the formula of Chakravarti & Li (1983).

For each unambiguous paternity assignment, we used the location of mother and father to compute the effective pollen dispersal distance. We tested the observed mean of effective pollen dispersal distances against a distribution of mean pollen dispersal distances obtained under random mating. To obtain this distribution we simulated replicate progeny samples by randomly drawing a father from the overall adult population for each progeny from each mother sampled in the field. For each of 100 replicates, we computed the mean pollen dispersal distance. The hypothesis of random mating was then tested by comparing the observed mean pollen dispersal distance with the simulated distribution and computing a *P*-value. We estimated the axial standard deviation of pollen dispersal by the formula: $(\sigma_{pollen})^2 = ½ M$, where M is the mean of squared father-mother distances. We estimated the diversity of fathers siring the seed on each mother by $P_{full\text{-}sibs}$, which is the mean proportion of pairs of progeny of a given mother sharing the same father.

The effective pollen dispersal distances estimated above depend on both the intrinsic capacity of pollen to disperse (the probability that a pollen grain travels a given distance also called pollen dispersal curve or pollen dispersal "kernel"), and on the relative spatial location

of individuals. To disentangled these two factors, we used a maximum likelihood method coined "the spatially-explicit mixed mating model" (SEMM, Oddou-Muratorio *et al.* 2005), to jointly estimate the pollen dispersal kernel, the selfing rate (*s*) and the proportion (*m*) of pollen from outside the sampled population. This method, which stems from the neighborhood model (Burczyk *et al.* 2002), allows estimation of the kernel by removing the effect of the position of adults on the landscape on the effective pollen dispersal curve. This hypothesis may not be fully relevant for an insect pollinated species because insects may stay longer in denser patches. Nevertheless, in other insect-pollinated species like *Sorbus torminalis* (Oddou-Muratorio *et al.* 2005) or partially insect-pollinated species like *Brassica napus* (Devaux *et al.* 2007), this model gave realistic estimations on pollen dispersal distance. The method assumes that the pollen dispersal kernel is part of a given family of dispersal kernels. We performed the estimates here for five different families (normal, exponential, exponential power, geometric and Weibull), which are described in detail in Austerlitz *et al.* (2004).

Given the spatial position and genotypes of mothers and potential fathers, and offspring genotype, the method estimates the most likely value of the parameters of the dispersal kernel. We estimated for each family the best fitting parameters (*a* for the one-parameter families, *a* and *b* for the two-parameter families), along with the migration (*m*) and selfing (*s*) rates. We used the Akaike information criterion (AIC) to estimate the best-fitting family. We allowed for the same genotyping error rate (5%) as above, assuming that an allele could be replaced by any other in case of genotyping error. For each family of dispersal kernels, we deduced from the estimated dispersal parameters (*a* and *b*), the mean distance of pollen dispersal (δ), the estimated axial variance of pollen dispersal (σ^2) using the formulas given in Austerlitz *et al.* (2004).

Genetic structure in the adult population:

The software FSTAT (Goudet 1995) was used to compute the observed and expected heterozygosities as well as the inbreeding coefficient F_I on the overall sample of adult plants, and to test for departure from Hardy-Weinberg genotypic proportions.

The extent of vegetative propagation among the exhaustively sampled individuals (group A) was quantified by considering that plants from different ramets with identical multilocus genotypes at all microsatellite loci belonged to the same genet. To assess the risk of error, we estimated the probability of finding twice the same multilocus genotype at random within the population (probability of identity, P.I., (Waits *et al.* 2001)) using the software GIMLET (Valiere 2002). We estimated clone size as the maximal distance between two ramets with identical genotype. Note that rosettes presenting signs of physical attachment or situated very close to each other (<0.5 cm) were not collected and considered as the same individual, thus biasing upwards the estimation of clone size. To avoid genetic replicates, and because clones were tightly clustered (see below), a single ramet was randomly chosen from each genet for further analyses.

Patterns of seed and pollen dispersal were first characterized indirectly using spatial genetic structure. Using a Mantel test (Mantel 1967), we tested for spatial genetic structure by comparing the matrix of pairwise spatial distances between individuals (on a log-scale) with that of pairwise genetic relatedness estimated using the kinship coefficient of Loiselle *et al.* (1995). The Mantel test was performed with 1,000 permutations on genetic distances with the software Genetix v4.02 (Belkhir *et al.* 2004). Under a model of isolation by distance in two dimensions, a negative linear relationship is expected between spatial distances (log-scale) and kinship. Assuming migration-drift equilibrium, the slope of the linear regression is expected to be inversely proportional to the variance of gene dispersal, and directly proportional to the level of spatial genetic structure (Hardy & Vekemans 1999). Accordingly,

we performed a linear regression between pairwise spatial distances and kinship coefficients and estimated the slope (b). We computed the axial variance of gene dispersal (σ^2) according to the following formula (Rousset 2000; Vekemans & Hardy 2004): $\sigma^2 = -(1-F_N)/(4\ \pi\ D_e\ b)$, where F_N is the average kinship coefficient between neighbouring individuals and D_e is the effective population density. This axial variance of gene dispersal (σ^2) depends both on seed and pollen dispersal (Crawford 1982). The variation of kinship between individuals according to spatial distance, $F(d)$ was estimated as the average kinship coefficients among pairs of individuals in each of ten distances classes. The upper boundary distances of the ten distance classes were 0.63 m; 1.15 m; 1.85 m; 2.73 m; 3.51 m; 4.21 m; 5.38 m; 7.63 m; 20.66 m; 47.56 m, which were determined so as to include the same number of pairs of individuals in each distance class. The value of F_N was taken as the average in the first distance class. In part A of the population, where sampling was exhaustive, the observed density was D_o=14 individuals/m^2. In absence of information on effective population density (D_e), we computed σ^2 assuming a ratio D_e/D_o equal to 1/5 or 1/10. The kinship coefficient computations were performed using the software Spagedi (Hardy & Vekemans 2002).

To infer the relative rates of pollen and seed dispersal from the spatial genetic structure, we applied the method developed by Heuertz *et al.* (2003) to identify the most likely combinations of axial standard deviations of pollen (σ_{pollen}) and seeds (σ_{seed}). To that end, we simulated spatial genetic structure in a theoretical population resembling as closely as possible to group A with a program kindly provided by O. Hardy which had already been used in Heuertz *et al.* (2003). We considered a strictly outcrossing two-dimensional population with the same density as in the observed population (14 individuals/m^2). In absence of direct field estimation, we used an annual survival rate based on greenhouse observations (= 0.8, personal observations), which takes into account that *A. halleri* is a perennial species. In this theoretical population, we considered the same number of

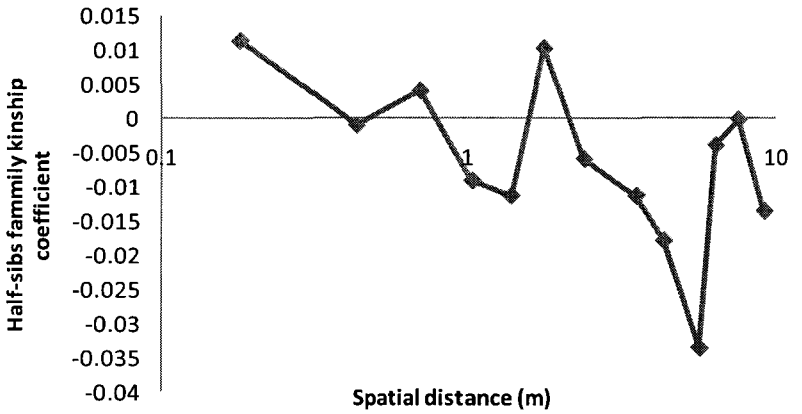


Figure. 2: Correlation between pairwise kinship coefficients between inferred paternal genes within half-sibs families and spatial distance (log scale).

microsatellite loci and the same number of alleles per locus as observed in our study population. Simulations were started with identical population frequencies for all alleles at every locus. Pollen and seed dispersal followed a normal distribution with parameters σ_{pollen} and σ_{seed}, respectively. Simulations of gene dispersal were made with values of σ_{pollen} and σ_{seed} ranging from 0 to 3 meters, and the σ_{pollen} - σ_{seed} combinations providing the best fit to the observed slope were retained as the most likely. We performed 1000 simulations for each [σ_{pollen}; σ_{seed}] combination. The slope observed in Nivelle was compared to the 1000 ranked slopes independently simulated for each [σ_{pollen}; σ_{seed}] combination. The parameter P, described by (Sokal & Rohlf 1995) was used to determine which [σ_{pollen}; σ_{seed}] pair gives the best fit to the data. $P = (n+1)/(N+1)$ if $n< N/2$ and $P=1-(n+1)/(N+1)$ if $n>N/2$ because it is a bilateral test; where n is the rank of the observed slope among the ordered simulated slopes, and N the number of replicates for each parameter set (here N=1000).

Results

Outcrossing rate and patterns of correlated paternity

The maximum likelihood multilocus estimate of outcrossing rate obtained under the mixed mating model (Ritland & Jain 1981) was very high, t_m = 0.987 +/- 0.052 (S.E.). Bootstrap analysis showed that the t_m statistic was not significantly different from 1, suggesting that selfing may be strictly avoided in this population assuming early-acting inbreeding depression is low. The multilocus estimate of the correlation of outcrossed paternity within maternal families (r_p) was significantly different from 0, but was strikingly low (r_p = 0.076 +/- 0.039 (S.E.)), meaning that only 7.6 % of sibs are actually full-sibs despite multiple sampling from single fruits. We observed a decrease of the pairwise kinship coefficients of the inferred paternal genes of maternal sibs with spatial distance (Fig. 2). The Mantel test on this auto-

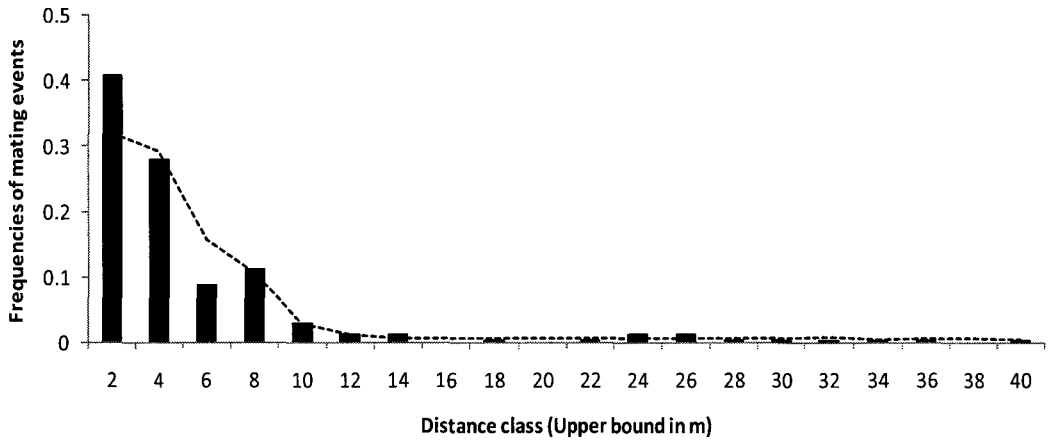


Figure 3: Distribution of effective pollen dispersal obtained by paternity analysis: frequency of mating events for each class of distance between mates. The broken line represents the numbers of effective pollen dispersal by distance class expected under random mating.

correlation was marginally significant (P=0.061), suggesting that plants located closer to each others were pollinated by more similar fathers than distant plants.

Analyses of pollen dispersal based on progeny genotypes

For the paternity analysis, the cumulative exclusion probability over all loci was 0.961. The proportion of confidently assigned paternity was exceptionally high (94% = 241 out of 256 offspring), thus providing solid ground for a detailed analysis of the distribution of realized pollen dispersal distances (Fig. 3). The observed mean pollen dispersal distance was significantly smaller than the mean pollen dispersal distance obtained under the hypothesis of random mating (P-value =0.024), thus indicating that pollen dispersal was spatially-restricted. Most successful pollination events occurred over limited distances, but successful population-wide pollen dispersal events were also observed occasionally. The mean effective pollen dispersal distance was 4.42 +/- 6.3 m (S.D.), giving an estimate of the axial standard deviation of pollen dispersal, σ_{pollen} = 5.44 m. Analysis of the diversity of paternal parents siring offspring from a given maternal individual showed a strikingly high diversity. Of the 26.7 (± 21.9 S.D.) offspring harvested per maternal plant on average, the estimated number of distinct fathers was 22.4 (± 16.4 S.D.), meaning that 84% of pollination events per maternal plant were accomplished by distinct fathers. On average, the fraction of maternal sibs that share the same father was very low, $P_{full\text{-}sibs}$ = 4.6 %. This remained true even within single fruits, since the mean number of seeds per fruit observed was 2.7 (± 1.8 S.D.) and the mean number of distinct fathers per fruit was 2.6 (± 1.7), which indicates that 98 % of pollination events per fruit were accomplished by distinct fathers.

Results from the analysis of the SEMM are summarized in Table 2 for different pollen dispersal kernels. For different shapes of the pollen dispersal kernel, we consistently estimated that only ~10% of pollen came from outside the sampled population (m = 0.099-

Table 2: Log-likelihood and parameter estimates from the spatially explicit mixed mating model for each of four shapes of the pollen dispersal kernel.

Curve	**-log likelihood**	***P***	**AIC**	***m***	***s***	***a***	***b***	***δ***	***σ***
Uniform (Random Mating)	-2498.0	2	5000.0	0.064	0				
Gaussian	-2486.4	3	4978.7	0.125	0	2.774	**2**	2.458	1.961
Exponential	-2482.9	3	4971.8	0.128	0	1.474	**1**	2.948	2.553
Exponential-Power	-2478.1	4	4964.2	0.103	0	0.001	0.209	142.10	233.168
Geometric	-2477.6	4	4963.2	0.099	0	0.098	0.989	∞	
Weibull	-2478.9	4	4965.9	0.114	0	5.144	1.421	4.678	4.064

P: number of parameters estimated, **AIC**: Akaike criterion, ***m*** : immigration rate, ***s*** : selfing rate, **1-*m*-*s*** : allo-fecundation rate in the population, ***a*** : scale parameter, ***b*** : shape parameter, ***δ*** : mean distance of pollen dispersal, **σ** : axial standard deviation of pollen dispersal.

Table 3: Single locus and multilocus statistics describing genetic diversity, departure from Hardy and Weinberg genotypic proportions, and spatial genetic structure in the adult population of *Arabidopsis halleri*.

Locus	Range in allele size (bp)	Number of alleles	H_o	H_e	F_I	Slope (*b*)
AthZFPG	132-170	3	0.3571	0.4831	0.279*	-0.0091
GC22	184-188	2	0.5287	0.4905	-0.074	-0.0012
H117	124-132	3	0.3291	0.3327	0.018	-0.0030
MDC16	113-115	2	0.3656	0.4162	0.124*	-0.0024
nga112	176-202	4	0.4217	0.4772	0.137*	-0.0040
nga361	123-127	3	0.2104	0.3099	0.296*	-0.0052
GC16	162-174	4	0.5774	0.6299	0.085*	-0.0156
ATH	169-206	7	0.5194	0.5602	0.079*	-0.0062
Lyr133	163-169	4	0.4889	0.5455	0.121*	-0.0049
lyr417	203-247	4	0.5845	0.5661	-0.028	-0.0092
Elf3	320-350	5	0.5472	0.6394	0.166*	-0.0288
Mean	-	3.7	0.4108	0.4542	0.106*	-0.0099

Number of alleles, observed heterozygosity (H_o) and expected heterozygosity (H_e), inbreeding coefficient (F_I) and exact test of departure from Hardy Weinberg genotypic proportions (* : $P < 0.05$), slope (*b*) of the regression of pairwise kinship coefficients on the logarithm of geographical distance.

0.128, Table 2). Consistent with the direct estimate from the progeny array, the SEMM model estimate of selfing rate (*s*) converged to 0% for each model of dispersal kernel. The random mating model (corresponding to uniform dispersal distribution) had an *AIC* much higher than all the other dispersal kernel, indicating that pollen dispersal distance distribution was significantly different from random mating in this population. The pollen dispersal kernel providing the best fit to the data was obtained with the geometric distribution (*AIC*=4963.2). This distribution has a fat tail, meaning that pollen dispersal events occurred over a large range of distances, including long distances relative to the population size. The estimated value of the *b* parameter for the geometric distribution was below 3, meaning that the estimated mean distance of pollen dispersal and axial standard deviation (σ_{pollen}) were infinite (Austerlitz *et al*, 2004). This result also observed in other cases (see e;g. Devaux *et al*, 2007) means mostly that the scale of the population is too small to estimate with high precision the shape of the pollen dispersal curve. However the fact that the second best-fitting dispersal curve, the exponential power, has a very low estimated *b* conforts the conclusion of fat-tailed pollen dispersal.

Genetic structure in the adult population

Statistics describing genetic diversity in the total adult population (group A and B) are given in Table 3. Locus Ice13 appeared monomorphic in Nivelle, and was thus discarded from further analyses. For the other loci, we observed from two to seven alleles per locus, with observed (H_o) and expected (H_e) heterozygosities ranging from 0.318 to 0.573, and from 0.331 to 0.644, respectively. The average inbreeding coefficient (F_I) in the adult population was 0.106 (95% CI: [0.081; 0.117]), showing an overall deficit in heterozygotes. This overall departure from Hardy-Weinberg genotypic proportions was significant at eight of the eleven loci analysed. The high number of moderately to highly variable microsatellite loci ensured

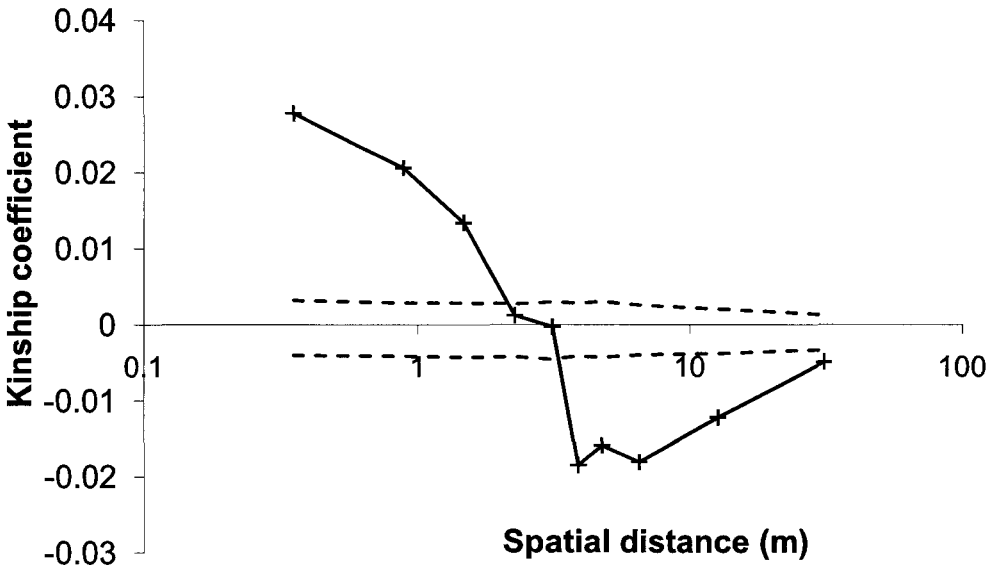


Figure 4: Correlogram showing spatial genetic structure. The y-axis represents the mean multilocus pairwise kinship coefficient between individuals; x-axis represents spatial distance intervals in meters (log-scale). The solid line represents the auto-correlogramm obtained on

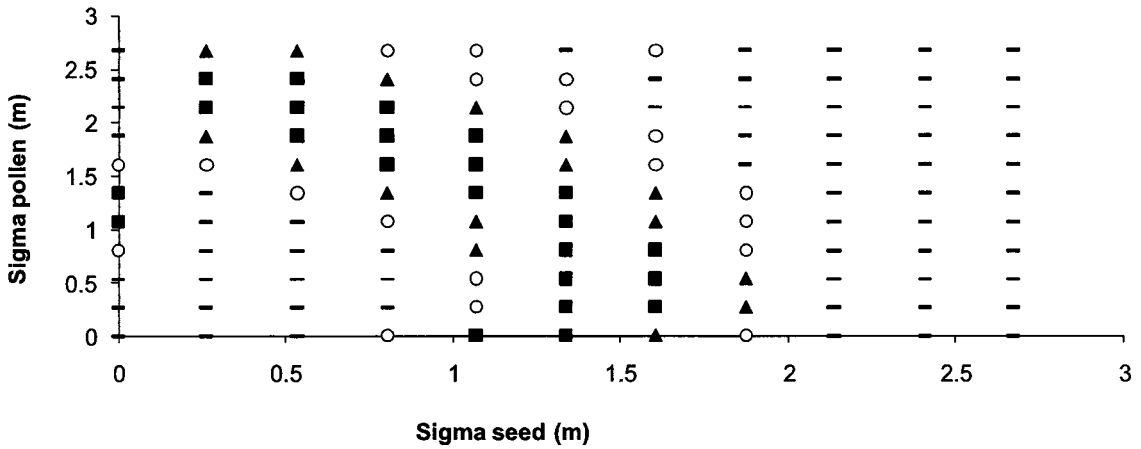


Figure 5 : Fit of the observed slope of kinship–distance plot for the statistic *P* with simulations under various combinations of pollen and seed dispersal standard deviations (σ_{seed} and σ_{pollen}) to the observed slope of kinship–distance plot in the population of *Arabidopsis halleri*. Hyphens : $P<0.05$, Circles : $0.05<P<0.2$, Triangles : $0.2<P<0.5$ and Squares : $P>0.5$

that our analysis had a high power to detect clonal reproduction: the probability of identity (P.I.) was 2.5×10^{-6}. Overall, we found multiple ramets for only 8 genets out of 355 distinct genotypes and the same genotype was never found in more than two ramets. Among those pairs, the mean distance between ramets was 0.10 m with a range of [0.004 m; 0.389 m]. Hence, the level of vegetative reproduction was rather limited in this population.

We found a significant correlation between pairwise multilocus kinship coefficients and spatial distance between individuals (Mantel test, $P < 0.001$), thus indicating a pattern of isolation by distance within the adult population. The average kinship showed an overall linear decrease up to about 8 m (Fig. 4). For eight class intervals out of ten, the average kinship coefficient between individuals was significantly different from that of random pairs of individuals obtained by permutations with a confidence interval of 95%. The slope of the linear regression (b-value) was -0.00993. Based on this slope and assuming an isolation by distance model, the indirect estimate of the average axial parent-offspring spatial distance was estimated to be $\sigma = 1.67$ or 2.36 m, assuming a ratio D_o/D_e equal to 1/5 or 1/10, respectively. The numerical simulations aimed at determining combinations of pollen and seed dispersal that best fit the observed pattern of isolation by distance suggested that the average axial pollen dispersal did not exceed 2.5 m, whereas the maximum value fitting the data for seed dispersal was 1.6 m (Fig. 5).

Discussion

We observed very high levels of multiple paternity even within single fruits in a small, isolated population of *Arabidopsis halleri*. As we detail below, our observations suggested that the effects of the self-incompatibility system and clonality on multiple paternity are minor as compared to that of pollen dispersal.

Outcrossing rate and multiple paternity

A. halleri has been described as a self-incompatible species, but formal investigation was lacking (Clauss & Koch 2006). Functional haplotypes have recently been identified and sequenced in *A. halleri* at a locus considered to be orthologous to the S-locus of the sister species *A. lyrata* (Bechsgaard *et al.* 2006, Castric & Vekemans 2007), but the level of outcrossing in a natural *A. halleri* population had never been studied. Our direct estimate of population outcrossing rate based on either the maximum likelihood method using maternal families (Ritland 2002) or the spatially-explicit mixed mating model using both maternal families and potential paternal genotypes (Oddou-Muratorio *et al.* 2005) are consistent and indicate strict outcrossing. As we genotyped offspring from seedlings which had germinated and grown enough to produce leaves, the hypothesis that some selfing actually occurred but remained masked by early-acting inbreeding depression remains possible. But since selfed progeny were successfully obtained in controlled conditions (V. Llaurens, unpublished data), the hypothesis of early-acting inbreeding depression alone could not explain the observed outcrossing rate. Our results are thus compatible with a fully functional self-incompatibility system in this population of *A. halleri*. In the closely related species *A. lyrata*, several populations with completely self-compatible individuals have been described recently, with population outcrossing rate as low as 0.2, which reveals the occurrence of substantial variation in the functionality of the SI system (Mable *et al.* 2005). It will therefore be important to extend the present mating system analysis to additional natural populations to determine whether such variation in SI also occurs in *A. halleri*.

The level of multiple paternity per mother plant observed in our study appeared to be remarkably high, Overall, the fraction of siblings sharing the same father observed here (0.046) was one of the lowest reported for an herbaceous self-incompatible species. In *Centaurea corymbosa*, this fraction was 0.196 (Hardy *et al.* 2004). The level of correlated paternity found in the self-incompatible perennial forb *Rutidosis leptorrhychoides* populations

ranged from r_p=0.37 to r_p=0.65 (Wells & Young 2002), about 10 times more than what we observed here. In the self-incompatible shrub *Grevillea iaspicula* McGill the average paternal diversity within open-pollinated sibships is low (r_p=0.31–0.54) (Hoebee & Young 2001). The level of correlated paternity varies from r_p =0.05 and r_p =0.64 in populations of the self-incompatible *Centaurea solstitialis* (Sun & Ritland 1998). In herbaceous species, the multiple paternity is also lower as observed here: for instance, in the herbaceous bee-pollinated species, *Mimulus guttatus*, Ritland (1989) found a correlated paternity (r_p =0.2) superior than observed here. The observed level of multiple paternity in the Nivelle population was indeed comparable to that typically observed for wind pollinated plants (Smouse and Sork 2004). Our observation that multiple paternity is also very frequent even within single fruits, suggests that pollinators' activity is remarkably high in this population. This high level of multiple paternity suggests that the SI system may not limit seed set here, as each mother plant sampled in this population seems to be pollinated by a high number of pollen sources, a large fraction of which being apparently compatible. The high rate of multiple paternity could indicate a high number of self-incompatibility alleles in this population. These results suggest that individuals of this *A. halleri* population have high mate availability, and it would now be interesting to perform similar investigations on different species with very different mate availability patterns (such as for instance species with either a low number of self-incompatibility alleles, with low density or very few pollinators) to determine which factors affect mate availability most strongly.

Vegetative propagation

Because clones typically occur close to each other, vegetative reproduction is a potentially important factor determining paternal diversity when pollen dispersal is restricted. Excluding the individuals presenting signs of physical attachment, the extent of vegetative

propagation seemed to be very limited in this population. Indeed, we found that only 2% of the multilocus genotypes exhibit more than a single ramet, with a distance between ramets not exceeding 0.39 m. In another population of *A. halleri* (Auby, France), about 22% of multilocus genotypes showed multiple ramets, with a maximum distance among ramets of a given clone below 1 m (Van Rossum *et al.* 2004). This suggests that clonal reproduction is particularly limited in the Nivelle population, but the discrepancy between these results could be due to the following three reasons: (1) genotypic diversity in the Auby population may have been underestimated because of the low number of microsatellite loci used in that study (5 loci as compared to 11 in the present study); (2) the sampling protocols differed slightly, i.e. rosettes were collected every 10 cm in the (Van Rossum *et al.* 2004) study, irrespective of whether traces of physical attachment were present or not, while we deliberately discarded such samples; (3) plant density was very contrasted in the two populations (higher in Nivelle than in Auby on average), a factor that was indeed reported to negatively affect the extent of clonal reproduction in *A. halleri* (Van Rossum *et al.* 2004). Our results indicate that in spite of a leptokurtic pollen dispersal, clonal reproduction is unlikely to negatively affect seed set in this *A. halleri* population.

Shape of pollen dispersal

The shape of pollen dispersal has a great influence on the multiple paternity because it determines the number of different pollen sources able to pollinated a given pistil of the population.

The spatially-explicit mixed mating (SEMM) model suggests that the observed pollen dispersal distribution shape was different from that expected under random mating and highly leptokurtic, suggesting an excess of pollination at short distances and occasionally at long distances relative to the population size.

The average estimate of effective pollen dispersal distances (*i. e.* the distance observed between the two parents of each seed) by the direct method of paternity analysis (σ_{pollen} = 4.72 m) is consistent with the σ_{pollen} found for real pollen dispersal distances estimated by the mixed mating model. Given the high density of the population, this would lead to a large neighborhood size, suggesting a high number of partners available for each plant in this population.

The paternity analysis confirmed that the population was effectively isolated, as only 10% of pollen came from outside the population. This estimate is indeed conservative as only half of the population was exhaustively sampled. This value is rather low as compared to the literature: pollen immigration rate in tree species is generally much higher, with 40%, 65%, and 69% estimates in *Sorbus terminalis* (Oddou-Muratorio *et al.* 2005), *Quercus robur* and *Quercus petraea* (Streiff *et al.* 1999), respectively although a lower value has been reported in a small isolated population of the tree species *Pinus sylvestris*, (4.3% of all observed matings, Robledo-Arnuncio & Gil 2005). Intermediate values have typically been reported in herbaceous plants: from 0% in *Heloniopsis orientalis* (Miyazaki & Isagi 2000), 8.2% or 17.9%, depending on the year of sampling *in Raphanus sativus* (Ellstrand & Marshall 1985), 38.1% in the herbaceous self-incompatible *Beta vulgaris* (Fenart *et al.* 2007), up to 54.3% in *Iris fulva* (Cornman *et al.* 2004). The pollen immigration rate also depend on the isolation of the population: Hoebee *et al.* (2007) observed that in *Sorbus torminalis* immigrant pollen is about 30% for continuous populations whereas it could be about 4% in a small isolated population. Altogether, these results suggest that the population studied is rather isolated. *A halleri* is indeed a heavy-metal tolerant species, growing preferentially on heavy-metals polluted soils outside its native distribution area. This population is growing on an embankment made from soil taken from a polluted area. Thus the low rate of immigrant

pollen could be due partly to the isolation of this population, in relation to the ecological requirement of the species.

Spatial genetic structure

More than any other factor, we identified the spatial location of individuals in the population as the main determinant of the composition of pollen siring a given maternal plant. Spatial genetic structure analysis revealed a typical pattern of isolation by distance due to restricted dispersal. The extent of within population spatial genetic structure as estimated by the slope of the regression of pairwise kinship coefficients against the log of spatial distance between individuals (Sp =0.010) was lower than observed in the other *A. halleri* population studied by Van Rossum *et al.* (2004) (Sp = 0.003), but was indeed strikingly similar to the average value obtained for 17 self-incompatible species, Sp = 0.013 +/- 0.08 (Vekemans & Hardy 2004). This is thus consistent with the general pattern of low extent of spatial genetic structure in strictly outcrossing species, in comparison with selfing species (mean Sp = 0.143) and species with a mixed mating system (mean Sp = 0.037).

In addition to functional self-incompatibility, the low isolation by distance observed in this population could also be due to non equilibrium conditions. The direct estimate of realized pollen dispersal distances (paternity analysis, σ_{pollen} = 4.72 m) was indeed higher than the maximum estimate based on spatial genetic structure (σ_{pollen} < 2.5 m.), which assumed drift/mutation equilibrium. The latter estimate thus integrates pollen dispersal over the long term, whereas the paternity estimate depends only on the last generation. Alternatively, the discrepancy could be due to the fact that only the initial part of the pollen dispersal distribution, i.e. with a shape similar to a negative exponential distribution, has an impact on the isolation by distance pattern, whereas the distal part, with a rather flat shape, contributes

little to the genetic structure. This would be consistent with the observation that the autocorrelogram flattens at the highest spatial scales (Fig. 4).

Despite a low extent of spatial genetic structure, the indirect estimate of gene dispersal is rather low, $\sigma = 2.36$ m, with a maximum value of $\sigma_{pollen} \approx 2.5$ m and of $\sigma_{seed} \approx 1.5$ m. The estimation of σ_{pollen} and σ_{seed} could also be biased to some extent by the normal distribution chosen for the pollen and seed dispersal curve in the model of Heuertz *et al.* (2003), which may appear unrealistic because it provided low fit for the pollen dispersal kernel in the mixed mating model. We suggest that the apparent contradiction between restricted dispersal distance and low extent of spatial genetic structure could be due to the high density of this population ($\approx$ 14 ind/m^2). Indeed it has been shown that within species, plant density has a paramount influence on patterns of spatial genetic structure and pollen dispersal (Vekemans & Hardy 2004); (Oddou-Muratorio *et al.* 2004). This is because although pollen travels short distances, it represents a large diversity of potential paternal individuals: for instance, assuming $\sigma_{pollen} = 2.5$ m 14 ind/m^2, an individual is surrounded on average by 275 potential fathers at a distance within a circle of radius σ_{pollen}, thereby decreasing the probability that the father and mother would be genetically related. Altogether, these results indicate that pollen and seed dispersal were restricted, but as the density found in this population was high, a large number of potential mates are available over short distance, allowing multiple paternity to be high.

Conclusions

Using genotypes of adult plants and maternal progenies at 12 microsatellite loci, we obtained very high estimates of paternal diversity within and among fruits of maternal individuals, despite spatially-restricted pollen dispersal, population isolation, and an apparently fully functional self-incompatibility system. We suggest that this result may be due

to the very high plant density and/or high pollinator activity in this population. This result suggests that the constraint on mating patterns due to self-incompatibility must be weak in this population, and more generally depends on the ecological context.

Acknowledgments

This work was supported by an ATIP research grant from the life science department of the CNRS, and by a PhD grant to VL from CNRS and Région Nord-Pas de Calais. The authors wish to thank Pierre Saumitou-Laprade and Cécile Godé for the initial trials of microsatellite primers in *A. halleri*, Virginie Legros and Adeline Courseaux for technical support for the genotyping, and Claire Feutrie, Eric Schmitt and Robert Dron for the plants care in greenhouse. They also thank Valeria Ruggiero and Sylvain Billiard for useful comments on the manuscript.

Author information box

VL performed this study during her PhD on the population genetics of the self-incompatibility system in *Arabidopsis halleri*. She uses empirical and modeling approaches to investigate evolutionary forces and constraints influencing allele frequencies at the sporophytic self-incompatibility locus. Together with VC, they work in the self-incompatibility research team at the University of Lille, headed by XV. This team aims at characterizing population genetics and molecular evolution properties of sporophytic self-incompatibility systems in plants. FA has a long standing interest in modelling the impact of demographic and adaptive processes on genetic diversity and in the estimation of dispersal from genetic data.

References

Aspinwall N, Christian T (1992) Clonal Structure, Genotypic Diversity, and Seed Production in Populations of Filipendula rubra (Rosaceae) from the Northcentral United States. *American Journal of Botany* **79**, 294-299.

Austerlitz F, Dick CW, Dutech C, *et al.* (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* **13**, 937–954.

Austerlitz F, Smouse PE (2001) Two-generation analysis of pollen flow across a landscape. II. Relation between Fft, pollen dispersal and inter-females distance. *Genetics* **157**, 851-857.

Bechsgaard JS, Castric V, Charlesworth D, Vekemans X, Schierup MH (2006) The transition to self-compatibility in Arabidopsis thaliana and evolution within S-haplotypes over 10 Myr. *Molecular Biology and Evolution* **23**, 1741-1750.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la genetique des populations.

Broyles SB and R Wyatt 1991 Effective pollen dispersal in a natural population of Asclepias exaltata: the influence of pollinator behavior, genetic similarity, and mating success. *American Naturalist* **138**: 1239-1249.

Burczyk J, Adams WT, Moran GF, Griffin AR (2002) Complex patterns of mating revealed in a Eucalyptus regnans seed orchard using allozyme markers and the neighbourhood model. *Molecular Ecology* **11**, 2379-2391.

Byers DL, Meagher TR (1992) Mate availability in small populations of plant species with homomorphic sporophytic self-incompatibility. *Heredity* **68**, 353-359.

Campbell DR (1998) Multiple paternity in fruits of Ipomopsis aggregata (Polemoniaceae). *American Journal of Botany* **85**, 1022-1027.

Castric V, Vekemans X (2007). Evolution under strong balancing selection: how many codons determine specificity at the female self-incompatibility gene *SRK* in Brassicaceae? BMC Evolutionary Biology in press.

Chakravarti A, Li CC (1983) The effect of linkage on paternity calculations. In: *Inclusion Probabilities in Parentage Testing* (ed. Walker R), pp. 411-422. American Association of Blood Banks, Arlington, Virginia.

Clauss MJ, Koch MA (2006) Poorly known relatives of Arabidopsis thaliana. *Trends in Plant Science* **11**, 449-459.

Cornman RS, Burke JM, Wesselingh RA, L. AM (2004) Constrasting genetic structure of adults and progeny in a Louisiana iris hybrid population. *Evolution* **58**, 2669-2681.

Crawford TJ (1982) What is a population? In: *The 23rd symposium of the british ecological society* (ed. Shorrocks B). Blackwell Scientific Publications, Leed.

Demauro MM (1993) Relationship of breeding system to rarity in the lakeside daisy (*Hymenoxys acaulis var. glabra*). *Conservation Biology* **7**, 542-550.

Devaux C, Lavigne C, Austerlitz F, Klein EK (2007) Modelling and estimating pollen movement in oilseed rape (*Brassica napus*) at the landscape scale using genetic markers. *Molecular Ecology* **16**, 487–499.

Ellstrand NC (1984) Multiple Paternity within the Fruits of the Wild Radish, Raphanus-Sativus. *American Naturalist* **123**, 819-828.

Ellstrand NC, Marshall DL (1985) Interpopulation Gene Flow by Pollen in Wild Radish, *Raphanus sativus*. *The American Naturalist* **126**, 606-616.

Fenart S, Austerlitz F, Cuguen J, Arnaud J-F (2007) Long distance pollen-mediated gene flow at a landscape level: the weed beet as a case study. *Molecular Ecology* **accepted**

Fernandez-M. JF, Sork VL (2005) Mating patterns of subdivided population of the andean oak (Quercus humboldtii Bonpl., Fagaceae). *Journal of Heredity* **96**, 635-643.

Goudet J (1995) Fstat version 1.2: a computer program to calculate Fstatistics. *Journal of Heredity* **86**, 485-486.

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**, 145–154.

Hardy OJ, Vekemans X (2002) A versatile computer program to analyse spatial genetic structure at the individual or population model. *Molecular Ecology Notes* **2**, 618-620.

Hardy OJ, González-Martínez SC, Colas B, *et al.* (2004) Fine-scale genetic structure and gene dispersal in Centaurea corymbosa (Asteraceae). II. Correlated paternity within and among sibships. *Genetics* **168**, 1601-1614.

Heuertz M, Vekemans X, Hausman J-F, Palada M, Hardy OJ (2003) Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology* **12**, 2483-2495.

Hoebee SE, Young AG (2001) Low neighbourhood size and high interpopulation differentiation in the endangered shrub Grevillea iaspicula McGill (Proteaceae). *Heredity* **86**, 489-496.

Hoebee SE, Arnold U, Duggelin C, *et al.* (2007) Mating patterns and contemporary gene flow by pollen in a large continuous and a small isolated population of the scattered forest tree Sorbus torminalis. *Heredity* **99**, 47-55.

Loiselle BA, Sork VL, J. N, C. G (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* **82**, 1420-1425.

Mable BK, Robertson AV, Dart S, Di Berardo C, Witham L (2005) Breakdown of self-incompatibility in the perennial Arabidopsis lyrata (Brassicaceae) and its genetic consequences. *Evolution* **59**, 1437-1448.

Mantel N (1967) The detection of disease clustering and generalized regression approach. *Journal of cancer research* **27**, 209-220.

Marshall DL, Ellstrand NC (1985) Proximal Causes of Multiple Paternity in Wild Radish, Raphanus sativus. *American Naturalist* **126**, 596-605.

Marshall TC, Slate J, Kruuk L, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular ecology* **7**, 639-655.

Miyazaki Y, Isagi Y (2000) Pollen flow and the intrapopulation genetic structure of Heloniopsis orientalis on the forest floor as determined using microsatellite markers. *Theoretical and Applied Genetics* **101**, 718-723.

Oddou-Muratorio S, Demesure-Musch B, Pelissier R, Gouyon P-H (2004) Impacts of gene flow and logging history on the local genetic structure of a scattered tree species, Sorbus torminalis L. Crantz. *Molecular Ecology* **13**, 3689–3702.

Oddou-Muratorio S, Klein EK, Austerlitz F (2005) Pollen flow in the wildservice tree, Sorbus torminalis (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent-offspring analysis. *Molecular Ecology* **14**, 4441-4452.

Oetting WS, Lee HK, Landers DJ (1995) Linkage analysis with mutiplexed short tandem polymorphism using infrared fluorescence and M13 tailed primers. *Genomics* **30**, 450-458.

Ricardo CT, Corrado CA, Mandujano MC, Molina-Freaner F (2006) Reproductive consequences of clonal growth in *Stenocereus eruca*, a rare clonal cactus of the Sonoran desert. *Evolutionary ecology* **20**, 131-142.

Ritland K, Jain S (1981) A model for the estimation of outcrossing rate and gene frequencies using n independant loci. *Heredity* **47**, 35-52.

Ritland K (1989) Correlated Matings in the Partial Selfer Mimulus-Guttatus. *Evolution* **43**, 848-859.

Ritland K (2002) Extensions of models for the estimation of mating systems using *n* independent loci. *Heredity* **88**, 221-228.

Robledo-Arnuncio JJ, Gil L (2005) Patterns of pollen dispersal in a small population of Pinus sylvestris L. revealed by total-exclusion paternity analysis. *Heredity* **94**, 13-22.

Robledo-Arnuncio JJ, Austerlitz F (2006) Pollen dispersal in spatially aggregated populations. *American Naturalist* **168**, 500-511.

Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology* **13**, 58-62.

Smouse PE, Sork VL (2004) Measuring pollen flow in forest trees: an exposition of alternative approaches. *Forest Ecology and Management* **191**, 21-38.

Sokal RR, Rohlf FJ (1995) *Biometry: the principles and practice of statistics in biological research* 3rd edn. W. H. Freeman & Co, New York.

Stoeckel S, Grange J, Fernandez-Manjarres JF, *et al.* (2006) Heterozygote excess in a self-incompatible and partially clonal forest tree species — Prunus avium L. *Molecular Ecology* **15**, 2109-2118.

Streiff R, Ducousso A, Lexer C, *et al.* (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl. *Molecular Ecology* **8**, 831-841.

Sun M, Ritland K (1998) Mating system of yellow starthistle (Centaurea solstitialis), a successful colonizer in North America. *Heredity* **80**, 225-232.

Thien LB, White DA, Yatsu LY (1983) The Reproductive Biology of a Relict-Illicium floridanum Ellis. *American Journal of Botany* **70**, 719-727.

Valiere N (2002) GIMLET : a computer program for analysing genetic individual identification data. *Molecular Ecology Notes* **2**, 377-379.

Van Rossum F, Bonnin I, Fénart S, *et al.* (2004) Spatial genetic structure within a metallicolous population of *Arabidopsis halleri*, a clonal, self-incompatible and heavy-metal-tolerant species. *Molecular Ecology* **13**, 2959-2967.

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

Vekemans X, Schierup MH, Christiansen FB (1998) Mate Availability and Fecundity Selection in Multi-Allelic Self- Incompatibility Systems in Plants. *Evolution* **52**, 19-29.

Wagenius S, Lonsdorf E, Neuhauser C (2007) Patch aging and the S-Allee effect: breeding system effects on the demographic response of plants to habitat fragmentation. *The american naturalist* **169**, 383-397.

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations : cautions and guidelines. *Molecular Ecology* **10**, 249-256.

Wells GP, Young AG (2002) Effetcs of seed dispersal on spatial genetic structure in populations of *Rutidosis leptorrhychoides* with different levels of correlated paternity. *Genetical research Cambridge* **79**, 219-226.

Yasui Y (1998) The 'genetic benefits' of female multiple mating reconsidered. *Trends in Ecology & Evolution* **13**, 246-250.

2. Sélection fréquence-dépendante au locus d'auto-incompatibilité et structuration de la variabilité génétique dans l'espace

La sélection fréquence dépendante augmentant la migration efficace des allèles d'auto-incompatibilité et pourrait donc modifier la structure génétique spatiale des allèles S. Nous avons donc comparé la structuration de la variabilité génétique pour des marqueurs neutres et pour le locus d'auto-incompatibilité ainsi de repérer s'il existe une signature de la sélection balancée agissant au locus S. Cette étude a fait l'objet du stage de M2R de Jean-Baptiste Leducq, que j'ai co-encadré avec X. Vekemans lors de ma deuxième année de thèse et fait l'objet de la publication ci-jointe. *Effect of balancing selection on spatial genetic structure within continuous populations: theory and empirical evidence from the self-incompatibility locus in Arabidopsis halleri.* Leducq, J. B., **Llaurens V.**, Hardy O. J., Vekemans X., en préparation pour le journal *Heredity.*

Effect of balancing selection on spatial genetic structure within continuous populations: theory and empirical evidence from the self-incompatibility locus in *Arabidopsis halleri*

Leducq J.-B.[1], Llaurens V.[1], Castric V.[1], Hardy O.[2] and Vekemans X.[1]

[1] Laboratoire de génétique et évolution des populations végétales, UMR CNRS 8016, Bâtiment SN2, Université des sciences et techniques de Lille - Lille1, F-59655 Villeneuve d'Ascq, France

[2] Laboratoire éco-éthologie évolutive - CP 160/12 Université Libre de Bruxelles50 Av. F. Roosevelt, B-1050 Bruxelles, Belgique

Corresponding author: Xavier Vekemans

Email : xavier.vekemans@univ-lille1.fr

Keywords: S-locus; frequency-dependent selection; spatial autocorrelation, SPAGeDi, numerical simulations

Running title: SGS in self-incompatible plants populations

Abstract

Within plant populations, as pollen and seed dispersal are commonly spatially restricted, genetic similarity among individuals is higher among neighbouring than among more distant individuals. This pattern of isolation by distance is influenced by mating systems, like selfing avoidance which reduces the spatial genetic structure (SGS). The SGS could also be influenced by the effect of natural selection, like strong balancing selection acting on self-incompatibility systems. By numerical simulations, we investigated the SGS within populations at neutral loci and the self-incompatibility locus (S-locus) under different conditions of seed and pollen dispersal, immigration rate and allelic diversity at the S-locus. In some circumstances, we found a significantly lower extent of SGS at the S-locus than at neutral markers, whatever the self-incompatibility system considered. In analogy to previous theoretical results on population structure at self-incompatibility genes, we interpreted these results as evidence for a higher effective dispersal of alleles at the S-locus as compared to neutral loci. We also investigated the SGS for the S-locus and for 11 microsatellite markers in three natural populations of *Arabidopsis halleri*, a species with sporophytic self-incompatibility. The observed patterns of SGS were highly variable, a result that we interpret because of different ecological and demographical conditions encountered in the three studied populations.

Introduction

Pollen and seed dispersal within and between plant populations are generally spatially restricted, and, through the action of genetic drift, this generates patterns of spatial genetic structure (SGS) that have been characterized theoretically as "isolation by distance" patterns (Malécot, 1950). These patterns of spatial genetic structure are marked by the fact that genetic similarity among individuals is higher among neighbouring than among more distant individuals. Many empirical studies in plant populations have investigated SGS, notably through the application of spatial autocorrelation methods (Heywood, 1991), and it has been shown that the extent of SGS is highly dependent on species' characteristics such as life form and mating system, as well as on local features such as population density (Vekemans & Hardy, 2004). For instance, mating systems characterized by mechanisms preventing self-fertilization experience on average higher pollen dispersal than mixed mating systems, and the resulting patterns of SGS are thus less pronounced (Doligez *et al.*, 1998).

Although patterns of SGS are mainly caused by interaction between gene flow and genetic drift, natural selection can also influence the spatial distribution of genetic variation. For instance, local adaptation under diversifying selection will tend to increase SGS at selected loci, as compared to neutral genes, whereas the reverse is true for balancing selection, a family of selection models leading to the maintenance of long-lasting polymorphisms within populations (Takahata, 1990), that is expected to counteract the effect of genetic drift on patterns of SGS (Slatkin, 1973; Sokal *et al.*, 1989). In plants, self-incompatibility systems are one of the best known examples of genes evolving under balancing selection (Castric & Vekemans, 2004). Self-incompatibility (SI) is a genetic system that impedes self-fertilization of hermaphrodite individuals due to the recognition and active rejection of self-pollen in the course of reproduction. As the SI reaction may also involve individuals sharing the same phenotype at SI genes, SI is also commonly viewed as a general system to avoid

consanguineous matings (Charlesworth & Charlesworth, 1987). In flowering plants, two different types of SI systems are found: gametophytic self-incompatibility (GSI), the most widespread system in which pollen specificity is determined gametophytically, and sporophytic self-incompatibility (SSI), in which pollen specificity is determined sporophytically by proteins expressed in the anther tissues (Takayama & Isogai, 2005). (Wright, 1939) showed that strong selection favouring rare alleles is acting on SI systems through an increase in pollen access to compatible mates, a mechanism corresponding to negative frequency-dependent selection. Polymorphism data from natural populations on the genes known to be involved in determination of SI in several plant families are in general agreement with expectations derived from theoretical models for GSI and SSI (Lawrence, 2000; Castric & Vekemans, 2004).

Empirical and theoretical studies on the effect of balancing selection on genetic structure at the locus controlling SI (S-locus) have mostly addressed the issue of the interaction between selection and migration on population subdivision. It has been shown that as a result of this interaction, migrant alleles at the S-locus that are rare in the recipient populations will be favoured over resident alleles, a process generating an increase in "effective migration rate" of selected alleles as compared to neutral alleles, and thus differentiation at the S-locus should remain low even under very restricted migration (Schierup *et al.*, 2000; Muirhead, 2001). These predictions have been rigorously tested only rarely in empirical studies, but the results quite convincingly showed that the S-locus has a lower population genetic structure than marker loci (Glémin *et al.*, 2005; Schierup *et al.*, 2007; Brennan *et al.*, 2006). Whether this high rate of effective migration for alleles at the S-locus also has an impact on spatial genetic structure within populations is still an open question. Brooks *et al.* (1996) showed that restricted pollen and seed dispersal within a continuous population caused an increase in the variance of allele frequencies at the S-locus

with GSI, but the effects on SGS could not be separated from effects of variation in plant size that were studied simultaneously. Neuhauser (1999) showed that the rate of loss of alleles at the S-locus, in a GSI system within a continuous population with restricted pollen dispersal, was higher as compared to a panmictic population, suggesting the occurrence of SGS for the S-locus. However, none of these theoretical studies did compare explicitly patterns of SGS expected at neutral loci versus at the S-locus. Only two empirical studies provided data on SGS within populations for both types of loci. In *Senecio squalidus*, Brennan *et al.* (2003) did not find a pattern compatible with the isolation by distance model within a single population, neither at the S-locus nor at allozyme markers. In contrast, in *Prunus avium*, Schueler *et al.* (2006) found evidence for local genetic structure at both the S-locus and microsatellite markers, but the extent of SGS was similar at both types of loci.

In this study, we performed numerical simulations to investigate the within population SGS in species with GSI or SSI with the following aims: (1) to distinguish the respective effects of strict allogamy, gametophytic and sporophytic self-incompatibility on the SGS of neutral loci; (2) to estimate the effect of balancing selection induced by GSI and SSI systems respectively, on the SGS of the S-locus as compared to unlinked neutral loci; and (3) to determine the effect of the extent of pollen and seed dispersal, of the number of alleles at the S-locus, and of the rate of gene flow into the population (immigration), on the extent of SGS at the S-locus and neutral loci. We also studied empirically SGS within three natural populations of *Arabidopis halleri* (Brassicaceae) at which SI is under SSI control. We estimated the extent of SGS at 11 microsatellite loci as well as at the S-locus, and compared the results with the predictions from the simulations.

Materials and methods

Numerical simulations

Simulations were performed using an extension of Hardy *et al* (2004)'s program. This program simulated finite populations of N hermaphroditic diploid individuals with five different mating systems: (1) selfing in proportion to pollen dispersal was allowed (relaxed allogamy); (2) selfing was avoided but there was no self-incompatible system (strict allogamy); (3) selfing and crosses between incompatible individuals are avoided through a fully functional GSI system; (4) SSI system with codominance among all S-alleles (SSI-COD); or (5) SSI system with hierarchical dominance among S-alleles (SSI-DOM). Individuals were simulated with a *S*-locus (when considering GSI or SSI) as well as ten unlinked neutral loci. All simulations were performed with a population size $N = 2500$. Plants were uniformly distributed on a 50 × 50 spatial grid, with a single individual per node. Initially, the genotype of each individual was randomly chosen for all loci. For each neutral locus, the initial number of alleles was set to six. For the *S*-locus, when considering both SSI and GSI system, the initial number of alleles was set to $n_S = 25$, a value expected under GSI for a population of 2500 individuals with a low mutation rate Vekemans & Slatkin (1994). We assumed that the migration dispersal distance of seeds and pollen follow a normal distribution, with initial parameters $\sigma_s = \sigma_p = 1$ for seeds and pollen respectively (dispersal's axial standard deviation in units of grid steps). Mutation was not considered in simulations, but in order to counteract the loss of diversity due to genetic drift, we introduced random immigration of pollen at rate m from a constant source of genotypes identical to the initial population. The immigration was initially fixed at $m = 0.06\%$ ($Nm = 1.5$ individuals per generation) corresponding to the minimal value for witch drift did not much affect allelic richness for both kind of loci in simulations for all mating systems.

To obtain genotypes of each individual i at each position in the grid at generation t, the simulation algorithm searched potential parents of i in the grid at generation t-1. The mother j was chosen in three steps: first a direction was randomly determined by drawing an angle α_{ij} in an uniform distribution; second, the distance between the mother and the offspring d_{ij} was randomly drawn from the normal distribution seed dispersal distances (with parametrer σ_s); finally the mother j was chosen as the nearest individual from the point given by the couple ($\alpha_{ij;}$ d_{ij}). The father was an immigrant from an outside population with probability m. With probability (1-m), the father was not an immigrant and it was chosen as described previously by using the mother j as the focal individual instead of the offspring i. Hence, the father k, was chosen as the individual closest to the point ($\alpha_{jk;}$ d_{jk}), with d_{jk} randomly drawn from the normal distribution of pollen dispersal distances (with parameter σ_p). If j or k were outside the grid, new coordinates were randomly taken. In the case of strict allogamy, GSI or SSI, we simulated self-incompatibility by checking whether j and k were compatible. For GSI, pollen with only one of both specificities at the S-locus was randomly chosen, and then j and k were considered compatible when the S allele of k was different than both S alleles of j. For SSI-COD, j and k were compatible only when both S alleles of k were different than both S alleles of j. For SSI-DOM, we first computed the self-incompatibility phenotype as a result of the dominance interactions of the two S-alleles of each mating partners and then checked the compatibility between the phenotype of j and k. In the case of incompatibility, a new potential father k was chosen with a probability m in the immigrant's source or 1-m in the grid and so on, until a compatible combination was found. This compatibility computation thus simulated the frequency dependent selection described in Wright (1939). When parents of i were defined, one allele of each locus from each parent was randomly taken in order to constitute the genotype of i. The process was repeated for each position in the grid, and then all

individuals of generation t-1 were replaced at once by the genotypes of generation t computed as just described.

We checked the time for stabilization of the population genetics statistics (H_e and F_{IS}) as well as the spatial genetic structure (SGS) statistics (to see bellow). For highly restricted pollen and seed dispersal, we found that F_{IS} and SGS statistics values were stabilized after about 400 generations, whereas H_e values only stabilized after about 5000 generations. Thereafter, all simulations were performed over 10000 generations to ensure that the genetic structure was in steady-state.

We then analysed the spatial genetic structure of the final populations. These analyses were carried out from a sub-sample of 1600 individuals centred on the grid, in order to minimize edge effects (Sokal *et al.*, 1997). At the end of each replicate simulation, the program computed for each locus the following statistics: the expected heterozygosity (H_e); the inbreeding coefficient (F_I); values of the pairwise kinship coefficients between individuals (F_r), using the formula by Loiselle *et al.*(1995) and mean kinship coefficients among pairs of individuals classified according to 40 inter-individual distance classes ($F_r(d)$, with d expressed in number of steps along the grid). From $F_r(d)$ values, we calculated the slope (b) of the linear regression of F_r values as a function of the logarithm of spatial distance between individuals (Hardy & Vekemans, 1999). In order to compare the spatial genetic structure of neutral loci vs S-locus and Neutral loci among different mating systems for initial parameters values tested, we computed the mean and standard deviation of the global slope b for neutral loci over 100 replicate runs, and the differences in the means were tested with a t-test. Thereafter, simulations were performed with different values of the following parameters: axial standard deviations of pollen (σ_p) and seeds (σ_s) dispersal distance compiled as σ_t so that $\sigma_t^2 = \sigma_p^2/2 + \sigma_s^2$ (σ_p in the range 0.3-8 and σ_s in the range 0.1-5), rate of immigration m in the range 0.00001%-60%, and number n_S of S-alleles in the initial population in the range 2-100.

The species

Arabidopsis halleri (Brassicaceae) is a perennial, diploid and herbaceous plant species which is highly tolerant to heavy metals like zinc and cadmium (Pauwels *et al.*, 2005). It is distributed among European heavy-metal-polluted zones and subalpin areas. It can form dense populations of clonal individuals flowering from late spring to summer. Like its sister species, *Arabidopsis lyrata*, *A. halleri* possesses a functional self-incompatibility system with sporophytic control of pollen phenotype (Clauss & Koch, 2006).

Study sites and sampling

Our study focused on three populations of *A. halleri* located respectively in Hautes-Fagnes (Belgium), Auby and Nivelle (North of France). The Auby population (France, latitude: 50.25, longitude: 3.05) is located in a site with high zinc content, due to soil pollution from local mining industry (Van Rossum *et al.*, 2004). In this very large population, we considered 210 individuals that were either sampled along a 400 m transect (134 individuals) or within a plot of a 0.5 × 3 m area (76 individuals, see Van Rossum *et al.* (2004) for a description of the sampling procedure). The Hautes-Fagnes population (latitude: 50.296, longitude: 4.600) grows in a non-polluted site and has a recent and uncertain origin with Central European affinities (Pauwels *et al.*, 2005). We sampled 162 individuals along a 400m transect (66 individuals) and within two plots of a 0.5 × 3 m area (68 and 28 individuals, respectively). The small and isolated Nivelle population (France, latitude: 50.467, longitude: 3.467) grows on embankments from heavy-metal polluted sites, approximately 40 km from the mining industry sites. In about one half of the population we sampled exhaustively all individuals. In the other half we sampled individuals along a 30m transect (364 individuals in total). In the three populations, each sampled leaf rosette was initially considered as a distinct individual.

In each population, spatial coordinates of all sampled individuals were noted. The plots and the exhaustive sample allowed us to estimate spatial genetic structure (SGS) at very short scale, whereas transect sampling were useful to estimate SGS at wider scale.

Genotyping of microsatellite markers and the S-locus

For the Nivelle population, DNA samples and genotypic data for 11 microsatellite loci (*ATH*, *ELF3*, *GC16*, *LYR133*, *LYR417*, *GC22*, *H117*, *ICE13*, *MDC16*, *NGA112* and *NGA361*) were taken from Llaurens *et al.*, (in prep.). For the Auby population, DNA samples and genotypic data from five microsatellite loci (*LYR132*, *LYR133*, *LYR417*, *GC16* and *ATH*) were already available (Van Rossum *et al.*, 2004). We genotyped six additional microsatellite loci already used in the Nivelle population (*GC22*, *H117*, *ICE13*, *MDC16*, *NGA112* and *NGA361*) by a multiplex PCR, as described in Llaurens *et al.* (in prep.). For the Hautes-Fagnes population, leaves taken from each sampled individual were dried at 65°C for 24h. DNA was extracted from 10-15 mg of dried leaf material using the extraction kits Dneasy® from Qiagen®. We genotyped 11 microsatellite loci (*GC22*, *H117*, *ICE113*, *MDC16*, *NGA112*, *NGA361*, *LYR132*, *LYR133*, *GC16*, *LYR104* and *ICE9*). For the first 6 markers (*GC22*, *H117*, *ICE113*, *MDC16*, *NGA112*, *NGA361*), we used the same multiplex PCR as in Nivelle and Auby described in Llaurens *et al.* (submitted). For *LYR132*, *LYR133*, *GC16,* we used the same PCR method as in Auby population (see Van Rossum *et al.* (2004)). We used two additional markers *LYR104* and *ICE9* kindly provided by Thomas Mitchell-Olds and Marc Hanikenne, respectively. The reaction mixture (15µL) contained 20 ng DNA, 1X buffer (Applied biosystem®), 2 mM of $MgCl_2$, 200 µM of Fermentas® dNTP mix, 200 µg/mL of BSA, 0.2 µM of each microsatellite primer, 0.15 µM of M13 primer (fluorescence-labelled with either IRD-700 or IRD-800) and 0.025 U/µL of *Taq* polymerase (Amplitaq DNA polymerase, Applied biosystem®). The amplification was carried out 5 min at 95°C, 8 cycles of 30 s at

95°C, 45 sec. at 50°C, 40 s at 72°C, then 30 cycles of 30 s at 95°C, 20 s at 50°C, 40 s at 72°C, and one cycle of 7 min at 72°C and performed in MJ research PTC 200 thermocycler®.

PCR products were separated on 6% polyacrylamide gels and visualized through fluorescence of M13 primers on a Li-Cor sequencer. Size standards were run in every third lane to allow accurate band sizing.

To determine the genotype at S-locus, we used PCR primers specifically targeted towards each S-haplotype to genotype each individual at the S-locus, thus generating data corresponding to a pattern of presence/absence for each S-haplotype. The genotyping at S-locus was already performed in the Nivelle population (Llaurens *et al.*, in prep). For the Auby and Hautes Fagnes populations, we thus used primers pairs and PCR method corresponding to the 25 S-alleles used in Llaurens *et al.* (submitted). We found a new allele amplified by the specific primer pair design for S-allele AhSRK03. We thus distinguished AhSRK03 and AhSRK28by performing a digestion on the PCR products by restriction enzyme TaqI. The reaction mixture (20µL) contained 10µL of the PCR product, 1 U of TaqI enzyme Biolabs®, 1X NE Buffer by Biolabs®, and 0.2mM of spermidine. The restriction was carried out at 65°C overnight in MJ research PTC 200 thermocycler®. Restriction products were mixed with loading dye and run at 110 V on 2% agarose gels in TBE buffer for 45 minutes. Fragments were fluorescently visualized by ethidium bromide under UV light and compared to a 100bp DNA ladder. We distinguished S-allele AhSRK03 and AhSRK28 by fragment size because digestion occurred only on AhSRK28 and not on AhSRK03.

As the genotyping method of genotyping was based on presence/absence, homozygote individuals couldn't be distinguished from heterozygote individuals for witch only one allele was detected. Individuals presenting more than 2 alleles at the *S*-locus were systematically removed (because regarded as errors of handling), while missing genotypes at the *S*-locus were taken into account in our analyses.

Data analysis

The extent of clonality among the exhaustively sampled individuals was quantified by considering that plants from different ramets with identical multilocus genotypes at all microsatellite loci belonged to the same genet. To assess the risk of error, we estimated the probability of finding twice the same multilocus genotype at random within the population (probability of identity, P.I., (Waits *et al.*, 2001)) using the software GIMLET (Valiere, 2002). All further analyses were performed on genet only. In order to assess the spatial genetic structure (SGS) in each population, we used the program SPAGeDi (Hardy & Vekemans, 2002). The standard deviation σ was estimated in zones of exhaustive sampling (patch or plots), from effective density D_e. Because effective density is difficult to estimate in natural populations, different values of D_e were tested ($D/2$, $D/4$ *and* $D/10$) (Hardy *et al.*, 2006). SGS was estimated by spatial autocorrelation analysis, performed with kinship coefficients plotted against the geographical distance separating all pairs of individuals (Loiselle *et al.*, 1995). These were calculated for 20 distance classes, so that each of them contained approximately the same number of pairs of individuals. In each population, for the 11 microsatellite loci on the one hand and for the *S*-locus on the other hand, SGS was characterized by the slope b of the profile of spatial autocorrelation, calculated for the logarithm of geographical distances ranging between σ and 20σ (Hardy *et al.*, 2006). Significances of slope and kinship coefficients per each distance class were obtained by a Mantel test after 1000 random permutations of the individuals among spatial positions. For microsatellite markers, mean and standard error of the slope were estimated using a jackknife procedure over the loci. For each population, significant difference between slope for all microsatellite loci and slope for the *S*-locus was tested using a t-test adapted to the comparison of a single observation with the mean of a sample (Sokal & Rohlf, 1995). In order to estimate the extend of SGS at neutral

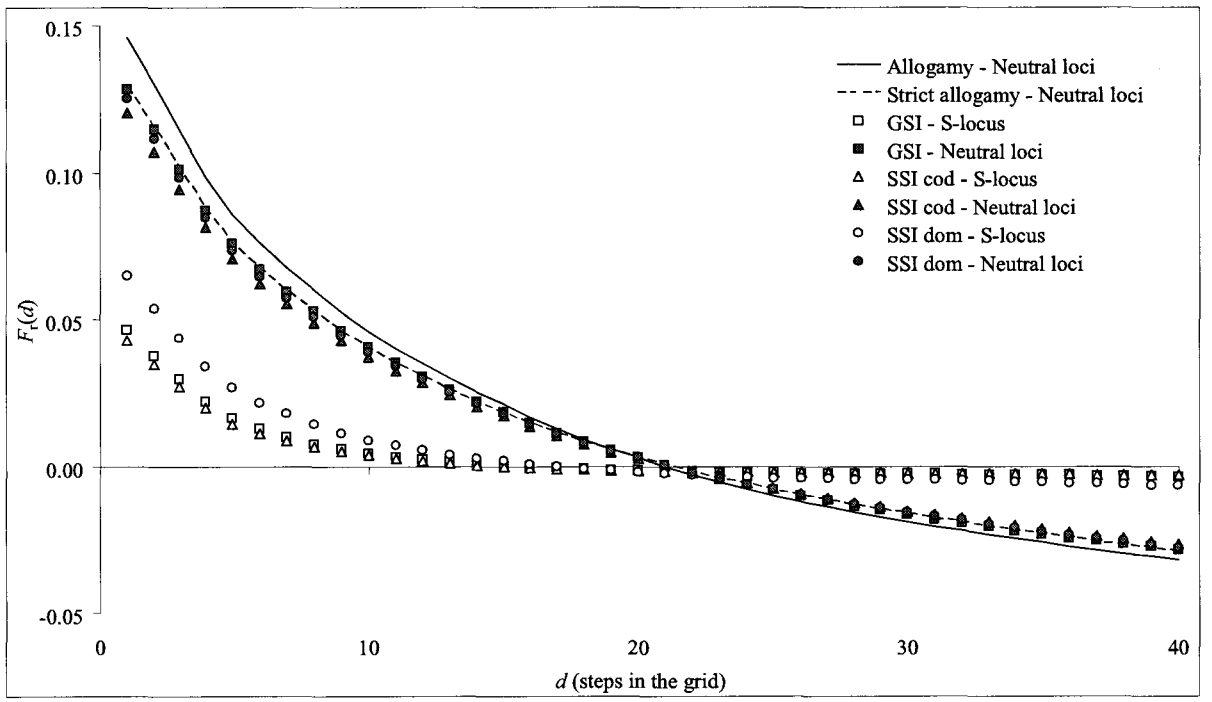


Figure 1: Autocorrelograms showing spatial genetic structure for the S-locus and for the average over 10 neutral loci for different mating systems based on 100 simulated continuous populations at 10000 generations. Population size, N = 2500 individuals; Standard axial deviation of dispersal, $\sigma_t = 1.225$ ($\sigma_s = \sigma_p = 1$); Migration rate $Nm = 1.5$, Number of S-alleles, $n_S = 25$. The y-axis represents the mean multilocus kinship coefficient between individuals, $F_r(d)$; the x-axis represents the spatial distance intervals between individuals, d, in meters (log-scale).

loci in *Arabidopsis halleri* populations, we calculated the S_p statistic for each population, estimated from $S_p = b/(F_r(1)-1)$, that is a quantification of the SGS (Vekemans & Hardy, 2004).

Results

Simulated spatial genetic structure at neutral loci and the S-locus

We first investigated the effect of mating systems on the SGS of unlinked neutral marker loci by comparing autocorrelograms representing the relationship between average kinship coefficients [$F_r(d)$] between individuals and their spatial distance (d) at neutral loci under five distinct mating system models (Fig. 1). The simulations were performed under very restricted dispersal ($\sigma_p = \sigma_s = 1$). A decreasing relationship between $F_r(d)$ and d was observed in each case, as predicted under the isolation by distance model. The slope b at neutral loci (computed using the logarithm of spatial distances) was steeper under panmixis (1), *i.e.* when selfing was allowed, as opposed to strict outcrossing without SI system (2), GSI (3), SSI-DOM (4), or SSI-COD (5) which were all very similar. The effect of the different SI systems on SGS of neutral markers was not significant. However, the slope of the correlograms was significantly higher (t-test: $p < 0.05$) under panmixis as compared to those under SSI-DOM or SSI-COD. This suggested that the main effect of the mating system on the SGS of neutral markers is through the avoidance of selfing.

We observed a significantly steeper slope for neutral markers than for the S-locus whatever the SI system (t-test: $p < 10^{-3}$). The highest slope at the S-locus was observed for the SSI-DOM model as compared to GSI and SSI-COD, suggesting that the effect strength of frequency dependent selection on the S-locus was weaker in SSI-DOM. This latter result is consistent with previous observations on SSI models by Schierup et al. (1997).

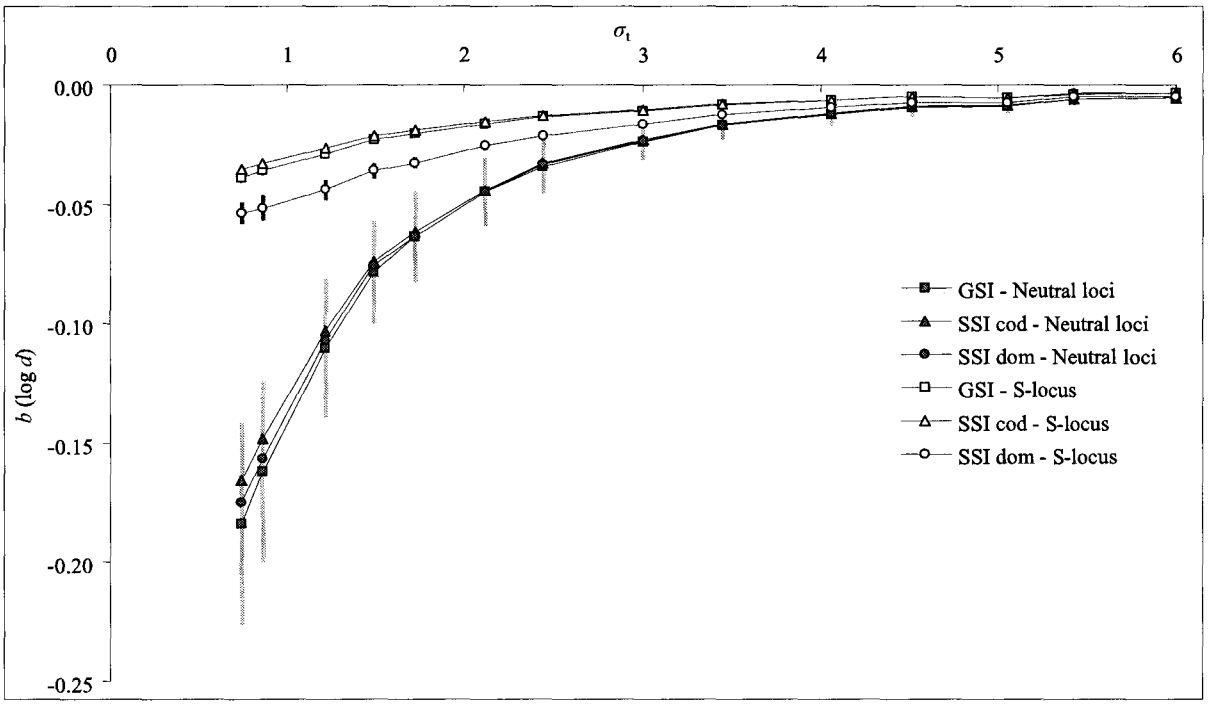


Figure 2: Effect of standard axial deviation of dispersal σ_t on slope b of the autocorrelogram for the S-locus and for 10 neutral loci based on 100 simulated continuous populations at 10000 generations, according to three self-incompatibility systems. Population size, $N = 2500$ individuals; Immigration rate, $Nm = 1.5$; Number of S-alleles, $n_S = 25$.

The effect of varying the seed and pollen dispersal capacities on the extent of SGS on both neutral markers and the S-locus was visualised by plotting the slope of the autocorrelograms as a function of the overall standard axial deviation of dispersal σ_t (Fig. 2). We tested the effect of varying the relative standard axial deviation of dispersal of the pollen σ_p and of the seeds σ_s, with identical σ_t, and found that when $\sigma_t > 0.755$, all combinations showed very similar results (data not shown), indicating that the slope of the autocorrelograms depends only on the overall σ_t when the linear regression is computed in the interval σ_t-20 σ_t (Heuertz *et al.*, 2003). Under very restricted dispersal ($\sigma_t < 2$), the slope *b* was substantially steeper for neutral loci as compared to the S locus. However, the difference decreased rapidly with increasing σ_t so that values of *b* for the two types of loci became undistinguishable for σ_t higher than about 2.5. All three SI systems showed a very similar pattern, although *b* values for the S-locus were slightly higher under SSI-DOM than under both GSI and SSI-COD models.

The immigration rate *m* also influenced the slope of autocorrelograms for neutral loci and the S-locus (Fig 3). The mean slope of autocorrelograms for neutral loci was not much affected by changes in the immigration rate below a value of $Nm = 2.5$. However, the loss of diversity at neutral loci due to genetic drift occurring under low values of *Nm* caused a substantial increase in standard deviation of *b*. Above $Nm = 2.5$, the extent of SGS for neutral loci was decreasing rapidly. For the S-locus, we observed a similar effect with lower amplitude as the slope b was substantially lower than at neutral loci. The loss of the SGS for high level of immigration is a consequence of the homogenizing effect of immigrating gene flow.

We also investigated the effect of the initial number n_S of S-alleles at the S-locus on the SGS (Fig 4). The number of S-alleles did not influence noticeably the SGS at the neutral loci. In contrast, increasing the number of S-alleles resulted in a substantial increase in the

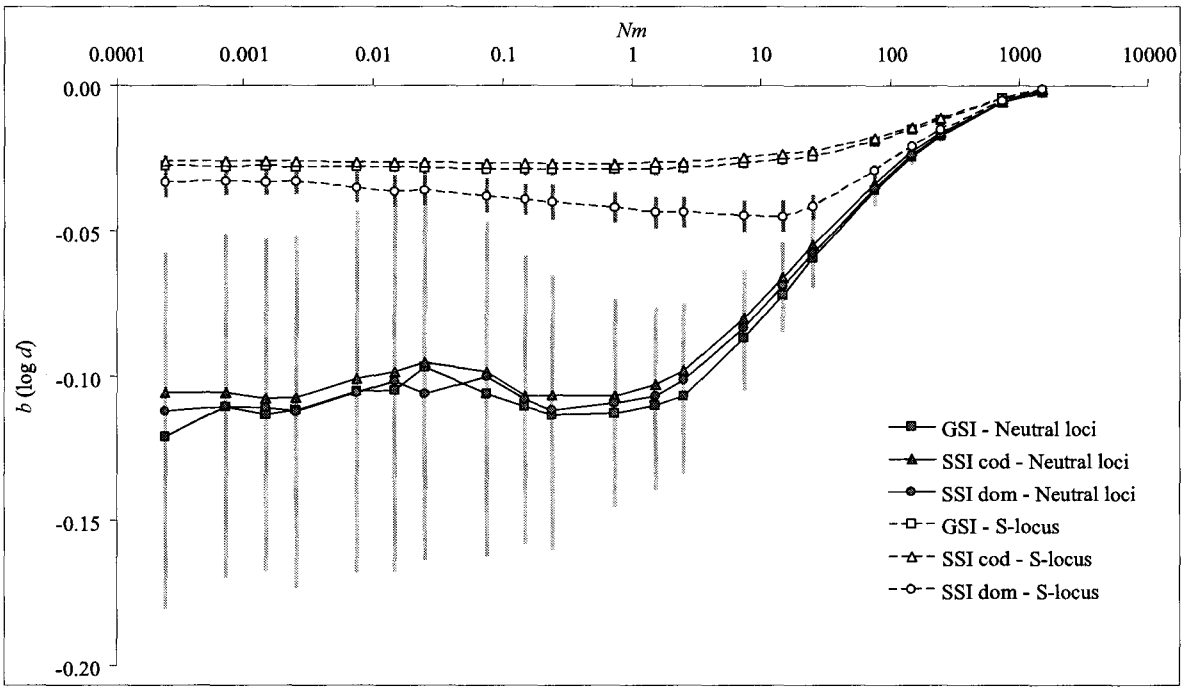


Figure 3: Effect of immigration rate *Nm* on slope *b* of the autocorrelogram for the S-locus and for 10 neutral loci based on 100 simulated continuous populations at 10000 generations, according to three self-incompatibility systems. Population size, $N = 2500$ individuals; axial standard deviation of dispersal, $\sigma_t = 1.225$; Number of S-alleles, $n_S = 25$.

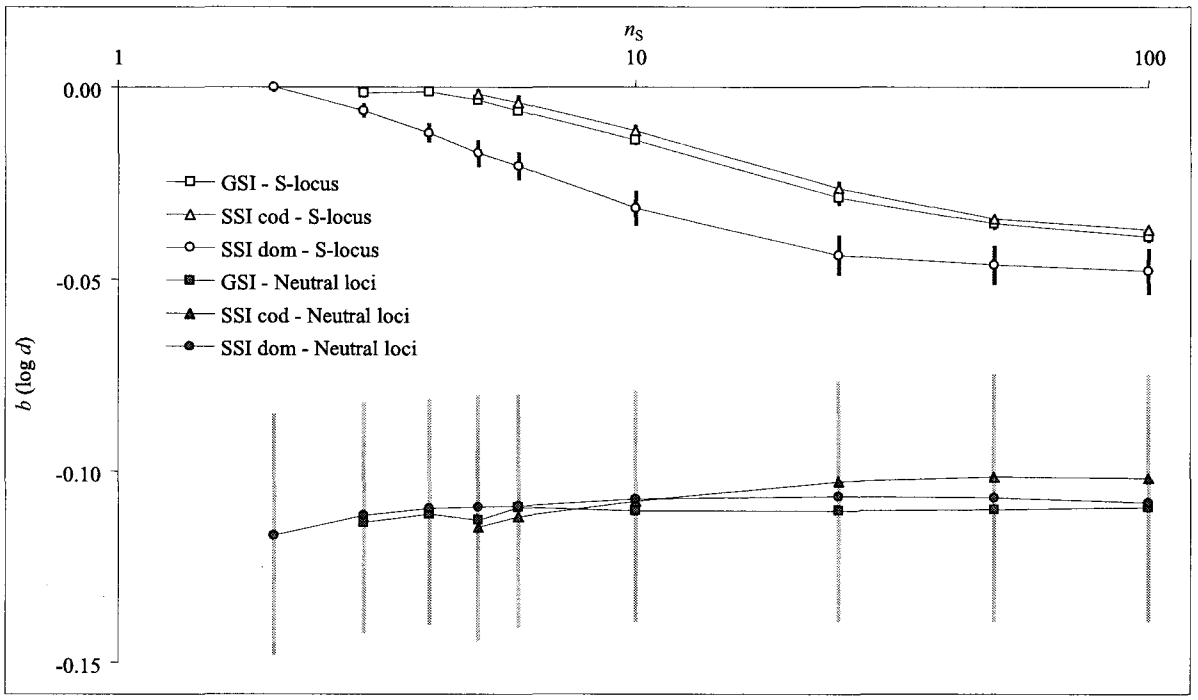


Figure 4: Effect of the number of alleles at the S-locus, n_S on slope b of the autocorrelogram for the S-locus and for 10 neutral loci based on 100 simulated continuous populations at 10000 generations, according to three self-incompatibility systems. Population size, $N = 2500$ individuals; axial standard deviation of dispersal, $\sigma_t = 1.225$; Immigration rate $Nm = 1.5$.

Table 1: Estimates of allelic diversity and observed heterozygosity H_{obs} at microsatellite loci and the S-locus; clonal ratio and efficient density D in three natural populations of *Arabidopsis halleri*. Estimates of gene dispersal standard deviation σ, slopes b of spatial autocorrelation analysis for the S-locus and neutral loci. Estimates of the S_p statistic for neutral loci. Slopes of autocorrelograms were estimated between σ and 20 σ, with considering only genets.

		Nivelle	Auby	Hautes Fagnes
Allelic diversity	S-locus (n_S)	8	9	4
	Neutral loci	3.67 ± 1.67	4.91 ± 2.30	3.45 ± 1.44
H_{obs}	S-locus (n_S)	0.82	0.77	0.69
	Neutral loci	0.46 ± 0.18	0.57 ± 0.20	0.39 ± 0.16
Clonal ratio		1.022	1.214	1.095
D (ind/m^2)		14	26	40
σ (m)		0.72	0.24	0.22
Mantel test	S-locus	0.000***	0.917	0.771
	Neutral loci	0.000***	0.000***	0.000***
b (log distance)	S-locus	-0.014		
	Neutral loci	-0.018 ± 0.010	-0.041 ± 0.020	-0.090 ± 0.049
	p-value (t-test)	0,718		
S_p (Neutral loci)		0.018	0.043	0.106

SGS of the S-locus for all SI models. The isolation by distance pattern was not detected when less than 5 alleles segregated at the S locus for both GSI and SSI-COD, whereas it was only the case with 2 alleles for SSI-DOM. For a higher number of S alleles, the SGS increased for the S-locus and remained higher for SSI-DOM than for GSI and SSI-COD. The number of S-alleles is inversely related to the strength of negative frequency-dependent selection on the S-locus (Wright, 1939), and it is shown that this effect has a strong impact on the extent of SGS.

Spatial genetic structure at neutral loci and at the S-locus in natural populations

Differences in genetic variation among populations were observed at microsatellite loci, with the highest polymorphism in the Auby population, which has a much higher size than the other populations, and the lowest polymorphism in the Hautes Fagnes population, which is highly isolated and has a recent origin. For the S-locus, we detected 4 to 9 alleles per population. In agreement with polymorphism at the neutral loci, the lowest number of alleles was observed in the Hautes Fagnes population, and the highest in the Auby population. As expected for a locus under balancing selection, allelic diversity and proportion of heterozygotes were higher for the S-locus than for marker loci.

Clonal propagation was observed in each population, with a ratio of the number of genets to the number of ramets ranging from 1.02 to 1.21. The highest level of clonal propagation was observed for the Auby population, and the lowest in the Nivelle population. Population density also varied among populations (range 14-40 ind./m^2) and was higher in Hautes Fagnes and lower in Nivelle.

For each population, we observed for microsatellite loci a decrease in kinship coefficients as the spatial distance increased (Fig. 5). This pattern of isolation by distance (IBD) at neutral loci was significant in all natural populations as revealed by the significant Mantel test (Table 1). The value of the S_p statistic, representing the extent of SGS, was lower

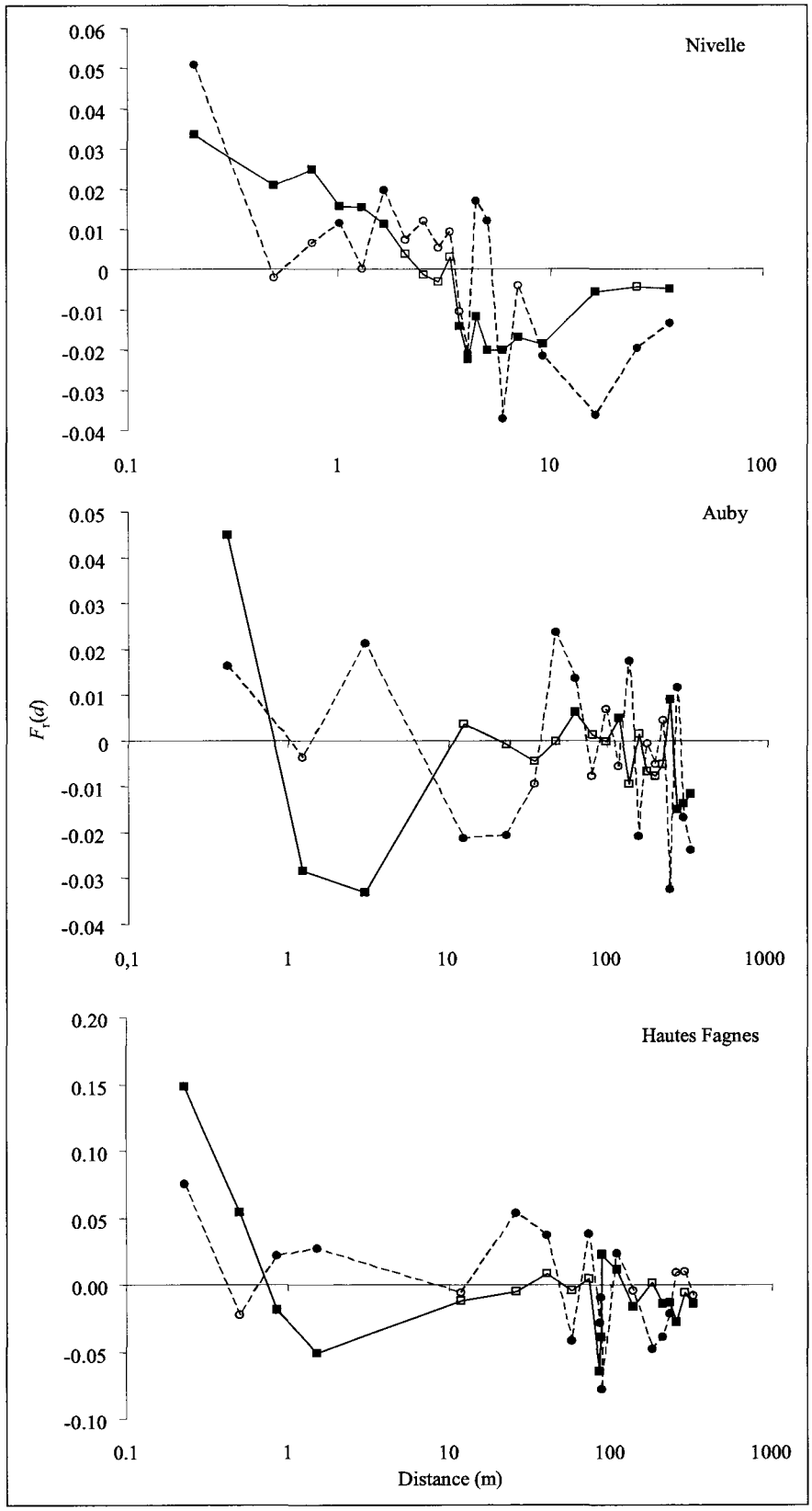


Figure 5: Spatial autocorrelation analysis of neutral loci (squares) and the *S*-locus (circles) in three natural populations of *Arabidopsis halleri* with considering only genets. Filled figures indicate a significant deviation of kinship coefficient from the average of 1000 permutations.

in Nivelle as compared to the other populations. Assuming that the spatial genetic structure at the neutral loci was at equilibrium in the three populations, and assuming that the effective population density was two times lower than the observed density ($D_e/D = 0.5$), we estimated the axial standard deviation of the distribution of gene dispersal distances σ_t. We found estimates of σ_t ranging from 0.2 to 0.7, with a value three times higher in Nivelle than in Auby and Hautes Fagnes (Table 1).

In contrast to the marked pattern of isolation by distance at microsatellite loci, the slope *b* for the S-locus was not significantly different from zero, with a positive value, in the Auby and Hautes Fagnes populations (Table 1). In the Nivelle population, the pattern of isolation by distance at the S-locus was highly significant, and the value of *b* for the S-locus was not significantly different than the value of the microsatellite loci (Table 1).

Discussion

Influence of mating system on SGS at neutral markers

The selfing avoidance for plants in a continuous population sensibly reduced its spatial genetic structure for neutral loci, although it was significant only for the very restricted dispersal parameters that we have tested ($\sigma_t = 0.612$; $p < 0.05$; data not shown). Doligez *et al.* (1998) found such a decrease of the SGS due to avoidance of selfing in simulated plants populations, but it was relevant only in populations with non-uniform distribution, where local dispersal was expected to be more restricted than in continuous populations, due to clumping of individuals. As suggested by Bos & Van der Haring (1988), GSI was found to have a barely noticeable influence on local differentiation at neutral loci compared to strict outcrossing. Our results were also very similar for SSI models. The absence of expected difference between strict outcrossing and all SI systems suggests that only the avoidance of selfing should be responsible for the lower SGS observed in a panel of SI plant species

(Vekemans & Hardy, 2004). Also, it suggests that, in contrary to the common belief, SI does not play an evolutionary relevant role in avoiding mating among close relatives. Hence we can conclude that SI is mainly selected for as a mechanism to prevent selfing.

Influence of the self-incompatibility model on SGS at the S-locus

Under restricted dispersal, our simulations showed a lower SGS at the S-locus than at neutral loci, suggesting that efficient pollen dispersal is expected to be higher for S-alleles than for neutral alleles, as already demonstrated under an island model of migration (Schierup et al. 2000, Muirhead, 2001). This effect was lower for the SSI-DOM model than for GSI or SSI-COD because the dominance relationships among S-alleles increased the number of compatible crosses in the population, thus decreasing the strength of frequency dependent selection acting at the S-locus (Schierup *et al.*, 1997). These quantitative differences among the three SI systems were also in close agreement with those obtained for the analysis of population subdivision under SI (Schierup *et al.* 2000).

Influence of dispersal, genetic drift and strength of balancing selection

Our results showed that the pollen and seed dispersal capacity, the isolation of the populations, and the allelic diversity at the S-locus are key factors influencing the detection of the signature of balancing selection on spatial genetic structure at the S-locus. Simulations showed an important decrease in the extent of SGS for both the S-locus and neutral loci with increasing local dispersal σ_t. Furthermore, under high dispersal, both types of loci exhibit no more difference in SGS. One thus expects that a significant difference in the extent of SGS between neutral loci and the S-locus would only be observed in populations or species with restricted pollen and seed dispersal. SGS patterns for both S-locus and neutral loci appeared to be affected in the same way by local dispersal than by immigration: difference in SGS

between the two types of loci was erased when the immigration rate was high. For low value of immigration the variation in the SGS of neutral markers increased because of the effect of genetic drift. The allelic richness at the S-locus (n_S) is known to influence greatly the strength of balancing selection and thus we found that it also has a large effect on the extent of SGS at the S-locus. As a result, the S-locus showed no more isolation by distance pattern for very low n_S values, and a strong SGS for high S-alleles diversity. On the contrary, the number of S-alleles did not affect much the SGS at neutral loci.

Interpretation of empirical results

In the natural populations of *A. halleri*, we could not find evidence for SGS at the S-locus in two populations (Auby and Hautes Fagnes), whereas a clear pattern of isolation by distance was observed for the microsatellite loci. In contrast, in the Nivelle population, the patterns of SGS were significant and similar at the S-locus and the microsatellite loci. We attempted to interpret these observations in the light of the theoretical results on the effects of dispersal distances and number of S-alleles on the relative patterns of SGS at neutral loci versus the S-locus. Firstly, we noted that the number of S-alleles in population Hautes Fagnes is about half that in the other two populations. In conditions of low allelic diversity at the S-locus, the strength of selection is maximised, and the simulation results showed that this would minimize the extent of SGS. This is thus in agreement with the absence of SGS observed in this population. Similarly, Van Rossum and Triest (2006) found a strong SGS at neutral loci, but no isolation by distance at the S-locus in a population of *Primula elatior*. Because the species' mating system is distyly, an heteromorphic SI system that is formally equivalent to a SSI-DOM with only two S-alleles, we could argue that the strong selection on the S-locus is responsible for the absence of a pattern of isolation by distance. However, other factors can also be responsible for observed patterns of SGS. The Hautes Fagnes population is

also by far the most isolated population of *A. halleri* as it has recently been founded outside of its natural distribution area. Hence, the absence of SGS at the S-locus can also be viewed as a consequence of a very low rate of immigration. Secondly, our indirect estimate of gene dispersal was about three times higher in the Nivelle population, as compared to the other populations. The theoretical results showed that the difference in the extent of SGS between neutral loci and the S-locus was decreasing with increasing overall level of gene dispersal. Hence, the observation of a significant pattern of SGS in this population, which is similar to that of the marker loci, may be interpreted as a result of a moderate to high level of dispersal. This interpretation would be in agreement with empirical results obtained by Schueler *et al.* (2006) in a population of *Prunus avium*, a temperate tree species with high pollen and seed dispersal and with GSI. These authors found a low but significant extent of SGS at the S-locus that was very similar to that at microsatellite loci, which can be due to the high level of dispersal. Some empirical results are more perplexing. The Auby population has intermediate properties in that the number of S-alleles is relatively high, but gene dispersal is relatively low, as compared to the other populations, so it is difficult to interpret the lack of pattern of SGS observed. Brennan *et al.*(2003) showed no significant SGS for both neutral loci and S-locus in a population of *Senecio squalidus*, but their study was based on a low number of individuals and allozyme markers. Indeed the detection of a lower spatial genetic structure at the S-locus as compared to marker loci is also made difficult by the large variances associated to the isolation by distance process, especially at marker loci that usually have lower allelic diversity than the S-locus.

Conclusion

By simulating the effect of gametophytic and sporophytic SI systems on the spatial genetic structure within plant populations, we found that the frequency dependent selection acting on

theses mating systems could lead to a lower spatial genetic structure at the S-locus than at neutral markers. However, this signature of selection would not always be detectable within populations because it depends on population characteristics such as population isolation, the extent of pollen and seed dispersal, and the allelic diversity at the S-locus. Empirical results presented here, as well as those reported in the literature, showed contrasted patterns of spatial genetic structure at the S-locus, and this could only partly be explained by the population characteristics.

Acknowledgments

This work was supported by an ATIP research grant from the life science department of the CNRS, and by a PhD grant to VL from CNRS and Région Nord-Pas de Calais. The authors wish to thank Pierre Saumitou-Laprade and Cécile Godé for the initial trials of microsatellite primers in *A. halleri*, Virginie Legros, Irene Canas and Adeline Courseaux for technical support for the genotyping, and Claire Feutrie, Eric Schmitt and Robert Dron for the plants care in greenhouse. They also thank Sylvain Billiard for useful comments on the manuscript.

References

Bos M, Van der Haring E (1988) Gene flow in Plantago. II gene flow pattern and population structure: a simulation study. *Heredity* **61**, 1-11.

Brennan AC, Harris SA, Hiscock SJ (2003) Population genetics of sporophytic self-incompatibility in Senecio squalidus L. (Asteraceae) II: a spatial autocorrelation approach to determining mating behaviour in the presence of low S allele diversity. *Heredity* **91**, 502-509.

Brennan AC, Harris SA, Hiscock SJ (2006) The population genetics of sporophytic self-incompatibility in Senecio squalidus L. (Asteraceae): the number, frequency, and dominance interactions of S alleles across its british range. *Evolution* **60**, 213–224.

Brooks RJ, Tobias AM, Lawrence MJ (1996) The population genetics of the self-incompatibility polymorphism in Papaver rhoeas. XI. The effects of limited pollen and seed dispersal, overlapping generations and variation in plant size on the variance of S-allele frequencies in populations at equilibrium. *Heredity* **76**, 367–376.

Castric V, Vekemans X (2004) Plant self incompatibility in natural populations : a critical assessment of recent theoritical and empirical advance. *Molecular Ecology* **13**, 2873–2889.

Charlesworth D, Charlesworth B (1987) Inbreeding Depression and its Evolutionary Consequences. *Annual Review of Ecology and Systematics* **18**, 237-268.

Clauss MJ, Koch MA (2006) Poorly known relatives of Arabidopsis thaliana. *Trends in Plant Science* **11**, 449-459.

Doligez A, Baril C, Joly HI (1998) Fine-scale spatial genetic structure with nonuniform distribution of individuals. *Genetics* **148**, 905-919.

Glémin S, Gaude T, Guillemin ML, *et al.* (2005) Balancing Selection in the Wild: Testing Population Genetics Theory of Self-Incompatibility in the Rare Species *Brassica insularis*. *Genetics* **171**, 279-289.

Hardy OJ, González-Martínez SC, Fréville H, *et al.* (2004) Fine-scale genetic structure and gene dispersal in Centaurea corymbosa (Asteraceae) I. Pattern of pollen dispersal. *Journal of Evolutionary Biology* **17**, 795.

Hardy OJ, Maggia L, Bandou E, *et al.* (2006) Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology* **15**, 559-571.

Hardy OJ, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**, 145–154.

Hardy OJ, Vekemans X (2002) A versatile computer program to analyse spatial genetic structure at the individual or population model. *Molecular Ecology Notes* **2**, 618-620.

Heuertz M, Vekemans X, Hausman J-F, Palada M, Hardy OJ (2003) Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash. *Molecular Ecology* **12**, 2483-2495.

Heywood JS (1991) Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics* **22**, 335-355.

Lawrence MJ (2000) Population Genetics of the Homomorphic Self-incompatibility Polymorphisms in Flowering Plants. *Annals of Botany* **85 (Supplement A)**, 221-226.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* **82**, 1420-1425.

Malécot G (1950) Quelques schémas probabilistes sur la variabilité des populations naturelles. *Ann. Univ. Lyon Sci. Sect. A* **13**, 37–60.

Muirhead CA (2001) Consequences of population structure on genes under balancing selection. *Evolution* **55**, 1532-1541.

Neuhauser C (1999) The ancestral graph and gene genealogy under frequency-dependent selection. *Theoretical Population Biology* **59**, 203-214.

Pauwels M, Saumitou-Laprade P, Holl A-C, Petit D, Bonnin I (2005) Multiple origin of metallicolous populations of the pseudometallophyte Arabidopsis halleri (Brassicaceae) in Central Europe: the cpDNA testimony. *Molecular Ecology* **14**, 4403-4414.

Schierup MH, Bechsgaard JS, Christiansen FB (2007) Selection at work in self-incompatible Arabidopsis lyrata. Spatial distribution of S-haplotypes in Iceland. *Genetics* **in press**.

Schierup MH, Vekemans X, Charlesworth D (2000) The effect of subdivision on variation at muti-allelic loci under balancing selection. *Genetical research Cambridge* **76**, 51-62.

Schierup MH, Vekemans X, Christiansen FB (1997) Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics* **147**, 835-846.

Schueler S, Tusch A, Scholz F (2006) Comparative analysis of the within-population genetic structure in wild cherry (Prunus avium L.) at the self-incompatibility locus and nuclear microsatellites. *Molecular Ecology* **15**, 3231-3243.

Slatkin M (1973) Gene flow and selection in a cline. *Genetics* **4**, 733-756.

Sokal RR, Jacquez GM, Wooten MC (1989) Spatial Auto-Correlation Analysis of Migration and Selection. *Genetics* **121**, 845-855.

Sokal RR, Oden NL, Thomson BA (1997) A simulation study of microevolutionary inferences by spatial autocorrelation analysis. *Biological journal of the linnean society* **60**, 73-93.

Sokal RR, Rohlf FJ (1995) *Biometry: the principles and practice of statistics in biological research* 3rd edn. W. H. Freeman & Co, New York.

Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proceeding of the National Academy of Science of the United States of America* **27**, 2419-2423.

Takayama S, Isogai A (2005) Self-incompatibility in plants. *Annual Review of Plant Biology* **59**, 467-489.

Valiere N (2002) GIMLET : a computer program for analysing genetic individual identification data. *Molecular Ecology Notes* **2**, 377-379.

Van Rossum F, Bonnin I, Fénart S, *et al.* (2004) Spatial genetic structure within a metallicolous population of *Arabidopsis halleri*, a clonal, self-incompatible and heavy-metal-tolerant species. *Molecular Ecology* **13**, 2959-2967.

Van Rossum F, Triest L (2006) Fine-scale genetic structure of the common Primula elatior (Primulaceae) at an early stage of population fragmentation. *American Journal of Botany* **93**, 1281-1288.

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

Vekemans X, Slatkin M (1994) Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* **137**, 1157-1165.

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations : cautions and guidelines. *Molecular Ecology* **10**, 249-256.

Wright S (1939) The distribution of self sterility alleles in populations. *Genetics* **24**, 538-552.

Chapitre 2 - Sélection fréquence-dépendante, dominance et dérive : Effets sur le locus d'auto-incompatibilité

Nous avons vu en introduction que la force de sélection principale conditionnant le polymorphisme au locus S était la sélection fréquence-dépendante. Nous avons voulu tester empiriquement l'influence de cette sélection fréquence-dépendante sur les allèles S. Nous avons également voulu quantifier l'effet d'autres facteurs évolutifs tels que la dérive génétique ou les relations de dominance entre allèles.

Grâce à une double approche théorique, d'une part par simulations de populations finies d'individus portant un système d'auto-incompatibilité sporophytique, et d'autre part par l'étude empirique des fréquences et des relations de dominance au locus en population naturelle, nous avons tenté de répondre aux questions suivantes :

Quel est le niveau de polymorphisme au locus S en population naturelle ?

- ✓ Quelle est la richesse allélique au locus S ?
- ✓ Quelle est la distribution des fréquences au locus S ?

Ce patron de polymorphisme est il expliqué :

- ✓ par la sélection fréquence-dépendante ?
- ✓ Par les relations de dominance entre allèles ?Par la dérive génétique ?Quelle est l'importance respective de ces trois facteurs évolutifs ?

Ce chapitre est traité au travers de l'article suivant : *Polymorphism at a locus under balancing selection: influence of genetic drift and dominance interactions. Example of the sporophytic self-incompatibility locus of Arabidopsis halleri.* **Llaurens V.**, Billiard S., Leducq J.B., Castric V., Klein, E., Vekemans X., en préparation pour le journal *Evolution.*

Polymorphism at a locus under balancing selection: influence of genetic drift and dominance interactions.

Example of the sporophytic self-incompatibility locus in *Arabidopsis halleri*

Llaurens V.*, Billiard S.*, Leducq J.B.*, Castric V.*, Klein E. K. +, Vekemans X.*

Corresponding author: Xavier Vekemans

Email : xavier.vekemans@univ-lille1.fr

* Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8016

Université des Sciences et Technologies de Lille 1

F-59655 Villeneuve d'Ascq

France

+ UR546, Unité de Biostatistique et Processus Spatiaux, INRA

Domaine St-Paul - Site Agroparc

F-84914 Avignon Cedex

France

Keywords: frequency-dependent selection, mating system, S-locus, allelic richness

Running title: Polymorphism at S-locus: frequency-dependent selection, dominance and genetic drift

Abstract

Balancing selection is an evolutionary process maintaining genetic diversity in natural populations, but determining its relative importance remains a major challenge in empirical population genetic studies. Here we estimated the extent to which polymorphism at a sporophytic self-incompatibility locus (S-locus) observed in a natural population could be explained by simple population genetic models including balancing selection, dominance relationships and genetic drift. We first used a molecular genotyping method to identify S-locus haplotypes in a natural isolated population of *Arabidospis halleri* and estimate their frequencies in a population sample in two different generations. We found eight different S-haplotypes and characterized their dominance relationships by controlled pollinations. We then used flexible stochastic simulations to theoretically predict the expected allelic frequencies in a finite population with the same dominance relationships as was observed, as well as the range of allele frequency variation between two successive generations. We finally compared observed and expected allelic frequencies in the population and in the offspring. Our results showed that the frequency of two S-haplotypes departed significantly from their theoretical expectation. Yet, likelihood comparisons across different genetic models of self-incompatibility showed that the data best fitted a model with (1) the observed dominance relationships and (2) no pollen limitation. Those results hence demonstrate that population genetics models including only basic features of balancing selection such as dominance relationships can reliably predict polymorphism at the S-locus. Balancing selection and genetic drift may thus be considered as the main evolutionary processes influencing polymorphism at this gene.

Introduction

Balancing selection is often invoked as a major evolutionary process maintaining polymorphism. The evolutionary mechanisms underlying balancing selection are very diverse, including overdominance (Lewontin, 1978), temporal and spatial heterogeneity of the environment (Levene, 1953), interspecific interactions such as host-parasite coevolution (Bergelson *et al.*, 2001) and negative frequency-dependent selection (FDS). The most obvious consequence of balancing selection is a deterministic increase in heterozygosity at the target locus. However, in finite populations, it is also expected to (1) counteract the loss of alleles through genetic drift and thus increase allelic richness at equilibrium (Robertson, 1962) and (2) modify the shape of the allele frequency spectrum from the classical U-shaped distribution for neutral alleles to a distribution with a peak at an intermediate gene frequency (Maruyama & Nei, 1981).

Theoretical developments have suggested different empirical approaches to detect loci under balancing selection in finite populations (Charlesworth *et al.*, 1997). However, only few studies have compared observed polymorphisms with theoretical predictions in known systems in order to determine the relative importance of balancing selection in comparison to other evolutionary processes such as genetic drift, migration, and other forms of selection. The results obtained so far are contrasted. Van Oosterhout *et al.* (2006) observed that levels of polymorphism at the major histocompatibility complex (MHC) in wild populations of *Poecilia reticulata* can be mainly explained by the balancing selection process, when taking into account local population size and strength of selection estimated by the local level of pathogen load. In contrast, Miller & Lambert (2004) showed that genetic drift outweighed balancing selection in determining MHC diversity in bottlenecked populations of the New Zealand robins (*Petroica traversi*).

Negative FDS is especially frequent among mating systems, such as sex chromosomes (Cho *et al.*, 2006), mating type determination in several fungi (Wu *et al.*, 1998) and plant self-incompatibility (Wright, 1939). Self-incompatibility (SI) is a widespread genetic system that prevents selfing in hermaphrodite plant species (de Nettancourt, 2001). It has been the subject of many theoretical and empirical investigations by population geneticists, as an example of a system subject to strong negative FDS (Castric & Vekemans, 2004). Self-incompatibility specificities are generally controlled by a single genetic factor, the S-locus, which exhibits a high level of polymorphism. In gametophytic SI, self-incompatibility phenotypes are determined gametophytically in pollen and sporophytically in pistils, with co-dominance among allelic specificities. Wright (1939) noted that alleles at the S-locus should be selectively equivalent, such that their transmission rate should only depend on their relative frequencies. In a review, Lawrence (2000) reported that for most species with gametophytic SI, the observed allele frequencies fitted the equal allele frequency distribution expected at equilibrium. In some plant species, however, strong departure from expectation were consistently reported because of migration (Stoeckel *et al.*), selection at a linked locus (Lane & Lawrence, 1995), or variance in reproductive success among individuals (Brooks *et al.*, 1996). Altogether, these studies on MHC and SI systems illustrate that in finite populations, a peak at intermediate gene frequency is indeed generally observed at loci subject to balancing selection, but other evolutionary processes may cause a substantial increase in the variance of allelic frequencies around this peak.

Additional complications arise in plants with sporophytic SI, where specificities at the S-locus are determined sporophytically both in pollen and pistil, because complex patterns of dominance relationships occur among alleles (Bateman, 1952). In particular, alleles may be ranked according to a linear hierarchy of dominance, or belong to a number of discrete classes of dominance, with co-dominance among alleles belonging to the same class. Dominance in

sporophytic SI introduces heterogeneity in effective selection coefficients among alleles, as highly recessive alleles will often be hidden from FDS whereas most dominant alleles will always be exposed to selection (Schierup *et al.*, 1997; Billiard *et al.*, 2007). In an infinite panmictic population, this will cause recessive alleles to reach higher frequencies at equilibrium than more dominant alleles. In a finite population, recessive alleles will be more prone to genetic drift, and are expected to exhibit higher variance in frequency than dominant alleles, so that the overall allele frequency spectrum may be much wider. When patterns of dominance relationships among alleles are organized into distinct classes of dominance, it was shown that the higher the dominance level, the higher the number of alleles expected to be maintained within the class, with a single allele in the most recessive class in general (Uyenoyama, 2000; Billiard *et al.*, 2007). In addition, Vekemans *et al.* (1998) suggested that FDS in SI systems may also act on female fitness, through an increase in seed production for individuals expressing a rare SI phenotype under ecological conditions of pollen limitation. This additional component of selection may induce substantial changes in the equilibrium allele frequency spectrum and this effect should vary according to the overall pattern of dominance relationships (Vekemans *et al.*, 1998).

In contrast to the considerable advances obtained in theoretical knowledge on sporophytic SI, detailed empirical studies devoted to disentangle the relative importance of FDS, genetic drift, and dominance relationships in determining allele frequencies in finite populations are strikingly scarce. This is in part due to the fact that molecular genotyping methods are needed for accurately estimating allele frequencies, whereas the S-locus in sporophytic SI has only been identified in a single family (Brassicaceae). Another reason is that, when the number of extant alleles in a population is large, determination of patterns of dominance relationships among alleles requires performing a considerable number of pollination tests among genotypic combinations. In a single population of *Arabidopsis lyrata* from Iceland, Mable *et*

al. (2003) identified at least 11 S-alleles in a sample of 20 individuals and noted that the allele with the highest frequency was also the most recessive, in agreement with theory. Prigoda *et al.* (2005) analysed more extensively patterns of dominance relationships among S-alleles in *A. lyrata*. They observed four hierarchical classes of dominance levels, with co-dominance among alleles within classes, and showed that these classes corresponded to phylogenetic clusters of alleles in an allelic genealogy. Moreover, they found identical patterns of dominance relationships in pollen and pistil. Using these patterns of dominance and newly identified S-alleles in the same Icelandic population, Schierup *et al.* (2006) estimated the allele frequency distribution and found that it fitted closely deterministic expectations. However, they did not take into account drift in their theoretical analysis, and more importantly, (Billiard, *in press*) showed that it is highly improbable that the observed number of alleles be maintained in such a small population (11 alleles for about 87 individuals), suggesting that this population is certainly not isolated and at equilibrium. In *Brassica insularis*, S-alleles belong to two classes of dominance, and (Glémin *et al.*, 2005) showed that alleles from the dominant classes are more numerous and have lower frequency than alleles from the recessive class. However, some uncertainties in the data precluded more formal tests against theoretical models.

In this paper we investigated diversity at the S-locus in a two generations sample from a small isolated natural population of *A. halleri*, a sister species of *A. lyrata* with sporophytic SI. We characterized the patterns of dominance interactions among the observed S-alleles thanks to controlled pollinations, and we used numerical simulations with the aim of evaluating the relative effects of negative FDS, genetic drift, and patterns of dominance on the determination of allele frequencies at the S-locus across two generations. Specifically we compared the observed patterns of dominance relationships among alleles in pollen and pistil against those observed in *A. lyrata* (Prigoda *et al.*, 2005). We tested whether the number of alleles in a

dominance class increased with its dominance level, as expected from theory. We also compared the observed allele frequencies at the S-locus with those predicted under several alternative models of frequency-dependent selection: deterministic models *vs* finite population models; models taking into account dominance interactions *vs* models assuming strict co-dominance among alleles; models with FDS acting only in pollen *vs* models with FDS in pollen and pistil. Finally, we compared for each individual allele the observed frequency change between parental and offspring generations with that expected under pure drift *vs* that predicted under negative FDS, specifically taking the observed patterns of dominance into account.

Materials and methods

Studied species

Arabidopsis halleri (Brassicaceae) is a diploid perennial herb with sporophytic SI that is closely related to the two model species *Arabidopsis thaliana*, and *A. lyrata* (Mable *et al.*, 2003). It is a European species growing in mountainous areas and on heavy-metal polluted sites, for example close to industrial areas in France.

Population sampling and DNA extraction

We focused on a single isolated population located in Nivelle (France, 50° 28' North, 3° 27' East). Sampling were collected as described in Llaurens *et al.* (submitted). In one part of the population we sampled exhaustively all individuals (N=328) by collecting leaves from each ramet that visually appeared as a distinct individual. We also sampled offspring on a small subset of these plants (n=256 from 110 fruits from a total of 22 maternal plants) and grew these offspring for eight weeks in a greenhouse. In the other part of the population, we sampled leaves from 36 regularly spaced individuals. All leaf samples (adult + offspring

populations) were oven-dried at 55°C, and DNA was extracted using the Nucleospin® Multi-96 Plant extraction kit from Macherey-Nagel®. Overall, we collected DNA from 364 adult individuals and 256 offspring. Total census size of the population is at least 800 individuals. In a separate study we showed that individuals from this population are largely outcrossing, suggesting the SI systems is fully functional (Llaurens *et al.*, submitted).

Screening for S-haplotypes

We used a molecular genotyping method to screen for S-haplotypes. In Brassicaceae, the S-locus contains two major genes (Schopfer *et al.*, 1999): *SCR*, encoding for a protein of the pollen coat, and *SRK*, encoding for a transmembrane kinase receptor in the papilla cells of pistils. A specific receptor-ligand interaction occurs between these two proteins, triggering a signaling cascade leading to pollen tube growth inhibition when pollen and pistil S-specificities are matching. *SCR* and *SRK* are typically tightly linked, so identification of the *SRK* allele is sufficient to characterize the whole S-haplotype. Accordingly, we will use the term S-haplotypes to name functional alleles at the S-locus that correspond to different specificities. In a previous study, twenty-two haplotypes of the *A. halleri SRK* gene (AhSRK1 to AhSRK22) have been identified in a European-wide population sample (Castric & Vekemans, 2007). Three additional S-haplotypes have subsequently been identified and sequenced (Llaurens & Ruggiero, unpublished results). We used PCR primers specifically targeted towards each S-haplotype (Table 1) to genotype each individual at the S-locus, thus generating data corresponding to a pattern of presence/absence for each S-haplotype. The reaction mixture (15µL) contained 20 ng DNA, 1X buffer (Applied biosystem®), 2 or 2.5 mM $MgCl_2$ (depending on the primer pair), 200 µM Fermentas® dNTP mix, 200 µg/mL BSA, 0.2 µM of each microsatellite primer, and 0.025 U/µL *Taq* polymerase (Amplitaq DNA polymerase, Applied biosystem®). The amplification profile was 5 min at 95°C, 35 cycles of

Table 1 Primer sequences and PCR conditions to amplify S-haplotype in *A. halleri*

Haplotype	5' Primer sequence 3'	Annealing temperature	[MgCl2] in PCR mix
AhSRK01	F TCAAGATTGAAGCTGAGTGA R TACACAACCCGTCCCGCCAA	57°C	2,5mM
AhSRK02	F GGGTTATTACAGCGATTTGA R GCCCACTCTTGCGGGTACGG	53°C	2,5mM
AhSRK03	F TGATATTTACATGGGGTGCA R CCTCTTGTCCACAATAGCAC	51°C	2,5mM
AhSRK04	F GACAATAAATACCGTAGACG R ACGCCGTACAGTTACAATGT	52°C	2,5mM
AhSRK05	F GATTTCAAAAGAACAGGATC R AATCCACATCAAACAACCAA	53°C	2,5mM
AhSRK06	F AAGGGTTTCAACCTCTGTAT R TACAATTGCAATTCTCCCTA	51°C	2,5mM
AhSRK07	F AGGATTGGGATTTTGGATTT R TCCTTACATTGTTCCAAAGA	53°C	2,5mM
AhSRK09	F GAGTCACCACTCCCAATGTG R TCAGTGTTAGTGTTATCTAG	54°C	2,5mM
AhSRK10	F GAGAATAGGGAGGAGGTCTG R CACCGCAATTCAATTGCCTA	56°C	2,5mM
AhSRK11	F CAACGGAGATGGATTTCACA R GTCCTACCATCCCGGGTAAT	53°C	2,5mM
AhSRK12	F ATCATGGCAGTGGAACACAG	58°C	2mM

	R CAAATCAGACAACCCGACCC		
	F GAGAAGATTGGAACCAACAT		
AhSRK13		51°C	2mM
	R CGTCTTCCTCACACACCCTT		
	F TTATACTTTCCGACTAATTA		
AhSRK14		50°C	2,5mM
	R CAGCCTTAGTAGCAGGAGAT		
	F ATTCCAACAGAACGTGAAAAT		
AhSRK15		55°C	2,5mM
	R ATCTGACATAAAGATCTTGACC		
	F GATTGGTATTCTGGAGATTG		
AhSRK16		53°C	2,5mM
	R CCTTCACATCAATTATTCTA		
	F GAGAATAGAGAGGAGGTCAT		
AhSRK17		57°C	2,5mM
	R GTACAAACCCTTTGATGCAT		
	F CGTGTGATATGTATGAAGAA		
AhSRK18		50°C	2,5mM
	R CTTTTTCCTTGCAGCTCAAG		
	F ATGTTCTGGCCTTCATCGGC		
AhSRK19		51°C	2mM
	R ATCCGGCACCTTCATTCTAC		
	F ACGTTCCTAATGACCAATTC		
AhSRK20		56°C	2,5mM
	R CTTTCACACCAATTCTCCGG		
	F CATCTACTCGAGATTGACCC		
AhSRK21		55°C	2,5mM
	R ATAACTCTTTGGTCCAGAAA		
	F ATTACATGGTTTACAACTTT		
AhSRK22		50°C	2,5mM
	R CCACTCCTGCTGGTTCTTGG		
	F CTGGGTTTCATCGGTACGCG		
AhSRK23		56°C	2,5mM
	R ACATAACTCCGAATATCGGT		
	F CAATGAGTTCCACAGGTGCT		
AhSRK24		56°C	2,5mM
	R ACCTCTTTTTGCATTCTCCA		

AhSRK26	F AACGGACCAGTGTGATATGA R GTCTTTCTCATCAATTTCCA		56°C	2,5mM

40s at 95°C, 40s at Tm°C (specific for each pair of primer, Tab.I), 40s at 72°C, and one cycle of 10 min at 72°C and performed in MJ research PTC 200 thermocycler®. PCR products were mixed with loading dye and run at 110 V on 2% agarose gels in TBE buffer for 45 minutes. Fragments were fluorescently visualized by ethidium bromide under UV light and compared to a 100bp DNA ladder. A positive control for each targeted S-haplotype was run on each gel. Because the presence/absence genotyping procedure is sensitive to PCR failure, we took advantage of the results from a previous study on the exact same individuals, where we amplified 12 microsatellite loci, to check DNA quality. When we observed no amplification for 3 or more microsatellite loci, we considered that DNA was not properly extracted and eliminated the individual from the data set. Overall, we screened the 364 parents and 256 offspring for presence/absence of each of the 25 S-haplotypes.

Dominance interactions between S-haplotypes

To assess patterns of dominance relationships among S-haplotypes, we performed controlled pollinations between individuals with known genotype. Compatibility was determined phenotypically based on fruits and seeds development. Thirty five plants from seeds collected in Nivelle were used for these crosses. Controlled pollinations were performed under binocular magnifying glass. An anther of the chosen male partner was collected with tweezers and pollen was deposited on the pistil of the chosen female partner. Pistils were marked individually and measured seven days after pollination, which was sufficient for total fruit elongation (data not shown). Pollination was classified as successful when the fruit was at least 6.1 mm long and seeds developed, according to a preliminary investigation. We performed at least 10 replicates per cross and controlled pollinations were done reciprocally. We considered a cross as "compatible" when successful pollination occurred in at least 80% of the replicates.

These compatibility tests then allowed us to infer the phenotype of the crossed individuals: if two heterozygote individuals sharing only one given S-haplotype were incompatible, we deduced that the haplotype they shared was expressed in both of them. By testing each S-haplotype this way, we were able to characterize the dominance interaction between S-haplotypes both in pollen and pistil. In some cases, we used plants with a homozygote genotype at s-locus obtained by controlled crosses among plants with one S-haplotype in common.

Estimation of S-haplotype frequencies

The genotyping method by presence/absence does not unambiguously assess zygosity: individuals in which only one S-haplotype was found can be either homozygotes or heterozygotes with another unidentified S-haplotype. To address this issue, we took advantage of the asymmetry in selective effects among S-haplotypes in sporophytic SI. More specifically, because recessive S-haplotypes are frequently not expressed when present in heterozygote genotypes, they are expected to occur frequently as homozygote in the population (Billiard *et al.*, 2007). We therefore considered that individuals for which a single haplotype had been identified were homozygote when the haplotype belonged to the recessive or intermediary dominance class and heterozygote when the haplotype belonged to the most dominant class.

Comparing observed versus expected distributions of S-haplotype frequencies

We obtained deterministic distributions of S-haplotype frequencies in an infinite panmictic population using the general model of sporophytic SI of Billiard et *al.* (2007). We used numerical simulations to generate distributions in finite panmictic populations. We then compared the observed distribution of allelic and genotypic frequencies in the population with

their theoretical prediction under the specific SSI model existing in the study population using the software NESSI (Billiard, in press). This program allows the estimation of allelic and genotypic frequencies distribution in an isolated population, for any dominance relationships pattern, under frequency-dependent selection, mutation and drift. Predictions were computed for two kinds of selection regime, the so-called fecundity selection model where selection occurs through both pollen and stigma (Vekemans *et al* 1998), and the Wright's model, where selection only occurs through pollen (Wright 1939). The fecundity selection model has been specifically derived to account for pollen limitation, *i.e.* when seed set is restricted by the amount of compatible pollen in the population. In cases where dominance relationships could not be fully resolved with the pollination tests performed, we ran predictions for all possible cases.

We computed (1) deterministic equilibrium frequencies, (2) allelic richness, allelic frequencies and genotypic frequencies distributions at drift-mutation-selection "equilibrium" in finite populations and (3) distributions of the allelic and genotypic frequency changes in one generation. Since simulations in finite population were chosen to be the closest to the sampled population we simulated a population of size $N = 800$, with a sample size of 300 individuals. Simulations were first run for 5,000 "forget generations" (discarded for burn-in), after which approximate mutation-drift equilibrium had been reached (data not shown). Samples were then drawn at 1,000 generations intervals during 100,000 generations, and the whole process was replicated 100 times. We used the k-alleles model of mutations with $k = 8$ with mutation rate $\mu = 10^{-6}$.

Likelihood computations

Simulations were done for various selection regime (fecundity *vs.* Wright's model), different dominance relationships (all codominant *vs.* observed dominance relationships) and for

different dominance levels for specific alleles. In order to quantitatively compare the different models and hypotheses, we estimated the likelihood L of the observed allelic frequencies. For each simulation x, the likelihood for random sampling was computed given the allelic frequencies in the population: $L_x = \prod_{i=1}^{a} \binom{m_{x,i}}{k_i} \Big/ \binom{2N}{2n}$ with N the total number of individuals in the simulated population, n the number of individuals in the sample, $m_{x,i}$ the number of copies of allele i among the $2N$ chromosomes in simulation x, k_i the number of copies of allele i in the observed sample and a the total number of alleles. This expression comes from the multivariate hypergeometric distribution adequate for random draw without replacement. Then the full likelihood including radom sampling and genetic drifs, defined as the expectation $\hat{L} = E[L_x]$, was computed as the average L_x of over 10,000 simulations, conditionally on the existence of L_x, in other words when $m_{x,i} \geq k_i$. It is important to notice that $\hat{L}$ includes two stochastic factors, drift and sampling, which is useful because drift interacts both with the dominance level of alleles and the selection regime.

Prediction of the changes in S-haplotype frequencies across two generations

We used the simulation procedure to compute the frequency change expected between two successive generations. We compared directly the expected and observed S-haplotype frequency in the set of sampled offspring, by entering in the simulations the genotypic composition of the actual parental population. For genotypic frequencies in the pool of ovules produced by the parental population, we restricted the analysis to mothers from which offspring had been collected and weighted the genotypic frequencies by the number of offspring sampled for each mother. For genotypic frequencies in pollen, we considered the whole parental population. We performed 1,000 simulations and then computed for each S-

Table 2 Occurrence of the S-haplotypes in the parental and offspring generations in the population of Nivelle.

Number of individuals with 0, 1, 2 or 3 S-haplotypes found, and carrying S-haplotypes AhSRK01, AhSRK02, AhSRK04, AhSRK12, AhSRK15, AhSRK20, AhSRK22 and AhSRK24 in the parental and offspring populations

	Number of individuals with											
	0 haplotype	**1 haplotype**	**2 haplotypes**	**3 haplotypes**	**S01**	**S02**	**S04**	**S12**	**S15**	**S20**	**S22**	**S24**
Parental population (n=322)	27	129	166	5	124	38	55	97	36	95	23	8
Offspring population (n=245)	6	71	168	0	94	27	75	72	35	75	15	13

allele the range of frequencies corresponding to 95% of the simulations and the percentage of the simulations where all alleles of the parental population were kept in the next generation.

We further used allelic frequencies changes between parents and offspring to compare the likelihood of different self-incompatibility scenarios. Using the genotypes of mothers sampled in Nivelle and an estimate of allelic frequencies in the pollen pool that have fertilized them (based on genotypic frequencies among all individuals obtained above), we computed the expected allelic frequencies in offspring according to different hypothesis: 1) no self-incompatibility system, 2) SSI with co-dominance among all S-alleles; 3) SSI with dominance relationships observed in Nivelle and Wright's selection regime; 4/ SSI with dominance relationships observed in Nivelle and fecundity selection regime. We computed the likelihood of the allelic frequencies data in offspring under the four hypotheses as $L_y = \prod_{i=1}^{a} p_{y,i}^{k_i}$ with $p_{y,i}$ the expected frequency of allele i in offspring under hypothesis y and k_i the observed number of alleles i observed in offspring, and a the total number of alleles.

Results

S-haplotype identification

We identified in the Nivelle population eight out of the twenty-five currently known S-haplotypes in *A. halleri* (AhSRK01, AhSRK02, AhSRK04, AhSRK12, AhSRK15, AhSRK20, AhSRK22, AhSRK24, Table 2). With our genotyping method, we obtained unambiguous characterization of heterozygote genotypes for 334 individuals out of 567 (Table 2). We identified only one S-haplotype for 200 individuals, but we assumed for 143 of them that they were homozygotes, as they carried the most recessive S-haplotypes (see below). Overall, 33/567 (5.8%) of all individuals produced no amplification with any of the primers. Since we discarded individuals who also showed amplification problems with microsatellites, this is unlikely to be due to poor DNA quality. Hence, we hypothesized that these individuals carry

Table 3 Compatibility of controlled pollination measured by proportion of compatible crosses on at least 10 replicates.

Cells in grey figure incompatible crosses

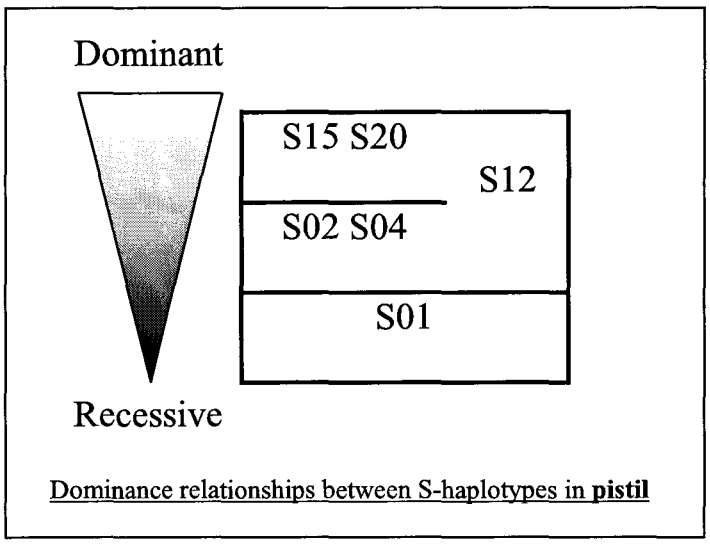
			STIGMA																	
		Stigma genotype	S01S01	S01S02	S01S04	S01S12	S01S15	S01S20	S01S22	S02S02	S02S04	S02S12	S02S15	S04S12	S04S20	S12S15	S12S20	S20S24	Cross with the same genotype	
	Pollen genotype	Inferred phenotype	S1	S2	S4	S12	S15	S20	S22	S2	S2-S4	S2-S12	S15	S4-S12	S20	S12-S15	S12-S20	S20 or S20-S24	Selfing	Outcrossing
POLLEN	S01S01	S1		1.00	1.00	1.00	1.00	1.00	1.00										0.07	0.05
	S01S02	S2	0.83			1.00				0.00	0.10	0.20	1.00						0.00	0.00
	S01S04	S4	1.00						1.00		0.20			0.00	0.32			1.00	0.00	0.00
	S01S12	S12	1.00	1.00				0.67	1.00			0.00		0.00		0.00	0.20	1.00	0.00	0.00
	S01S15	S15	1.00										0.00			0.00			0.00	0.17
	S01S20	S20	0.90						1.00						0.06		0.20	0.20	0.00	0.00
	S01S22	S22	1.00																0.00	
	S02S02	S2		0.00					1.00			0.00						1.00	0.00	
	S02S04	S4		1.00	0.10														0.00	0.00
	S02S12	S12		1.00		0.00				1.00				0.00					0.00	
	S02S15	S15		1.00			0.00									0.00			0.17	0.00
	S04S12	S12			0,80	0,12						0,25							0,00	
	S04S20	S20			0,97			0,00											0,00	0,21

S12S15	**S15**				1,00	0,00		0,80				0,00							0,00	0,13
S12S20	**S20**				1,00		0,10												0,20	
S20S24	**S20 or S20-S24**						0,00	1,00											0,10	
																		All	**0,03**	**0,07**

S-haplotypes that are not currently known. Five individuals were found to show 3 different S-haplotypes. No allele was consistently associated with these five genotypes, so we hypothesized that this observation was probably caused by mixing of plant material at the time of sampling or to DNA contamination problems. We therefore discarded these individuals from further analyses.

Dominance relationships between S-haplotypes

Table 3 shows the proportion of compatible crosses obtained by controlled pollination for individuals with known genotypes at the S-locus. For most genotypes, self-pollination and pollinations with plants carrying the same genotype were performed. These crosses were all classified as incompatible, confirming that the determination of the self-incompatibility genotype was correct and that the self-incompatibility system was functional. Crosses between different genotypes allowed the characterization of dominance relationships among S-haplotypes: incompatibility implying that at least one haplotype was expressed in common in both mates. We first pointed out that the crosses between homozygotes [AhSRK01; AhSRK01] with heterozygous individuals carrying only one copy of AhSRK01 were always compatibles. We thus inferred that the S-haplotype AhSRK01 was never expressed at heterozygous state, meaning that it must be recessive with respect to every other S-haplotype tested. We then used crosses of different combinations of heterozygotes with heterozygotes carrying the S-haplotype AhSRK01 to test the dominance relationship among the other S-haplotype. By the incompatibility pattern, we thus inferred dominance relationships among S-haplotypes in both pollen and pistil. Dominance relationships among AhSRK01, AhSRK02, AhSRK04, AhSRK12 and AhSRK15 were fully determined in pollen and pistil. The levels of dominance of S-haplotype in pollen and pistil were correlated overall, but dominance in pollen was strictly hierarchical whereas dominance in pistil was subdivided in three main



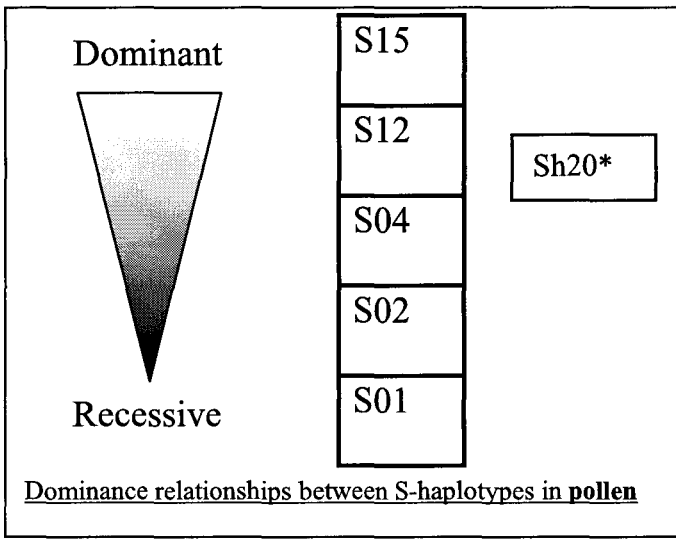


Figure 1: Dominance relationships determined between S-haplotypes by controlled pollinations. Haplotypes in the same cell of the table were co-dominant; *: S-Haplotype Sh20 is dominant over Sh12, Sh04, Sh02, Sh01 but relationship with Sh15 remains unknown.

classes with co-dominance within class and full dominance between classes (Figure 1). In pistils, a single haplotype (AhSRK01) belonged to the most recessive class, two haplotypes (AhSRK02 and AhSRK04) belonged to the second most recessive class and two haplotypes (AhSRK15 and AhSRK20) belonged to the most dominant class. Haplotype AhSRK12 had a particular behavior in pistil, since it was co-dominant with haplotypes AhSRK02, AhSRK04, AhSRK15 and AhSRK20, which belonged to different dominance classes. Dominance of the S-haplotype AhSRK20 in pollen is only partially known because we had not any available plant carrying the necessary genotypic combination allowing determination with all S-haplotype. We were able to check that AhSRK20 was dominant over AhSRK01, AhSRK02, AhSRK04, AhSRK12 but the relationship with AhSRK15 remained undefined. In pistils, AhSRK20 was co-dominant with AhSRK12 and AhSRK15. In both pollen and pistil, AhSRK24 was recessive or co-dominant with respect to AhSRK20 and AhSRK22 was dominant over AhSRK01, but information on the dominance of AhSRK22 and AhSRK24 remained scarce for the same reason.

Allelic frequencies distribution

For the computation of the S-haplotypes frequencies, we assumed that when the recessive S-haplotypes AhSRK01, AhSRK02 and AhSRK04 were found alone in an individual, this individual was a homozygote.

For the prediction of S-haplotypes frequencies by the NESSI model, we introduced in the dominance interactions among S-haplotypes observed in the natural population. Because dominance interactions were not fully determined for some haplotypes, we made three main hypotheses: (1) in pollen, AhSRK20 is dominant over AhSRK12 and recessive to AhSRK15, AhSRK24 is dominant over AhSRK15, AhSRK22 is dominant over all the other haplotypes. In pistil, AhSRK24 and AhSRK22 are co-dominant with AhSRK20, AhSRK12 and

Table 4: Likelihood of the observed frequencies in Nivelle with different selection model and dominance pattern with population size N=800, and mutation rate $\mu=10^{-6}$

Selection model	Dominance pattern	Log (likelihood)
Wright	hyp. 1	-175.046
Wright	hyp. 2	-176.853
Wright	hyp. 3	-224.732
Wright	all codominants	-251.459
Fecundity selection	hyp. 1	-189.067
Fecundity selection	hyp. 2	-197.138
Fecundity selection	hyp. 3	-228.572
Fecundity selection	all codominants	-251.541

AhSRK15; (2) same as hypothesis (1) except that in pollen AhSRK20 is dominant over AhSRK15 and recessive to AhSRK24; (3) same as hypothesis (2) except that in pollen, AhSRK24 is dominant over AhSRK01 and AhSRK22 situated below AhSRK02. In pistils, AhSRK24 and AhSRK22 are co-dominant with AhSRK02, AhSRK04 and AhSRK12. Hypotheses (1) and (2) allowed us to test the consequences of the AhSRK20 position in pollen and hypothesis (3) was designed to test if AhSRK22 and AhSRK24 were rather recessive or dominant with respect to the other S-haplotypes.

As shown on Table 4, the model with Wright's model of selection under dominance pattern from hypothesis (1) was the most likely to explain our empirical data. The likelihood of the model with full codominance was very low, suggesting that the dominance pattern had a very important influence on S-haplotype frequencies. The comparison of dominance hypotheses showed that the alternate hypothesis (2) on AhSRK20 was slightly less likely than hypothesis (1) suggesting that AhSRK20 was more likely recessive than dominant with respect to AhSRK15. The model with dominance pattern from hypothesis (3) had an even lower likelihood, suggesting that AhSRK22 and AhSRK24 should be rather dominant than recessive like assumed in hypothesis (1) and (2). Altogether, the dominance relationship described in hypothesis (1) allowed the best predictions for our empirical data. As shown in Table 4, the likelihood obtained with Wright's model of selection was always higher than with the fecundity selection model. The prediction of frequency-dependent selection given by Wright's model thus seemed to better fit the observed data. We checked the sensibility of the likelihood computation to changes in mutation rate μ and population size N and pointed out that the rank order of likelihoods of the different models was robust to uncertainty in these parameters (results not shown).

Figure 2 shows the haplotype frequency in the parental population and in theoretical predictions under hypothesis (1). The confidence interval was computed only on simulations

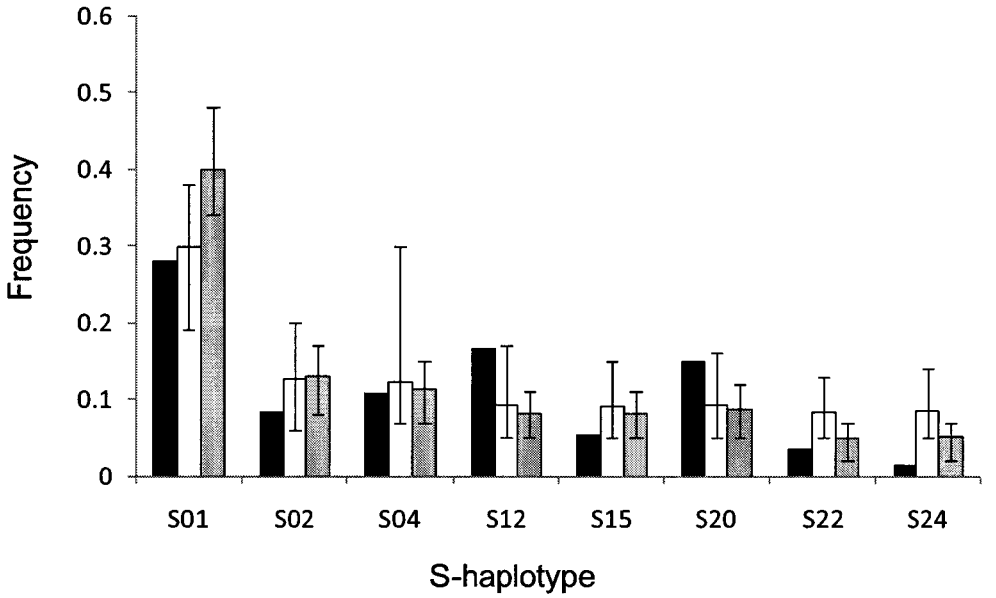


Figure 2: S-haplotypes frequencies observed in Nivelle and predicted under dominance interaction hypothesis 1.

Computations performed only on simulations where the eight S-haplotypes were conserved. Black bars: frequencies observed in natural population; White bars: frequencies under Wright model; Grey bars: frequencies under fecundity selection model; Standard error bars represents the 95% confidence interval obtained with the simulations (n=10,000) in finite population (N=800).

Table 5: Probability of losss of one S-haplotype in 10,000 simulations with population size N=800 and mutation rate $\mu = 10^{-6}$ based on under dominance relationships from hypothesis 1

Selection model	S1	S2	S4	S12	S15	S20	S22	S24
Fecundity selection	0	0	0	0	0	0	0.15	0.12
Wright	0.10	0.18	$1.\ 10^{-3}$	$9.\ 10^{-3}$	7.10^{-4}	0	0	0

with no loss of S-haplotype. For the observed frequencies, according to the expected recessive effect, AhSRK01, the most recessive haplotype had the highest frequency, followed by the two intermediate haplotypes AhSRK02 and AhSRK04, and by AhSRK15, which is the most dominant haplotype. However, AhSRK12 and AhSRK20 had higher frequency than AhSRK02 and AhSRK04, which are more recessive. Although this observation apparently conflicted the recessive effect, their frequency nevertheless belonged to the 95% confidence interval obtained by simulation in finite population in Wright model under hypothesis (1). Under Wright model, two S-haplotypes were outside the predicted range: AhSRK22 and AhSRK24 which had a slighty lower frequency than expected.

Allelic richness

Table 5 shows the proportion of simulations where each S-haplotype was lost. In Wright's model, only recessive S-haplotypes (AhSRK01, AhSRK02 or AhSRK04) were likely to be lost, and loss only ocurred in a small proportion of the simulations (≈10%). On the contrary, dominant S-haplotypes AhSRK22 and AhSRK24 were more susceptible to drift in the fecundity selection model, with haplotype losses observed in about one third of all simulations. In both models of selection the same allelic richness was essentially retained: all eight S-haplotypes were maintained in 79% and 76% of simulations in Wright's and fecundity selection models. Wright's model was thus slightly more likely to explain the level of diversity observed in this population.

The predicted number of alleles for a neutral locus should tend to n_a=1 based on Kimura's equation Kimura (1983) (assuming effective population size N=800 and mutation rate $\mu=10^{-6}$) whereas for a locus under heterozygous advantage, the equation described in Takahata (1990) predicted n_a=8 alleles (assuming effective population size N=800, selection coefficient s=0.33 and mutation rate $\mu=10^{-6}$) so that balancing selection could explain a higher polymorphism

Table 6: Observed S-haplotypes in the offspring and expected range of frequencies under Wright model and dominance relationships from hypothesis (1).

95% boundaries were obtained by 10,000 simulations in finite population (N=800, μ=10^{-6}), * showed that the observed frequency had a significant departure from the 95% boundaries

S-haplotype	Observed frequency	95% boundaries in a finite population
S1	0.26	0.25 - 0.32
S2	0.06	0.05 - 0.09
S4	0.18*	0.08 - 0.13
S12	0.16	0.16 - 0.22
S15	0.07	0.06 - 0.1
S20	0.15	0.14 - 0.2
S22	0.03*	0.04 - 0.08
S24	0.03	0 - 0.03

than the neutral case. In particular case of balancing selection acting on the gametophytic self-incompatibility locus, the predicted number of alleles based on Yokohama & Hetherington (1982) was n_a=16 (assuming effective population size N=800 and mutation rate μ=10^{-6}).

Frequency change in one generation

Table 6 shows the frequency change in one generation, taking into account maternal genotypes and FDS. The observed S-haplotype frequency changes were consistent with the 95% boundaries obtained by 10,000 simulations in a finite population with dominance relationships corresponding to hypothesis (1) except for AhSRK04 and AhSRK22. This suggested that S-haplotype frequencies observed in the offspring generation were mostly consistent with the frequency change predicted by the model.

The log-likelihood under the four models tested were respectively: -31.41 with no self-incompatibility system (1), -38.6 with SSI with all S-alleles codominant (2), -28.52 with SSI with the observed dominance relationships and under Wright selection model (3) and -29.13 with SSI with the observed dominance relationships and under fecundity selection model (4). The best fit was thus obtained for the case (3), suggesting further that the mating system and the dominance relationship of S-alleles have a significant effect on genetic diversity. Remember indeed that the difference in AIC between two models (i. e. twice the difference in log-likelihood for our model without parameters fitted) is considered "significant" and "very significant" at values of 5 and 10 respectively (Anderson & Burnham, 2002). Furthermore, it seems that Wright's selection regime provides a better fit to the data than the fecundity selection model.

Discussion

Dominance relationships observed

Controlled pollinations confirmed the functionality of the self-incompatibility system in *A. halleri*, which was described in the literature (Clauss & Koch, 2006) and suggested by the high outcrossing rate estimated in the Nivelle population by paternity analysis on neutral markers (Llaurens *et al.*, submitted).

Dominance relationships between S-haplotypes in *Arabidopsis halleri* seemed to be different in pollen and in pistils. More specifically, although co-dominance among S-haplotypes seemed to be more common in pistils, full hierarchical dominance among S-haplotypes was mainly observed in pollen. Patterns of dominance interactions among S-haplotypes have been compared between pollen and pistils in a few sporophytic self-incompatible species, and several common features are emerging. In most cases studied, the levels of dominance in pollen and pistil were generally similar, like in *Ipomoea trifida* (Kowyama *et al.*, 1994) or *Arabidopsis lyrata* (Schierup *et al.*, 2006). However, co-dominance was found more common in the pistil than in the pollen: in *Brassica campestris*, Hatakeyama *et al.* (1998) found non-linear dominance relationships among 24 S-haplotypes with two dominance groups in the pistil and three in the pollen. Five S-haplotypes found in the population of Nivelle were trans-specific with S-haplotypes found in *A. lyrata* (Castric *et al.*, in prep.): the *A. halleri* S-haplotype AhSRK01, AhSRK02, AhSRK04, AhSRK12, AhSRK24 should be orthologous to the *A. lyrata* AlSRK01, AlSRK17, AlSRK37, AlSRK42, AlSRK20 respectively. The *A. lyrata* S-haplotypes were classified in four phylogenetic group A1, B, A3 and A2 by Prigoda *et al.* (2005) which supposed hierarchical dominance relationships among groups and codominance within groups. The S-haplotype AhSRK01/ AlSRK01 belonged to phylogenetic group A1, known as the most recessive, the S-haplotypes AhSRK02/ AlSRK17 and AhSRK04/ AlSRK37 belonged to group A3 which is supposed to have an intermediary dominance level and AhSRK12/ AlSRK42 and AhSRK24/ AlSRK20 belonged to phylogenetic group A2 which is supposed to be the most dominant group. The dominance

found here in *A. halleri* is thus consistent with the one found for the trans-specific S-haplotypes of *A. lyrata.* However, we pointed out that the presumed codominance within phylogenetic group was mostly true in the pistil but not in the pollen since we have observed a fully linear hierarchy among S-haplotypes.

Allelic richness

The determination of S-haplotypes by sequences comparisons was used to design the primers allowing genotyping at the S-locus for the individuals of this population. This method is currently used in the study of self-incompatibility haplotypes but should be combined to segregation analysis of combination of S-haplotype to make sure that all S-haplotype found by sequence analysis belong to the same locus, *i.e.* the S-locus. Such direct evidence was not available, but some indirect evidence could be highlighted. Some S-haplotypes found in *A. halleri* were trans-specific with S-haplotype of *A. lyrata* as mentioned previously. Moreover, the observation of self-incompatibility for controlled crosses between plants carrying the same S-haplotypes also suggested that S-haplotypes truly belonged to S-locus. We may thus be reasonably confident that the S-haplotypes found during this study referred to self-incompatibility haplotypes.

We found eight S-haplotypes in this population out of 25 identified in the European populations of *A. halleri.* As we observed individual with only one S-haplotype, it could mean that they were homozygotes or heterozygotes with another unknown S-haplotype. We may thus suppose that the number of S-haplotype is actually slightly higher in this population. But as the frequency of fully unknown genotypes was rather low, the total number of unknown S-haplotypes should not be so high.

This number belonged to the range of the S-haplotypes numbers observed in natural populations of other species with sporophytic self-incompatibility system: Glémin *et al.*

(2005) observed three to sixteen putative S-alleles in different populations of *Brassica insularis*, and Schierup *et al.* (2006) found eleven distinct SRK alleles in a natural population of *A. lyrata*. In this last study, they found no differences between observed and deterministic frequencies but by simulation in finite population; Billiard *et al.* (in press) pointed out that the allelic richness observed in this population is too high to be maintained by balancing selection. It was then hypothesized that the high level of allelic richness should be maintained by immigration. In our study, the level of allelic richness observed at self-incompatibility locus in natural population was convenient with the prediction in finite population.

Results from a previous paternity analysis (Llaurens *et al.*, submitted) showed that pollen immigration was about 10 % corresponding to pollen coming from unsampled plants of the same population allowing us to consider the population of Nivelle as rather small and isolated. The number of S-haplotypes that a given finite population could maintain is not predictable because it depends tightly on the level of dominance of the arisen alleles that is fully unknown. However, our results show that the observed allelic richness is consistent with the hypothesis that the Nivelle population is isolated, of approximate size $N = 800$, and that the present S-haplotypes have the assessed dominance relationships. Allelic richness level observed could be explained by Wright's frequency-dependent selection model and genetic drift.

S-haplotype frequencies distribution

The observed frequencies indicated that S-haplotypes followed the recessive effect. We showed that our observed S-alleles frequencies were closer to the prediction based on Wright's frequency-dependent selection than to those based on the fecundity selection model. The fecundity selection model assumes that frequency-dependent selection is acting both on pistils and pollen, which could take place in limiting pollen situation, where it may be difficult

for pistils to be fertilized. A previous study on pollen dispersal in this population has suggested that this population was well supplied with pollen (Llaurens *et al.*, submitted), making the fecundity selection model far from the ecological condition encountered by this population. Wright's model of frequency dependent selection acting on male function only was thus adapted to our ecological conditions, *i.e.* for a small isolated population with a high density.

Our simulations in finite population showed that genetic drift was a key parameter to explain the S-haplotype frequencies in natural populations of sporophytic self-incompatible species. But genetic drift did not manage to explain all the differences observed between theoretical and empirical data for the offspring: under Wright model, we have observed two out of eight S-haplotypes outside the predicted range of frequency in the natural population and one S-haplotype outside the predicted range for the frequency change in one generation. Our limited knowledge of the S-haplotypes number and dominance relationships could also lead to departure of the empirical result from the theoretical expectations. Some recent changes in demography or the spatial distribution of S-haplotypes around the sampled mother, leading to non random mating, may also be part of the explanation. Another explanation could be that the hypothesis of the selective equivalence between S-haplotypes in the frequency-dependent model was not respected. We may imagine that S-haplotype have variable fitness inducing a segregation bias. For instance, the S-locus could be linked to another selected locus, thus modifying the S-haplotype frequency by hitchhiking effect: in *Papaver rhoeas*, Lane & Lawrence (1993) found that the S-locus may be linked to one or more genes that control seed dormancy and supposed that this linkage disequilibrium could modify the S-haplotype frequency distribution. Segregation distortions may occur for some S-haplotypes (Bechsgaard *et al.*, 2004), allowing them to modify their frequency evolution. Or, as suggested by Uyenoyama (2005) and empirically demonstrated by Stone (2004), a sheltered genetic load

could be associated to some S-haplotype and consequently lead to selective heterogeneity between S-haplotypes.

In our population, for most of the S-haplotypes, these other evolutionary forces should have a limited importance because both the allelic frequencies distribution and the allelic richness could be explained by the frequency-dependent selection, the genetic drift and the dominance interactions.

Conclusions

Our combined approach by theoretical simulations and empirical observation allowed us to provide indirect evidence that the major evolutionary processes explaining the polymorphism at the S-locus are negative frequency dependent selection and dominance relationships. We also highlighted that it is necessary to take genetic drift into account to understand allelic richness at a locus under balancing selection.

Acknowledgements

This work was supported by an ATIP research grant from the life science department of the CNRS, and by a PhD grant to VL from CNRS and Région Nord-Pas de Calais. The authors wish to thank Adeline Courseaux for technical support on the S-alleles specific primers PCR conditions and Claire Feutrie, Eric Schmitt and Robert Dron for the plants care in greenhouse.

References

Anderson DR, Burnham KR (2002) Avoiding pitfalls when using information-theoretic methods. *Journal of Wildlife Management* **66**, 912-918.

Bateman AJ (1952) Self-incompatibility systems in angiosperms. 1. Theory. *Heredity* **6**, 285-310.

Bechsgaard J, Bataillon T, Schierup MH (2004) Uneven segregation of sporophytic self-incompatibility alleles in *Arabidopsis lyrata*. *Journal of Evolutionary Biology* **17**, 554-561.

Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant R-genes. *Science* **292**, 2281–2285.

Billiard S (in press) NESSI: A program for numerical estimations in sporophytic self-incompatibility genetic systems. *Molecular Ecology Notes*.

Billiard S, Castric V, Vekemans X (2007) A General Model to Explore Complex Dominance Patterns in Plant Sporophytic Self-Incompatibility Systems. *Genetics* **175**, 1351–1369.

Brooks RJ, Tobias AM, Lawrence MJ (1996) The population genetics of the self-incompatibility polymorphism in Papaver rhoeas. XI. The effects of limited pollen and seed dispersal, overlapping generations and variation in plant size on the variance of S-allele frequencies in populations at equilibrium. *Heredity* **76**, 367–376.

Castric V, Vekemans X (2004) Plant self incompatibility in natural populations : a critical assessment of recent theoritical and empirical advance. *Molecular Ecology* **13**, 2873–2889.

Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research* **70**, 155-174.

Cho S, Huang ZY, Green DR, Smith DR, Zhang J (2006) Evolution of the complementary sex-determination gene of honey bees: Balancing selection and trans-species polymorphisms. *Genome Research* **16**, 1366–1375.

Clauss MJ, Koch MA (2006) Poorly known relatives of Arabidopsis thaliana. *Trends in Plant Science* **11**, 449-459.

de Nettancourt D (2001) *Incompatibility and Incongruity in Wild and Cultivated Plants* Springer-Verlag, Berlin.

Glémin S, Gaude T, Guillemin ML, *et al.* (2005) Balancing Selection in the Wild: Testing Population Genetics Theory of Self-Incompatibility in the Rare Species *Brassica insularis*. *Genetics* **171**, 279-289.

Hatakeyama K, Watanabe M, Takasaki T, Ojima K, Hinata K (1998) Dominance relationships between S-alleles in self-incompatible Brassica campestris L. *Heredity* **80**, 241-247.

Kimura M (1983) *The Neutral Theory of Molecular Evolution* Cambridge University Press.

Kowyama Y, Takashasi H, Muraoka K, *et al.* (1994) Number, frequency & dominance relationships of dominance relationships of S-alleles in diploid Ipomoea trifida. *Heredity* **73**, 275-283.

Lane MD, Lawrence MJ (1993) The population genetics of the self-incompatibility polymorphism in Papaver rhoeas. VII. The number of S-alleles in the species. *Heredity* **71**, 596-602.

Lane MD, Lawrence MJ (1995) The Population-Genetics of the Self-Incompatibility Polymorphism in Papaver Rhoeas .10. An Association between Incompatibility Genotype and Seed Dormancy. *Heredity* **75**, 92-97.

Levene H (1953) Genetic Equilibrium When More Than One Ecological Niche is Available. *The American Naturalist* **87**, 331-333.

Lewontin RC (1978) Heterosis as an explanation for large amounts of genic polymorphism. *Genetics* **88**, 149-170.

Llaurens V, Castric V., Austerlitz F, Vekemans X (submitted) High paternal diversity in the herbaceous perennial herb Arabidopsis halleri despite spatially-restricted pollen dispersal, population isolation, and sporophytic self-incompatibility. *Molecular Ecology*.

Mable BK, Schierup MH, Charlesworth D (2003) Estimating the number, frequency, and dominance of S-alleles in a natural population of Arabidopsis lyrata (Brassicaceae) with sporophytic control of self-incompatibility. *Heredity* **90**, 422-431.

Maruyama T, Nei M (1981) Genetic vaviability maintained by mutation and overdominant selection in finite populations. *Genetics* **98**, 441-459.

Miller HC, Lambert DM (2004) Genetic drift outweighs balancing selection in shaping post-bottleneck major histocompatibility complex variation in New Zealand robins (Petroicidae). *Molecular Ecology* **13**, 3709-3721.

Prigoda NL, Nassuth A, Mable BK (2005) Phenotypic and genotypic expression of self-incompatibility haplotypes in Arabidopsis lyrata suggests unique origin of alleles in different dominance classes. *Molecular Biology and Evolution* **22**, 1609-1620.

Robertson A (1962) Selection for heterozygotes in small populations. *Genetics* **47**, 1291-1300.

Schierup MH, Bechsgaard JS, Nielsen LH, Christiansen FB (2006) Selection at Work in Self-Incompatible Arabidopsis lyrata: Mating Patterns in a Natural Population. *Genetics* **172**, 477–484.

Schierup MH, Vekemans X, Christiansen FB (1997) Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics* **147**, 835-846.

Schopfer CR, Nasrallah ME, Nasrallah JB (1999) The male determinant of self-incompatibility in Brassica. *Science* **286**, 1697–1700.

Stoeckel S, Castric V, Mariette S, Vekemans X (*in prep.*) Unequal allelic frequencies at the self-incompatibility locus within local populations of Prunus avium L.: an effect of population structure?

Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proceeding of the National Academy of Science of the United States of America* **27**, 2419-2423.

Uyenoyama MK (2000) Evolutionary dynamics of self-incompatibility alleles in Brassica. *Genetics* **156**, 351-359.

Uyenoyama MK (2005) Evolution under tight linkage to mating type. *New Phytologist* **165**, 63–70.

Van Oosterhout C, Joyce DA, Cummings SM, *et al.* (2006) Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (Poecilia reticulata). *Evolution* **60**, 2562–2574.

Vekemans X, Schierup MH, Christiansen FB (1998) Mate Availability and Fecundity Selection in Multi-Allelic Self- Incompatibility Systems in Plants. *Evolution* **52**, 19-29.

Wright S (1939) The distribution of self sterility alleles in populations. *Genetics* **24**, 538-552.

Wu J, Saupe SJ, Glass NL (1998) Evidence for balancing selection operating at the het-c heterokaryon incompatibility locus in a group of filamentous fungi. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 12398-12403.

Yokohama S, Hetherington LE (1982) The expected number of self-incompatibility alleles in finite plant populations. *Heredity* **48**, 299-303.

Chapitre 3 - Dépression de consanguinité et fardeau génétique lié au locus d'auto-incompatibilité

Le fardeau génétique lié au locus S semble être une des forces évolutives importantes agissant sur le locus d'auto-incompatibilité. Dans le cadre du stage de M2R de Lucy Gonthier, que j'ai co-encadré avec Sylvain Billiard (MCF-Laboratoire GEPV, Université de Lille1), nous avons modélisé l'accumulation de mutations délétères liées au locus S dans un système d'auto-incompatibilité sporophytique, et également testé l'importance du fardeau lié en fonction du niveau de dominance associé. Nous avons ensuite tenté de mettre empiriquement en évidence ce fardeau en utilisant un protocole de rupture de l'auto-incompatibilité par traitement au dioxyde de carbone. Nous avons mesuré ce fardeau pour des allèles S de dominance différente.

Cette étude a fait l'objet d'une publication : *The sheltered genetic load linked to the S-locus: new insights from theoretical and empirical approaches in sporophytic self-incompatibility.* **Llaurens V.**, Gonthier L, Billiard S. en préparation pour le journal *Genetics*.

The sheltered genetic load linked to the S-locus: new insights from theoretical and empirical approaches in sporophytic self-incompatibility

Llaurens V.*, Gonthier L*, Billiard S.*

Corresponding author: Sylvain Billiard

Email : sylvain.billiard@univ-lille1.fr

* Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8016

Université des Sciences et Technologies de Lille - Lille1

F-59655 Villeneuve d'Ascq cedex

France

Keywords: inbreeding depression, genetic load, self-incompatibility overcome

Running title: Sheltered load linked to sporophytic self-incompatibility system

Abstract

Self-incompatibility (SI) is a common phenomenon in Angiosperm that allows hermaphrodite plants to avoid selfing and reproduction with relatives and thus limit the expression of the inbreeding depression. In homomorphic SI, each individual carries a specificity, encoded by the S-locus, and pistil can only be fertilized by pollen carrying a different specificity. In Brassicaceae, the SI is sporophytic (SSI), meaning that the SI phenotype is encoded by the diploid parental genome at the locus S, so that dominance among S-alleles determined the specificity. Since the heterozygosity is very high, and the recombination rate very low at S-locus, it has been suggested that a sheltered genetic load could be accumulated in the S-locus region. We aimed at demonstrate theoretically and empirically the existence of a sheltered load linked to the S-locus in SSI. Simulations in finite population showed that deleterious mutations could be accumulated and linked to specific S-alleles. We also demonstrated that this load is very high when associated with dominant S-allele, low when associated with intermediary S-allele and null for recessive S-alleles. By CO_2 gas treatment, we overcame self-incompatibility in *Arabidopsis halleri* plants and checked for the expression of a sheltered load in seed and plant development. We found a significant sheltered load associated to the most dominant S-allele (S15) out of three S-alleles tested, in accordance with the theoretical prediction. This sheltered load seemed to be expressed at the early stage of the plant development and have a stronger impact on plant fitness than background inbreeding depression. The existence of a sheltered load in SSI could have important evolutionary consequence like diversification rate in S-lineage, viability of small populations or SI persistence.

Introduction

Inbreeding depression is the decline in fitness of inbred individuals compared to outbred individuals. The main genetic mechanism behind inbreeding depression is believe to be the expression in homozygotes of partially recessive deleterious mutations Charlesworth (1999). Interactions between inbreeding depression and mating systems have an important impact on their respective evolution. On one hand, inbreeding depression is an essential factor often invoked in the evolution of mating systems, especially for the maintenance of outcrossing versus selfing despite the two-fold cost of sex (Charlesworth & Charlesworth, 1987). On the other hand the mating system determines homozygosity in a population which affects the purging of recessive deleterious alleles (Byers & Waller, 1999).

Among all genetic systems favoring outcrossing in plants, homomorphic self-incompatibility is the most widely distributed among Angiosperm families (de Nettancourt, 2001). It allows the avoidance of selfing and mating with relatives through a specific recognition between pollen and pistil. It is controlled by proteins generally encoded by a single genomic region, the S-locus. Each plant of a self-incompatible population expresses an S-specificity and is unable to mate with other plants expressing the same specificity. The S-specificity can be either controlled by the pollen haploid genome and the parental diploid genome for the pistil in gametophytic self-incompatibility (GSI) or both the diploid parental genomes in sporophytic self-incompatibility (SSI). As a consequence of this mating system a high heterozygosity is expected in populations at the S-locus in particular but also at other loci in the S-locus genomic region {Kamau, 2007 #247}. Consequently, partially recessive deleterious mutations are less efficiently purged, especially at loci close to the locus controlling self-incompatibility, because of a "sheltering effect", which can drastically increase inbreeding depression in small populations (Glémin *et al.*, 2001). This effect can thus lead to the generation of a "sheltered load" (Uyenoyama, 1997). Such a sheltered load may

have important evolutionary consequences. First, the appearance of new S-alleles can be slowed down because of shared deleterious mutations between the ancestral and the new S-allele (Uyenoyama, 2003). Second, a strong sheltered load can largely extend the conditions of maintenance of GSI systems (Porcher & Lande, 2005). Third, a sheltered load can largely increase the inbreeding depression in a small population (Glemin *et al.*, 2001) what can have large consequences for endangered species and the viability of their populations. The sheltered load could thus have stronger effect on small population than the genomic inbreeding depression that is supposed to be lower in small than in large populations (Bataillon & Kirkpatrick, 2000).

The magnitude of the sheltered load should depend on the size of the genomic region experiencing enhanced heterozygosity because of the linkage to the S-locus and also on the number of genes affecting fitness in that region. From an analysis of recombination rates in the S-locus genomic region in *Arabidopsis lyrata*, a species with SSI, Kawabe (2006) suggested that the number of genes in such a genomic region is not high enough for a sheltered load to be large compared to the overall genomic load. To the best of our knowledge, a single study specifically demonstrated the existence of a sheltered load in a SI species, linked to two out of seven S-alleles investigated in *Solanum carolinense*, a species with GSI (Stone, 2004). Hence, the plausibility of a sheltered load is far from complete, and its existence still remains debated.

In SSI, there are dominance interactions among S-alleles as observed in *Ipomoea trifida* (Kowyama *et al.*, 1994), *Brassica campestris* (Hatakeyama *et al.*, 1998), *A. lyrata* (Mable *et al.*, 2003) and in *A. halleri* (Llaurens *et al.*, *submitted*). The impact of these dominance interactions on the formation of a sheltered load have not been investigated neither theoretically nor empirically, but may potentially be strong. Indeed, recessive S-alleles are expected to be more often at homozygous state in natural populations than dominant alleles

(Schierup *et al.*, 1997), and so may be more prone to purging of deleterious recessive mutations, what should limit the sheltering effect. The sheltered genetic load could thus differ depending on the dominance level of the S-allele to which it is linked.

In this study, we aimed at testing the existence and the strength of the sheltered genetic load linked to the S-locus in relation with the dominance level in the sporophytic self-incompatibility system. We focused on *Arabidopsis halleri*, a Brassicaceae with SSI. In Brassicaceae, the S-locus is composed of two main genes: *SCR* (also called *SP-11*), encoding a cystein-rich protein of the pollen envelope and *SRK*, encoding a kinase receptor situated across the membrane of the papilla cell. As expected, the heterozygosity at the S-locus has been reported to be very high in different species: in *A. halleri*, we have observed an heterozygosity of H_{obs} between 0.39 and 0.57 at the S-locus, depending on the population considered (Leducq *et al.*, *in prep.*). Alleles at *SRK* and *SCR* genes are tightly linked since they are located close to each other and since recombination suppression in the S-locus region has been suggested in different studies: In *Brassica* (Casselman *et al.*, 2000), in *Arabidopsis lyrata* (Kawabe *et al.*, 2006 and Kamau & Charlesworth, 2005). Since the recombination rate is low and heterozygosity is high in the S-locus region in Brassicaceae, conditions may be encoutered for the existence of a substantial sheltered genetic load.

In this paper, we first used stochastic simulations of the dynamics of mutations linked to the S-locus in a general SSI system to investigate the theoretical conditions for the accumulation of a sheltered load of deleterious mutations. We used these results to test the prediction that a greater load should accumulate linked to dominant S-alleles than linked to recessive S-alleles. We then empirically tested these predictions by performing controlled pollinations in the self-incompatible *A. halleri* to specifically measure the magnitude of the sheltered load of three S-alleles with different dominance levels.

Materials and methods

Simulations

We aimed at testing if the accumulation of sheltered genetic load in the genomic region of the S-locus is possible. We also would like to know if the strength of this sheltered load depends on the dominance of the linked S-allele. We thus looked at the evolution of frequencies of deleterious mutation appearing at genetic loci fully linked to a sporophytic self-incompatibility locus. We performed simulations of a panmictic population of size N diploid individuals with non-overlapping generations. Each individual is defined by its genotype in a given genomic region within which there was no recombination. This region contains the S-locus, which controls the S-specificity expressed in pollen and stigma, as well as a region we called "region D" and which contains L loci at which two alleles could segregate: 0 and 1, allele 1 being deleterious and completely recessive. The life cycle in simulations had four steps: (i) gametogenesis: N adult individuals produced an infinity of ovules and pollen; (ii) syngamy: an infinity of zygotes are produced from compatible crosses in the population. Crosses between ovules and pollen were compatible when the expressed specificities of pollen and stigma were different. The frequency of each genotype at the S-locus in seeds was computed following the general equations given in Billiard *et al.* (2007). Since this step only depends on the genotype at the S-locus, the same equations were used to compute the genotypic frequencies at the L loci. The reproduction regime was supposed to follow the "fecundity selection" model (Vekemans *et al.*, 1998), which assumes that frequency-dependent selection acts through both pollen and stigma; (iii) viability selection and regulation: we assumed that the survival probability p of a zygote depends on the number n of loci among the L which are homozygotes for the deleterious mutation (1) as follows: we assumed each individual mutation at the homozygous state to decrease the survival probability by s and mutations at different loci to interact multiplicatively. Finally, the

survival probability was computed as $p = (1\text{-}s)^n$. We randomly drew one zygote following a multinomial sampling with the genotypic frequencies after syngamy as parameters; we computed p for this zygote and we randomly determined whether this individual survived; we repeated these steps until N surviving individuals were obtained; (iv) mutation: we defined the mutation rate μ from allele 0 to allele 1 at each locus in region D, and the reverse mutation rate η from allele 1 to allele 0. We randomly drew the total number of mutations occurring in the population following a Poisson distribution. We then randomly assigned the mutation to a chromosome and a locus in region D.

We considered a simple model of SSI, with classical dominance pattern, similar to observation performed in *A. halleri* (Llaurens *et al.*, in prep). We simulated a population with five S-alleles and we assumed that no new S-allele could appear by mutation. These S-alleles had different dominance levels and belonged to three different dominance classes: two were dominant over all the others and codominant relatively to each other, one was recessive towards all others, and two were intermediate and codominant to each other. Those dominance relationships were assumed in both pollen and stigma.

At initial state, the whole D region was monomorphic and had the neutral allele 0 at all L loci. We first ran simulations without deleterious mutations until deterministic equilibrium for S-alleles frequencies were reached, defined as all allelic frequencies changes in one generation lower than 10^{-3}. Mutations at the L loci were then allowed. We performed simulations for different values of L, N and s. The mutation rates were fixed at $\mu=10^{-4}$ and $\eta=10^{-5}$ mutations per generation. Each simulation was performed with 100 independent replicates of 100,000 generations and the frequency of the deleterious alleles at all L loci was measured every 1,000 generations. In the case where L=1 we estimated the frequency spectrum of the deleterious allele in the total population and inside a dominance class. We will use the term "fixation

inside an allelic class" when all copies of a given S-allele in the population were linked to a deleterious mutation at a given locus among the *L* locus.

Plant material

Arabidopsis halleri (Brassicaceae) is a diploid species taxonomically closely related to the two model species *Arabidopsis thaliana*, and *Arabidopsis lyrata* (Clauss & Koch, 2006). It is a European species growing both in mountainous areas and on heavy-metal polluted areas. We used plants issued from seeds collected in a single natural population of *A. halleri* located in Nivelle (France) whose matting pattern has been extensively studied (Llaurens *et al.*, *submitted*). Estimates of the selfing rate were 0% based on progeny analyses from this natural population (Llaurens *et al.*, *submitted*). In controlled self-pollination performed in greenhouse on *A. halleri* plants from this population, fruits were formed in 3% of all attempts (Llaurens *et al.*, *in prep.*). These results suggested that the self-incompatibility system was thus apparently fully functional in individuals from this population, allowing a sheltered load to be potentially accumulated.

Self-incompatibility overcoming

We used the following experimental procedure to disentangle the sheltered load of deleterious mutations from background inbreeding depression, which is defined here as the effect of mildly deleterious mutations located in the whole genome. Three different types of crosses were performed to measure genetic load, by manual transfer of pollen from different origins to stigmas: (1) in enforced selfing crosses, with pollen from the same plant; (2) in enforced incompatible crosses with pollen from a different individual sharing the same genotype at the S-locus, which were thus naturally incompatible and (3) in outcrossing controls, with pollen from a different individual with a different genotype at the S-locus. The decline in fitness in

offspring from enforced selfing and enforced incompatible crosses compared to the outcrossing control was interpreted as the effect of inbreeding depression. We hypothesized that offspring from enforced selfing crosses should be affected by both background inbreeding depression and sheltered genetic load while offspring from enforced incompatible crosses should only be affected by the sheltered load. Finally, offspring from outcrossing controls should be affected neither by the background inbreeding depression nor the sheltered load. Hence, if the sheltered load is high enough relatively to the background inbreeding depression to be distinguished, we should observe a decline in fitness significantly higher for the enforced selfing crosses than for the enforced incompatible crosses, and also significantly higher for the enforced incompatible crosses compared to the outcrossing control crosses.

We used plants with known genotype at the S-locus: genotypes were determined by PCR with S-allele specific primers as described below. We will use the abbreviation S instead of S-allele AhSRK throughout the paper for simplification. We used two different genotypes: a first class of individuals (n=5) with genotype [S01; S15] and a second class of individuals (n=5) with genotype [S01; S02]. A previous study on the individuals of this population showed that the S-allele S01 was the most recessive allele of the population, the S-allele S02 had intermediary dominance level and the S-allele S15 had one of the highest dominance level (Llaurens et al., *in prep.*). The individual used for the control crosses had genotype [S04; S20]. We performed twenty selfing crosses per individuals, eight incompatible crosses per plant, namely two with each other plant of the same genotype at the S-locus, and four control crosses per plant.

All crosses were performed under CO_2 gas treatment in order to overcome self-incompatibility as described in Nakanishii *et al.* (1969). Although this method was commonly used in seed companies to maintain inbred lines of Brassicaceae, we used this method for the

first time in *Arabidopsis halleri*. We performed controlled pollinations under binocular magnifying glass. Anthers of pollen donor were collected with tweezers. The pistil was distempered with the pollen of the collected anthers. Then the pollinated plants were placed overnight in an atmosphere with 5% CO_2. Pistils were marked individually and measured seven days after pollination. We considered a cross as successful, either between compatible or incompatible genotypes, when the size of the silic after 7 days was larger than 6.1 mm and when seed development was observed. When the threshold of 6.1 mm was reached, the fruit was developing and we could collect seeds. We then grew the offspring in greenhouse with constant temperature T=20°C and photoperiod 16h day/ 8h night. Plants were randomly located in the greenhouse table.

To check if our controlled pollination were not contaminated by pollen coming from other sources, we checked if the offspring genotype at seven neutral markers were convenient with the genotype of the parents used for the pollination. This test was performed on 87 offsprings from six Selfing crosses, 55 offsprings from six Incompatible crosses and 13 offsprings from one Control cross. We genotyped these individuals for seven microsatellite loci (NGA112, NGA361, GC22, MDC16, ICE13, H117 and Athzfpg) following the method described in Llaurens *et al.* (*Submitted*). We observed contamination in less than 3% of the offspring analyzed.

Estimations of background inbreeding depression and sheltered genetic load

We assessed the reproductive success of offspring from the different types of crosses, based on several proxis of fitness: the number of seeds per fruit and the germination rate were compared between the selfing and the incompatible crosses with a χ^2 test with the expected value given by the control crosses; the germination time, the time of cotyledons appearance, the death rate after appearance of cotyledons and the time of appearance of the first leaf, were

compared with a Kruskal-Wallis test since the data did not follow a normal distribution; The length and width of the biggest leaf were finally measured two weeks after the germination, and we compared them between all crosses by an ANOVA. To check if the decline in fitness observed for phenotypic variables between selfing and incompatible crosses is specifically due to one allele, in other words if the sheltered load is allele-specific, we performed two-factors tests. The first factor was the type of cross (selfing or incompatible) and the second was the genotype at S-locus. For parameters with a non-Normal distribution (germination time, cotyledon appearance time and first leaf appearance time), we used the Sheirer-Ray-Hare extension of the Kruskal-Wallis test described in Sokal (1995). In other cases (length and width of the largest leaf), we used a two-factors ANOVA.

To detect early-acting inbreeding depression, we computed the segregation of S-alleles in the offspring. We extracted DNA from offspring leaves collected 2 weeks after germination. Leaf samples were then oven-dried at 55°C, and DNA was extracted using the Nucleospin® Multi-96 Plant extraction kit from Macherey-Nagel®. We used primers specifically targeted towards each S-allele: each pair of primer allowed amplifying specially one single S-allele when present in the sample considered (for details see Llaurens *et al.*, *in prep.*). The reaction mixture (15µL) contained 20 ng DNA, 1X buffer (Applied biosystem®), 2 or 2.5 mM of $MgCl_2$ (depending on the primer pair), 200 µM of Fermentas® DNTP mix, 200 µg/mL of BSA, 0.2 µM of each microsatellite primer, and 0.025 U/µL of *Taq* polymerase (Amplitaq DNA polymerase, Applied biosystem®). The amplification was carried out 5 min at 95°C, 35 cycles of 40 s at 95°C, 40 s at Tm°C (specific for each pair of primer, Tab.I), 40 s at 72°C, and one cycle of 10 min at 72°C and performed in MJ research PTC 200 thermocycler®. PCR products were mixed with loading dye then deposed in 2% agarose gel, colored by ethidium bromure. The migration was performed in TBE buffer at 110mV during 45 minutes. The PCR products were visualized with fluorescence in UV and sized thanks to migration with sizing

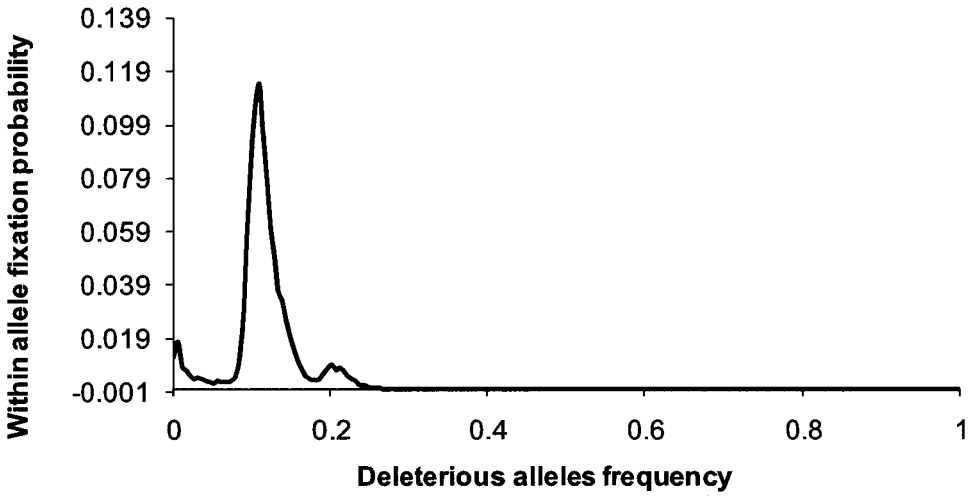


Figure 1:Frequencies distribution of deleterious mutations in the region D, linked to all S-alleles (population size: N=1,000; number of loci in the D-region: L=1 and selection coefficient: s=0.1)

marker. We used DNA from an individual carrying the targeted S-allele as positive control for each PCR. We finally tested if the observed segregation at the S-locus followed Mendelian expectation by a Chi^2 test.

Results

Predicted sheltered genetic load

Our model of evolution of the D-region, fully linked to the S-locus, allowed us to analyse the frequency spectrum of deleterious alleles linked to the S-locus. Figure 1 summarized the allele frequency spectrum for linked deleterious at a single linked locus (L=1). Overall, deleterious alleles associated with an S-alleles were observed at intermediate frequency (0.11) in about 10% of the simulations. We then looked at the influence of the dominance level of the associated S-allele on the probability of fixation of the deleterious allele inside an allelic class at a single locus of the D-region (Fig.2). As shown on the left part of Fig. 2, all S-alleles regardless of their dominance level were linked to some deleterious alleles in low frequency due to mutation. In sharp contrast, the right part of Fig. 2 revealed that S-alleles of different dominance levels differed greatly in their probability to be entirely associated with deleterious alleles. While the most recessive S-allele was never associated with a fixed deleterious allele, alleles at intermediate dominance level were occasionally linked to a deleterious allele that was fixed within an S-allele (about 8% of cases), and dominant S-alleles were commonly (38%) associated with a fixed deleterious allele. Since we supposed two S-alleles in the most dominant class and since they are selectively equivalent, this result means that there is a probability of 76% that all copies of one of the two dominant S-alleles are associated with the deleterious allele. This probability is lower for the intermediate S-alleles (8% of fixation with the intermediate dominance class and hence 16% that one of the intermediate S-allele is fixed for the deleterious allele). Finally, we never observed fixation of the deleterious allele for the

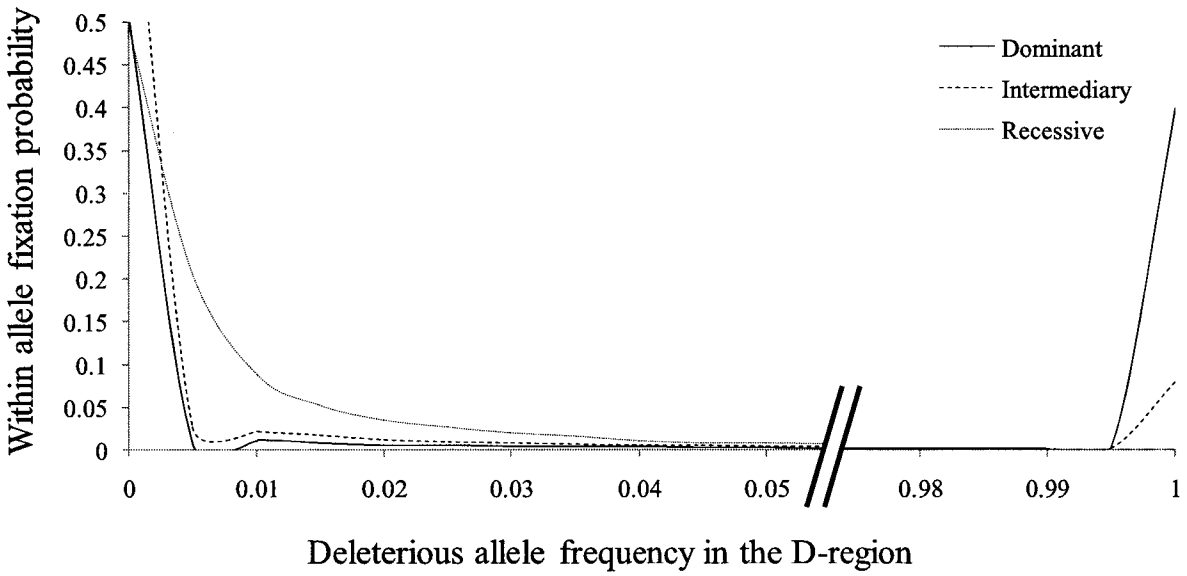


Figure 2: Frequencies distribution of deleterious mutations in the region D, within linked S-alleles belonging to different dominance classes.
(population size: N=1,000; number of loci in the D-region: L=1 and selection coefficient: s=0.1)

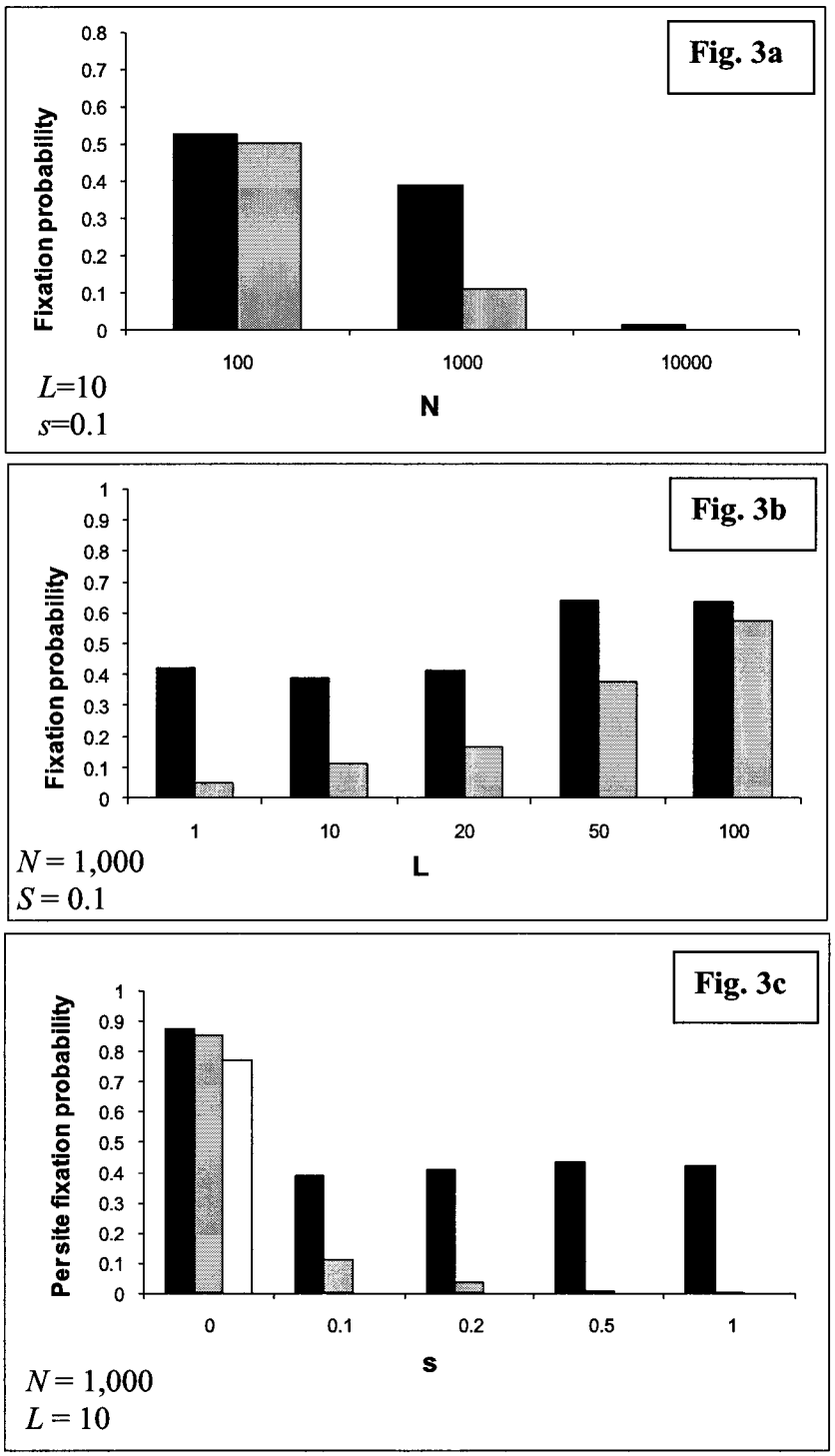


Figure 3: Deleterious allele fixation probability in the region D with respect to population size *N* (**Fig. 3a**), number of loci *L* at in the region D (**Fig. 3b**) and selection coefficient *s* (**Fig. 3c**). Black bars represents region D associated with a dominant S-allele, grey bars with an intermediary S-allele and white bars with a recessive S-allele.

most recessive S-allele. The probability of fixation within an S-allelic class of the deleterious allele thus depended on the dominance of the linked S-allele. We therefore predict that the sheltered load should be mainly associated to the most dominant S-alleles.

Figure 3 revealed that this prediction held true for a large range of parameters including the population size N, the number of linked fitness loci L and the strength of selection s. However, all these parameters had important impact on the accumulation of the linked load. The fixation probability of deleterious alleles linked to the S-alleles decreased when N increased (Fig. 3a), hence small populations should show a more important linked load. On the contrary, the fixation of deleterious alleles was more likely when L increased, (Fig. 3b), deleterious alleles should therefore accumulate more readily when a large number of loci contribute to the sheltered load, allowing a sheltered load with stronger effect on homozygote fitness. Figure 3c shows that the strength of the purifying selection limited the accumulation of the sheltered genetic load linked to the intermediary S-alleles but had small effect on the sheltered load linked to the dominant S-alleles.

Estimations of background inbreeding depression and sheltered genetic load

In Table 1 we compared six fitness variables in the offspring from the three kinds of crosses for two different maternal genotypes: [S01; S15] and [S01; S02]. Concerning the maternal genotype [S01; S15] the mean number of seeds per cross, and the length and width of the largest leaf were different between cross types (Tab. 1a) while the other variables, except cotyledons appearance times, show no significant difference between selfing and incompatible crosses but both were significantly different from control. The difference between selfing and incompatible cross should be interpreted as a consequence of the background inbreeding depression whereas difference between incompatible and control

Table 1: Comparison of fitness parameters of the offspring of the 3 type of crossing (Selfing, Incompatible and Control)

Crossing with the maternal genotype [S01; S15] (**Tab. 1a**), crossing with the maternal genotype [S01; S02] (**Tab. 1b**)

NS (non-significant): $P>0.05$; ***: $P \leq 0.05$

Table 1a: Crossing with the maternal genotype [S01; S15]

	Selfing	Incompatible	Control	Significance
Mean seeds number	2.42 (a)	3.95 (b)	9.86 (c)	***
Germination rate	0.88	0.94	0.87	NS
Germination time (days)	8.86 (a)	8.35 (a)	10.34 (b)	***
Cotyledons appearance time (days)	1.83 (a)	1.78 (ab)	1.56 (b)	***
First leaf appearance time (days)	6.65 (a)	6.35 (a)	5.33 (b)	***
Length of the largest leaf (cm)	0.67 (a)	0.77 (b)	0.86 (c)	***
Width of the largest leaf (cm)	0.57 (a)	0.67 (b)	0.80 (c)	***

Table 1b: Crossing with the maternal genotype [S01; S02]

	Selfing	Incompatible	Control	Significance
Mean seeds number	3.39 (a)	4.24 (ab)	7.16 (b)	***
Germination rate	0.74	0.81	0.86	NS
Germination time (days)	9.30	9.51	9.41	NS
Cotyledons appearance time (days)	1.84 (a)	1.69 (a)	1.34 (b)	***
First leaf appearance time (days)	5.81 (a)	5.34 (b)	5.23 (b)	***
Length of the largest leaf (cm)	0.76	0.78	0.76	NS
Width of the largest leaf (cm)	0.70	0.71	0.71	NS

Table 2: Effect of genotype at the S-locus on fitness parameters for the maternal genotype [S01; S15] (**Tab. IIIa**) and for the maternal genotype [S01; S02] (**Tab. IIIb**)

NS (non-significant): $P>0.05$; ***: $P \leq 0.05$

Table 2a: Effect of genotype at the S-locus on fitness parameters for the maternal genotype [S01; S15]

	S01-S01	**S01-S15**	**S15-S15**	**Significance**	
Germination time	8.61	8.27	8.94	P=0.363	NS
Cotyledons appearance time	1.86	1.71	1.78	P=0.210	NS
First leaf appearance time	6,220 [(a)]	6,441 [(ab)]	7 [(b)]	$P=9.8.10^{-4}$	***
Length of the largest leaf	0.624	0.618	0.568	P=0.055	NS
Width of the largest leaf	0.729	0.728	0.650	P=0.042	***

Table 2b: Effect of genotype at the S-locus on fitness parameters for the maternal genotype [S01; S02]

	S01-S01	**S01-S02**	**S02-S02**	**Significance**	
Germination time	9.61	8.53	9.56	P=0.278	NS
Cotyledons appearance time	11.52	10.17	11.27	P=0.084	NS
First leaf appearance time	15.34	14.08	15.38	P=0.109	NS
Length of the largest leaf	0.674	0.731	0.683	P=0.052	NS
Width of the largest leaf	0,739 [(ab)]	0,799 [(a)]	0,732 [(b)]	P=0.0122	***

should be due to the sheltered load only. Our results thus suggest an effect of both background inbreeding depression and sheltered load on fitness. However, the sheltered load seemed to have greater effect than the background inbreeding depression: for example for the seed number, the sheltered load caused a 59.9% reduction of fitness compared to the control whereas the background inbreeding depression caused only a 15.5% reduction of fitness. For the maternal genotype [S01; S02], no clear patterns between the three types of crosses were found except for the number of seeds, the appearance time of cotyledons and first leaf where results for selfing crosses were significantly different from control crosses (Tab. 1b).

Table 2 shows in what extent the differences in fitness-related variables between selfing and incompatible crosses were associated to the S-locus genotype. For the offspring of maternal genotype [S01; S15], the individuals with the homozygous genotype [S15; S15] showed a significantly slower first leaf appearance time (Tab. 2a). A similar trend was observed for leaf size although it was not significant (P=0.055). The heterozygote had intermediary values between the homozygotes [S01; S01] and [S15; S15]. For the offspring of the maternal genotype [S01; S02], we observed a significant decrease of the leaf width for the individuals with the homozygous genotype [S02; S02] (Tab. 2b).

The segregation of S-alleles in selfing and incompatible crosses in the case of the maternal genotype [S01; S15] highly significantly departed from mendelian segregation with an excess of homozygotes [S01; S01] and a lack of homozygotes [S15; S15] and heterozygotes [S01; S15] (Tab. 3a). There was thus a segregation bias against allele S15 that could be interpreted as an effect of sheltered load on seed formation. This sheltered also to act on heterozygotes but with less intensity, suggesting that deleterious alleles should be partially recessive. No such significant departure from the mendelian segregation was found for the genotype [S01; S02] except for the incompatible crossing, with an excess of homozygotes [S01; S01] and a corresponding lack of heterozygotes [S01;S02] (Tab. 3b). The test of segregation at S-locus in

Table 3: Proportions of each genotype at the S-locus in the offspring of the Selfing and Incompatible crossing for the maternal genotype [S01; S15] (**Tab. 3a**) and for the maternal genotype [S01; S02] (**Tab. 3b**)

NS (non-significant): $P>0.05$; ***: $P \leq 0.05$

Table 3a: Segregation at the S-locus in the offspring of the crossing with the maternal genotype [S01; S15]

	S01-S01	S01-S015	S15-S15	Mendelian segregation	
Selfing	0.391	0.466	0.143	$P = 3.3.10^{-5}$	***
Incompatible	0.367	0.438	0.195	$P = 8.4.10^{-3}$	***
Mean	0.381	0.453	0.166	$P = 4.7.\ 10^{-7}$	***

Table 3b: Segregation at the S-locus in the offspring of the crossing with the maternal genotype [S01; S02]

	S01-S01	S01-S02	S02-S02	Mendelian segregation	
Selfing	0.208	0.528	0.264	$P = 0.33$	*NS*
Incompatible	0.350	0.390	0.260	$P = 0.02$	***
Mean	0.257	0.480	0.262	$P = 0.75$	*NS*

the control crosses allowed us to check the segregation of the studied S-allele (S01, S02 and S15) in compatible crossings. For the crossing [S01; S15] x [S04; S20], we showed no significant departure from the mendelian expectation (P=0.06); S15 had a slightly lower frequency than expected but it was not significant. For the crossing [S01; S02] x [S04; S20] a negative segregation bias for S04 was detected (P=0.02). Our results hence suggest that the segregation of S-alleles we are focused on is not biased.

Discussion

Existence of genomic and sheltered load

The model showed that deleterious recessive mutations could accumulate in a region closely linked to S-alleles and so even more easily when the number of susceptible genes where deleterious mutations could occur is large. The results obtained with controlled pollination and forced fertilization between compatible and incompatible individuals of *A. halleri* are consistent with the model predictions. First we observed that there is a background inbreeding depression since for the two maternal genotypes and especially for [S01; S15] since we showed that the fitness-related variables of offspring from selfing crosses were significantly different from control crosses and, in certain cases, from incompatible crosses (Tab. 1). Second, our results suggest the existence of a strong sheltered load associated to the S-allele S15. Indeed, we showed that most fitness-related variables in offspring from incompatible crosses had intermediate values relatively to offspring from selfing and control crosses (Tab. 1). This observation is consistent with the hypothesis of the existence of a sheltered load. If there was no sheltered load or if it was not strong enough relatively to the genomic load we should have observed no difference between offspring from incompatible or control crosses. Furthermore, we showed a negative segregation bias against S15 in offspring from both selfing and incompatible crosses (Tab. 3). We showed as well that the differences between

fitness-related variables were mainly due to genotypes bearing S15 (Tab. 2). Hence, it is likely that the sheltered load is mainly linked to S15. Our results also suggest that at least some deleterious alleles associated with S15 are only partially recessive since there are fewer heterozygotes than expected and since S15 is in lower frequency in offspring from control crosses. Finally our results suggest that there may be no sheltered load linked to allele S02 since neither a clear pattern of difference between crosses have been found nor a clear segregation bias.

When is the sheltered load preponderant in the life cycle?

The major differences between the three types of crosses were found for the seeds number and the segregation bias. These results suggest that sheltered load effect on fitness is the largest between fertilization and germination. However, one could argue that the difference in the number of seeds between the three types of crosses can be due to partial self-incompatibility overcoming. Indeed, if our method of self-incompatibility overcoming is only partial, only a small proportion of pollen could germinate and be candidate for the fertilization of ovules, what could decrease the number of fertilization events possible and thus reduce the seed set. Nonetheless, observations on *Brassica*, suggested that a large amount of pollen managed to enter the pistil after CO_2 gas treatment (Nakanishii & Hinata, 1973) so that even if pollen germination was somewhat reduced it should not limit the number of fertilization events. But Nakanishii & Hinata (1973) also pointed out that the self-incompatibility overcome though CO_2 gas treatment also depends on the strain of the plant treated and could thus differ among genotype at S-locus. Our segregation results could thus be biased if the response to CO_2 gas treatment depends on the S-alleles. Several studies have tried to identify the genes implied in the CO_2 response: genetic analyses in *Raphanus sativus* indicated that this gene should not correspond to S-genes (Niikura & Matsuura, 2000) whereas Kwun *et al.* (2004) showed that

in *Brassica campestris* the expression of 2% of pistil genes, including the *SLG* gene, which belong to the S-locus were modified by CO_2 gas treatment. Since *SLG* gene has never been found in the Arabidopsis genus (Schierup *et al.*, 2001), and since several genes seemed to be implied in the CO_2 response, we could thus assumed the segregation bias observed in this study did not only depend on efficiency of the CO_2 gas treatment. Moreover, the difference in the number of seeds among incompatible and control in the maternal genotype [S01; S02] was not significant, suggesting that at least in this case the CO_2 gas treatment should not limit seed set (Tab. 1b). Moreover, the significant difference in seed number between selfing and incompatible crosses in maternal genotype [S01; S15] could neither been explained by the efficiency of the CO_2 gas treatment (Tab. 1a). Finally, our results show that it is likely that the sheltered load in the S-allele S15 has a strong effect during the early stage of the plant development.

Does the sheltered load depend on the dominance level of the considered S-allele?

We found no sheltered load linked to the S-allele S1. However, we found a strong sheltered genetic load linked to the S-allele S15 whereas the evidence for a sheltered load linked to the S-allele S02 was unclear. We could thus infer that the sheltered load linked to S02 was probably null or very low compared to the sheltered load carried by the S-allele S15. The S-allele S15 was one of the most dominant alleles in the considered population of *A. halleri*, whereas S01 was the most recessive and S02 had an intermediate dominance level (Llaurens et al., *in prep.*). The results of our model suggest that these differences in sheltered load level between S-alleles may be due to their dominance level. Indeed, our model showed it was more likely to accumulate deleterious mutations linked with the most dominant S-alleles than with more recessive S-alleles and moreover no sheltered load was expected with the most recessive alleles (Fig. 1). In our theoretical model however, we did not consider the

expression of deleterious effect at heterozygote state. In the model developed for gametophytic self-incompatibility, Glémin *et al.* (2001) considered an expression level *h* and found that the accumulation of a sheltered genetic load was possible for partially recessive mutations (h=0.2). We thus do not expect that taking into account partial recessivity should change our qualitative results.

Genetic bases of the sheltered genetic load.

The existence of a sheltered genetic load linked to the self-incompatibility locus remains debated: Kawabe (2006) argued that in *A. lyrata*, the low recombinant region surrounding the S-locus was about 600 kb which roughly correspond to 120 genes, based on the *A. thaliana* genome. They concluded that so few genes potentially susceptible to be responsible of the sheltered load could certainly not generate any strong enough decline in fitness to be distinguishable from the genomic load. However, our theoretical model suggested that the fixation of deleterious mutations was possible even if the number of possible sites was low (Fig. 3b).

Our empirical results also suggest that the sheltered genetic load has a greater effect than the background inbreeding depression in *A. halleri.* Hence, since the number of genes implied is certainly low according to the different estimation of the size of the region under the influence of the S-locus in Brassicaceae (based on the review of Uyenoyama, 2005), 45 genes seemed to co-segregate with the S-locus in *Brassica*), we conclude that the deleterious mutations linked to S15 in *A. halleri* must have a strong effect, certainly during the early developmental stages. Uyenoyama (2005) also reviewed that several genes with known functions co-segregated with S-loci in different species. These genes had sometimes important functions like seed dormancy in *Papaver rhoeas* (Lane & Lawrence, 1995), so that a single mutation in one of these genes could have a severe effect at the individual level. Our results also suggest

that it is not necessary that strongly deleterious mutations are totally recessive to accumulate closely to S-alleles since the deleterious mutation associated to the allele S15 seems only partially recessive. Indeed, if an S-allele is linked with any strongly deleterious allele, this S-allele will decrease in frequency. Then, since it is rarer than expected it will be favored by the negative frequency dependent selection acting during the reproductive stage. This effect will further contribute to increase the frequency of deleterious alleles and thus allowed the maintenance of the sheltered load.

Evolutionary consequences of the sheltered genetic load

Our results showed that the sheltered load is likely to exist in *A. halleri*. This sheltered load could play an important role in the maintenance of self-incompatibility system. Porcher & Lande (2005) pointed out that a strong sheltered load combined to a high background inbreeding depression were necessary to understand the persistence of the gametophytic self-incompatibility system. It could now be interesting to estimate if the sheltered load and background inbreeding depression observed here allow the persistence of the sporophytic self-incompatibility system.

We have also demonstrated that an important sheltered load could be accumulated in small populations. We have observed in our simulations analyses that some deleterious alleles linked to the S-locus could be shared among S-alleles, in particularly for dominant S-alleles (data not shown). The sheltered load could be therefore expressed in natural population. It could then be a threat for small self-incompatible populations.

The sheltered load could also influence the dynamics of self-incompatibility alleles: since a new S-allele should share deleterious alleles with its ancestor, the sheltered load is thought to slow down the appearance of new alleles along phylogenetic branches (Uyenoyama, 2003).

Our results indicate that the sheltered genetic load depend on the level of dominance of the associated S-allele. Prigoda *et al.* (2005) suggested that in *A. lyrata*, S-alleles belonging to a class of dominance have a unique origin. Hence, if the sheltered load is large, we can expect that the differentiation rate of the S-alleles should differ among dominance classes, with a slower rate in dominant than in recessive S-allelic class.

Conclusions

This study is the first demonstration of the existence of a sheltered genetic load linked to the S-locus in sporophytic self-incompatibility system. We have also suggested a link between the strength of the sheltered load and the dominance of the associated S-allele. This should be confirmed by further experiments with S-alleles at different dominance level.

Acknowledgements

This work was supported by an ATIP research grant from the life science department of the CNRS, and by a PhD grant to VL from CNRS and Région Nord-Pas de Calais. The authors wish to thank Adeline Courseaux for technical support and Angelique Bourseaux, Eric Schmitt and Robert Dron for the plants care in greenhouse. We also thanks Vincent Castric and Xavier Vekemans for useful comments on the manuscript.

References

Bataillon T, Kirkpatrick M (2000) Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter. *Genetical Research* 75, 75-21.

Billiard S, Castric V, Vekemans X (2007) A General Model to Explore Complex Dominance Patterns in Plant Sporophytic Self-Incompatibility Systems. *Genetics* 175, 1351–1369.

Byers DL, Waller DM (1999) Do plant populations purge their genetic load ? Effects of population size and mating history on inbreeding depression. *Annual Review of Ecology, Evolution, and Systematics* 30, 479-513.

Casselman AL, Vrebalov J, Conner JA, *et al.* (2000) Determining the Physical Limits of the Brassica S Locus by Recombinational Analysis. *The Plant Cell* 12, 23–33.

Charlesworth B, Charlesworth D (1999) The genetic basis of inbreeding depression. *Genetical Research* 74, 329-340.

Charlesworth D, Charlesworth B (1987) Inbreeding Depression and its Evolutionary Consequences. *Annual Review of Ecology and Systematics* 18, 237-268.

Clauss MJ, Koch MA (2006) Poorly known relatives of Arabidopsis thaliana. *Trends in Plant Science* 11, 449-459.

de Nettancourt D (2001) *Incompatibility and Incongruity in Wild and Cultivated Plants* Springer-Verlag, Berlin.

Glemin S, Bataillon T, Ronfort J, Mignot A, Olivieri I (2001) Inbreeding depression in small populations of self-incompatible plants. *Genetics* 159, 1217-1229.

Hatakeyama K, Watanabe M, Takasaki T, Ojima K, Hinata K (1998) Dominance relationships between S-alleles in self-incompatible Brassica campestris L. *Heredity* 80, 241-247.

Kamau E, Charlesworth D (2005) Balancing Selection and Low Recombination Affect Diversity near the Self-Incompatibility Loci of the Plant Arabidopsis lyrata. *Current Biology* 15, 1773–1778.

Kawabe A, Hansson B, Forrest A, Hagenblad J, Charlesworth D (2006) Comparative gene mapping in Arabidopsis lyrata chromosomes 6 and 7 and A. thaliana chromosome IV: evolutionary history, rearrangements and local recombination rates. *Genetical Research Cambridge* 88, 45-56.

Kowyama Y, Takashasi H, Muraoka K, *et al.* (1994) Number, frequency & dominance relationships of dominance relationships of S-alleles in diploid Ipomoea trifida. *Heredity* 73, 275-283.

Kwun M, Choi Y, Yoon H, *et al.* (2004) Expression analysis of the pistil genes in controlling self-incompatibility of Brassica campestris by CO2 gas using microarray. *Molecules and Cells* 18, 94-99.

Lane MD, Lawrence MJ (1995) The Population-Genetics of the Self-Incompatibility Polymorphism in Papaver Rhoeas .10. An Association between Incompatibility Genotype and Seed Dormancy. *Heredity* 75, 92-97.

Mable BK, Schierup MH, Charlesworth D (2003) Estimating the number, frequency, and dominance of S-alleles in a natural population of Arabidopsis lyrata (Brassicaceae) with sporophytic control of self-incompatibility. *Heredity* 90, 422-431.

Nakanishii T, Esashi Y, Hinata K (1969) Control of self-incompatibility by CO_2 gas in *Brassica. Plant and cell physiology* 10, 925-927.

Nakanishii T, Hinata K (1973) An effective time for CO_2 gas treatment in overcoming self-incompatibility in *Brassica. Plant and cell physiology* 14, 873-879.

Niikura S, Matsuura S (2000) Genetic analysis of the reaction level of self-incompatibility to a 4% CO2 gas treatment in the radish (Raphanus sativus L.). *Theoretical and Applied Genetics* 101, 1189-1193.

Porcher E, Lande R (2005) Loss of gametophytic self-incompatibility with evolution of inbreeding depression. *Evolution* 59, 46–60.

Prigoda NL, Nassuth A, Mable BK (2005) Phenotypic and genotypic expression of self-incompatibility haplotypes in Arabidopsis lyrata suggests unique origin of alleles in different dominance classes. *Molecular Biology and Evolution* 22, 1609-1620.

Schierup MH, Mable BK, Awadalla P, Charlesworth D (2001) Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of Arabidopsis lyrata. *Genetics* 158, 387-399.

Schierup MH, Vekemans X, Christiansen FB (1997) Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics* 147, 835-846.

Sokal RR, Rohlf FJ (1995) *Biometry: the principles and practice of statistics in biological research* 3rd edn. W. H. Freeman & Co, New York.

Stone JL (2004) Sheltered load associated with S-alleles in *Solanum carolinense*. *Heredity* 92, 335-342.

Uyenoyama MK (1997) Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics* 147, 1389-1400.

Uyenoyama MK (2003) Genealogy-dependent variation in viability among self-incompatibility genotypes. *Theoretical Population Biology* 63, 281-293.

Uyenoyama MK (2005) Evolution under tight linkage to mating type. *New Phytologist* 165, 63–70.

Vekemans X, Schierup MH, Christiansen FB (1998) Mate Availability and Fecundity Selection in Multi-Allelic Self- Incompatibility Systems in Plants. *Evolution* 52, 19-29.

Chapitre 4 - Evolution de la dominance au locus d'auto-incompatibilité

Dans les chapitres précédents, nous avons pu constater l'importance des relations de dominance sur l'évolution des allèles d'auto-incompatibilité. En définissant le phénotype d'auto-incompatibilité, les relations de dominance conditionnent notamment la disponibilité en partenaires, c'est-à-dire la capacité de reproduction des individus dans les populations. Nous nous sommes donc demandé dans quelle mesure ces relations peuvent évoluer. Par l'utilisation d'un modèle déterministe simple, nous avons étudié l'évolution de la dominance au locus S, et tenté d'expliquer les patrons de dominance observés en population naturelle :

- La dominance est elle sélectionnée par rapport à la co-dominance ?
- Est-ce que la dominance peut évoluer au sein d'une hiérarchie ?
- Y a-t-il une coévolution entre les niveaux de dominance dans le pollen et le pistil ?

Cette étude fait l'objet d'une publication : *Evolution of dominance at a sporophytic self-incompatibility locus*. Llaurens, V., Billiard, S., Vekemans, X., en préparation pour le journal *PNAS*.

Evolution of dominance at a sporophytic self-incompatibility locus

Llaurens V.*, Billiard S.*, Vekemans X.*

Corresponding author: Sylvain Billiard

Email : sylvain.billiard@univ-lille1.fr

* Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8016

Universite des Sciences et Technologies de Lille - Lille1

F-59655 Villeneuve d'Ascq cedex

France

Keywords: gene expression, balancing selection, S-locus, « hidding effect »

Running title: Evolution of dominance at the S-locus

Abstract

Evolution of dominance is a long-stand debate among evolutionary biology: few cases of evolution of dominance have been reported for locus under balancing selection. Since the sporophytic self-incompatibility locus is under strong balancing selection, we suggested here that it could represent a new case where dominance could evolve. In self-incompatible plants, reproduction is possible only between individuals expressing different self-incompatible specificities. These specificities are encoded by the S-locus. In sporophytic SI, specificities of pollen and pistil are determined by the diploid genome of the parent and dominance interactions among haplotypes at S-locus have been observed. These dominance relationships thus conditioned the possible mating in population. We investigated the possible evolution of dominance at the sporophytic self-incompatibility locus. We found that dominance was selected against co-dominance, thus favoring dominance ladder among S-haplotype. We then demonstrated that changes in dominance ladder are limited. And finally, we showed that the invasion of mutant S-haplotypes with opposite dominance level in pollen and pistil are counter selected under high level of inbreeding depression. Inbreeding depression could thus be an importance force for the co-evolution of dominance level in pollen and pistil.

Introduction

Evolution of dominance is a long-stand debate among evolutionary biologist (Fisher, (1928); Wright, (1929); Veitia, (2005)). Haldane (1930) suggested that the evolution of dominance because of natural selection should be more easily encountered in natural populations when heterozygotes are expected to be frequent. More recently, Otto & Bourguet (1999) showed that dominance could easily evolve for loci under balancing selection. However, only few specific cases have been studied so far: balancing selection due to heterogeneous environment like in evolution of resistance in mosquitoes (Bourguet *et al.*, 1997) or in hosts-parasites interactions (Nuismer & Otto, 2005a). Our aim here is to study an original example which has never been considered so far in the eye of the evolution of dominance: the sporophytic self-incompatibility system (SSI) in hermaphroditic plants.

SSI avoids selfing and reproduction with relatives, through a molecular recognition between pollen and stigmate: when pollen expresses the same molecular specificity than the receptor stigmate, fertilization is avoided. This specificity is generally encoded by a single locus: the S-locus. This system puts constraints on the mating opportunities within populations, that are determined by the genotypes at the self-incompatibility locus. Individuals carrying rare haplotypes at the S-locus (S-haplotypes) have access to a higher number of possible mates than individuals carrying common S-haplotypes, a process generating negative frequency dependent selection on the S-locus (Wright, 1939). In SSI, the mating phenotypes of both pollen and pistils are determined by the diploid parental genotypes at the S-locus. Complex dominance relationships among S-haplotypes usually occur in sporophytic SI systems (Bateman, 1952), which dramatically influence patterns of mating within populations (Vekemans *et al.*, 1998). The dominance level of a given S-haplotype was found indeed to determine its evolutionary properties such as its ability to invade, its propensity to be lost by drift, or its frequency at equilibrium ((Schierup *et al.*, 1997);

(Uyenoyama, 2000) and (Billiard *et al.*, 2007)). Dominance relationships between alleles at the S-locus are also implied in a species demography, especially in small populations viability (Kirchner *et al.*, 2006) or in an invasion front (Brennan *et al.*, 2006). In spite of the recognition of the large impact of dominance on evolutionary dynamics of sporophytic SI, the question of the mechanisms influencing patterns of dominance evolution, as well as their interactions, have not been formally studied.

The molecular characterization of sporophytic SI has only been achieved in a single family of flowering plants, the Brassicaceae (Takayama & Isogai, 2003). Phenotypes of pollen and pistils at the S-locus have been found to be determined by two separate, yet highly linked, genes, *SCR* and *SRK*, respectively. The proteins produced by these two genes interact as a lock-key system, allowing recognition and active rejection of self-pollen at the stigma surface. These two genes have independent origins and recent studies suggested that their mechanisms of determination of S-haplotype dominance are not homologous ((Shiba *et al.*, 2006); (Naithani *et al.*, 2007). Experimental investigations on patterns of dominance among S-haplotypes within the Brassicaceae family showed contrasted results. In most studies, complex dominance relationships among S-haplotypes were observed, with a structure marked by discrete dominance classes, where co-dominance usually occurs within classes and dominance between classes. However, in *Arabidopsis lyrata*, and *A. halleri*, the results seemed to indicate an overall correlation in patterns of dominance between pollen and pistil (Mable *et al.*, 2003); (Prigoda *et al.*, 2005); (Schierup *et al.*, 2006); Llaurens *et al.*, *in prep.*), whereas in Brassica and Raphanus, contrasted dominance relationships between pairs of alleles in pollen and pistil were reported ((Hatakeyama *et al.*, 1998); (Stevens & Kay, 1989)). Another common observation is that co-dominant relationships among S-haplotypes occur more often in the pistil than in the pollen (Llaurens *et al.*, *in prep.*; (Hatakeyama *et al.*, 1998)).

Can we explain those complex dominance relationships patterns by the action of selection? Two main types of selection acts on the S-locus and hence could potentially act on dominance as well. First negative frequency dependent selection or advantage of the rare: an allele is favored when the expression of its specificity is rare in the population, either on pollen or pistil, during reproduction. Second inbreeding depression which is invoked as the main factor explaining the maintenance and evolution of self-incompatibility systems (Porcher & Lande, 2005). This phenomenon is assumed to be widespread in species with SI (Mable *et al.*, 2005), and could be due either to background inbreeding depression because of mildly deleterious mutations distributed across the genome, and/or to a sheltered genetic load linked to S-haplotypes (Uyenoyama, 1997), (Glémin *et al.*, 2001). A sheltered load is indeed thought to be associated to some S-alleles because recessive deleterious mutations linked to S-alleles could not be purged (Uyenoyama, 1997): dominant S-alleles are rarely at homozygous state and thus avoid the expression of the associated deleterious mutations. As recombination is usually very low in the S-locus genomic region (Charlesworth, 2006), a sheltered genetic load could thus be conserved associated to the S-locus.

In this paper, we investigated the evolution of dominance in sporophytic SI systems. Based on a general model of evolution of allele frequencies at a sporophytic SI locus (Billiard *et al.*, 2007), we built a deterministic model where the dominance level of each S-haplotype could evolve without changing the S-specificity encoded. We also took into account the existence of inbreeding depression in cases of dominance relationships that allow self crosses, mostly when the expression of an S-allele is asymmetric between pistil and pollen. Finally, we also looked at the effect of an hypothetical sheltered load. We aimed at answering the following questions (1) Is it possible to maintain dominance classes, in other words codominance, when there is neither inbreeding depression nor sheltered load? (2) If the dominance pattern follows

a ladder, is it fixed or could the S-alleles move across it? (3) Under which conditions should we expect no correlation between dominance patterns in pistil and in pollen?

Materials and methods

The basic model

The model is derived from Billiard *et al.* (2007). We assumed an infinite population of hermaphroditic diploid individuals with a fully fonctional SSI system. We assumed first that n different alleles S_i segregate at the S-locus, and that each allele encodes for a different specificity when it is expressed. A given individual is characterized by its genotype S_iS_j and its phenotype A_{ij} in anthers and P_{ij} in pistil. The phenotype is defined as a vector of size n containing the relative quantity of specificity i and j expressed by a genotype S_iS_j: if S_i and S_j are codominant in the anthers, the individual produces one half of specificity i, one half of specificity j and no other specificity in the pollen. In this case, the phenotype of genotype S_iS_j in anthers is therefore $A_{ij}=\{0 \ldots 0.5 \ldots 0 \ldots 0.5 \ldots 0\}$, with the value 0.5 at position i and j only. If allele S_i is dominant over S_j in the pistil, the individual only expresses specificity i and therefore its phenotype is $P_{ij}=\{0 \ldots 1 \ldots 0\}$, with 1 at position i and 0 elsewhere. This notation allows to determine easily if a cross between pollen from an individual S_iS_j and the pistil of an individual S_kS_l is compatible: if $A_{ij}.P_{ij}^T = 0$ and incompatible otherwise.

We computed the frequency change of the genoype S_iS_j in a generation given the frequency of all genotypes in the population f_{kl} as

$$\begin{cases} f'_{ij} = \sum_{k=1}^{n} \frac{1}{2}\left(w_{iijk} + w_{jkii}\right) + \sum_{k=1}^{n} \frac{1}{2}\left(w_{ikjj} + w_{jjik}\right) + \sum_{\substack{k=1 \\ k\neq i}}^{n} \sum_{\substack{l=1 \\ l\neq j}}^{n} \frac{1}{4}\left(w_{ikjl} + w_{jlik}\right) & \text{for } i \neq j \\ f'_{ii} = \sum_{j=1}^{n} \frac{1}{2}\left(w_{iiij} + w_{ijii}\right) + \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{1}{4} w_{ijij} + \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{\substack{k=1 \\ k\neq i\neq j}}^{n} \frac{1}{4}\left(w_{ijik} + w_{ikij}\right) \end{cases}$$

with w_{ijkl} the fraction of seeds produced by a cross between S_iS_j pollen and S_kS_l ovules among all seeds. The w_{ijkl} depend on genotypic frequencies, compatibility between S_iS_j pollen and

S_kS_l pistil and on the selection regime as well: Wright's model if selection only acts through pollen or fecundity selection model is it acts through both pollen and pistil (see Billiard et al. 2007 equations 2 to 4 for details). The latter case could be realistic in some particular ecological situations, like in pollen limitation.

Introduction of the mutant allele

We aim to compute the fate of a mutant allele entering the population which does not change the specificity it encodes for but which has a different dominance relationships than the allele it is derived, and so with at least one other alleles, in pollen or pistil or both. Hence, once the mutant is entered in the population, there is n+1 alleles but still only n specificities are potentially expressed by the individuals of the population. In other words, phenotype vectors have still only n elements but the sums in equation (1) now goes until n+1 instead of n. For instance if the mutant allele m is derived from allele i such that it is dominant over all other alleles instead of being codominant in the pollen, its phenotype in anthers is A_{mj} = {0...1..0} with 1 at position i while the phenotype of its ancestor is A_{ij} = {0 ... 0.5 ...0 ... 0.5 ...0} with 0.5 at position i and j.

Fate of the mutant

To determine the deterministic fate of the mutant, we first determined the genotype frequencies at deterministic equilibrium of the population without the mutant using recursions (1). We then introduced the mutant at low frequency ε (ε=0.005 for if n<=10, ε=0.0025 if n=15) in a population supposed at deterministic equilibrium. .We finaly used recursions (1) taking into account the new allele until the frequency change in one generation of the mutant is lower than 10^{-3}.

Patterns of dominance

We considered two classical patterns of dominance relationships; the COD and the DOM systems which are commonly used in the study of sporophytic self-incompatibility (Schierup *et al.*, 1997). In the COD system, all S-haplotypes are fully co-dominant in both pollen and pistil; we assumed here a COD system with four S-haplotypes which is the minimum number of S-haplotypes for self-incompatibility to function in a COD system. Mutant S-haplotype in this case could evolve to be recessive or dominant with respect to all the other S-ahplotype. In the DOM model, dominance relationships among S-haplotypes are linear; we assumed here the simplest case with three S-haplotypes: S-haplotype S_3 dominant one S-haplotype S_2 itself dominant over S-haplotype S_1 in both the pollen and the pistil. These S-haplotypes were ordered in a dominance ladder with seven steps, numbered from I the most recessive steps to VII, the most dominant steps (see initial state, Fig. 3) by considering dominance among steps and co-dominance within steps. The mutant S-haplotype could move to every possible steps, the four possible mutations of each three ancestral S-haplotypes (*i.e.* twelve differents cases) were studied. Finally, we considered three different dominance evolution: in pollen only, in pistil only or in both simultaneously. We will denote S_i^+ and S_i^- mutant derived from S_i which are respectively more dominant or more recessive than the ancestor.

Sheltered genetic load linked to the S-locus

We also introduced a sheltered genetic load due to a recessive deleterious allele at a locus fully linked to the S-locus. This sheltered load was only expressed in homozygote genotypes at the linked locus. In the particular case of the mutant-haplotype, we considered that it shares the same deleterious alleles as its ancestor and so that the sheltered load is expressed in heterozygotes bearing the mutant and the ancestor alleles. We assumed that the sheltered laod acts after reproduction by decreasing the survival of seeds homozygotes at the linked locus by

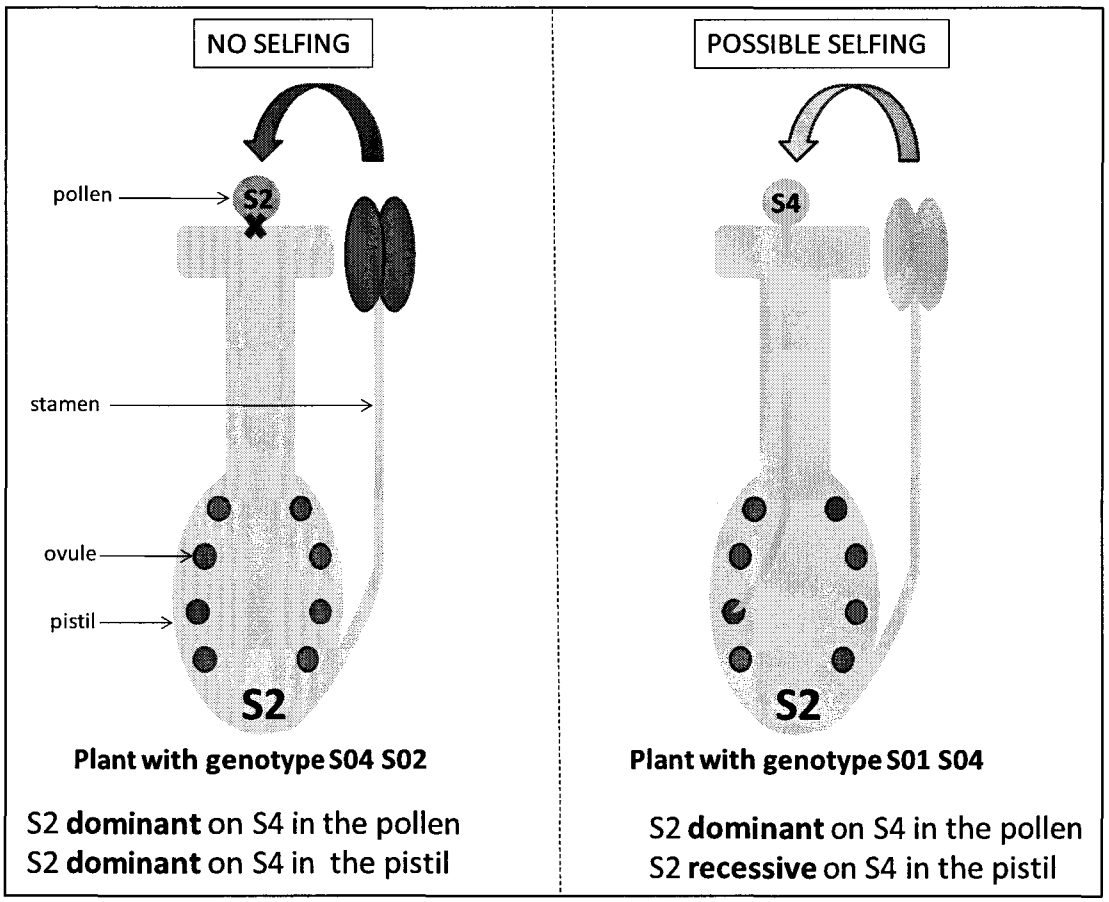


Figure 1: Evolution of selfing through discordant dominance relationships in the pollen and the pistil.

a factor (1 - d) with $0 \leq d \leq 1$, We simulated the evolution of dominance in the absence of genetic load ($d = 0$), when the genetic load was equal for all S-alleles (d=1), or correlated to the dominance level of the allele: high in the dominant allele (d_3); intermediate for the allele with medium dominance level (d_2); and low for the recessive allele (d_1). The latter hypothesis were justified because it is expected that sheltered load is lower for recessive than dominant alleles since there are more homozygotes for recessive alleles which should allow more purging of linked deleterious alleles.

Background inbreeding depression

Under some combinations of dominance relationships in pollen and pistil, the heterozygous genotypes may become self-compatible, when dominance for a given allele is asymetrical between pollen and pistil (see Fig. 1). We assumed that background inbreeding depression could act in this case: self-compatible genotypes were enforced to self cross at rate s with a cost of selfing δ. Practically, we assumed that among all seeds produced by a self-compatible genotype, a fixed fraction s were from selfcrossing and 1-s from outcrossing. We further assumed that the fraction of selfed seeds had a survival probability lowered by a factor (1- δ).

Results

Co-dominance is counterselected

Fig. 2a shows the evolution of S-haplotype frequencies towards equilibrium in a COD system, with n=4, under *Wright*'s model. We introduced a mutant S-haplotype S_3^+ that is dominant over all other S-haplotypes in both pollen and pistil. At initial state, all four S-haplotypes had identical frequencies (0.25), as expected in a COD model of sporophytic SI. The dominant mutant S-haplotype (S_3^+) reached rapidly high frequency, whereas its parental S-haplotype (S_3) decreased in frequency. The result was similar when introducing a recessive S-haplotype

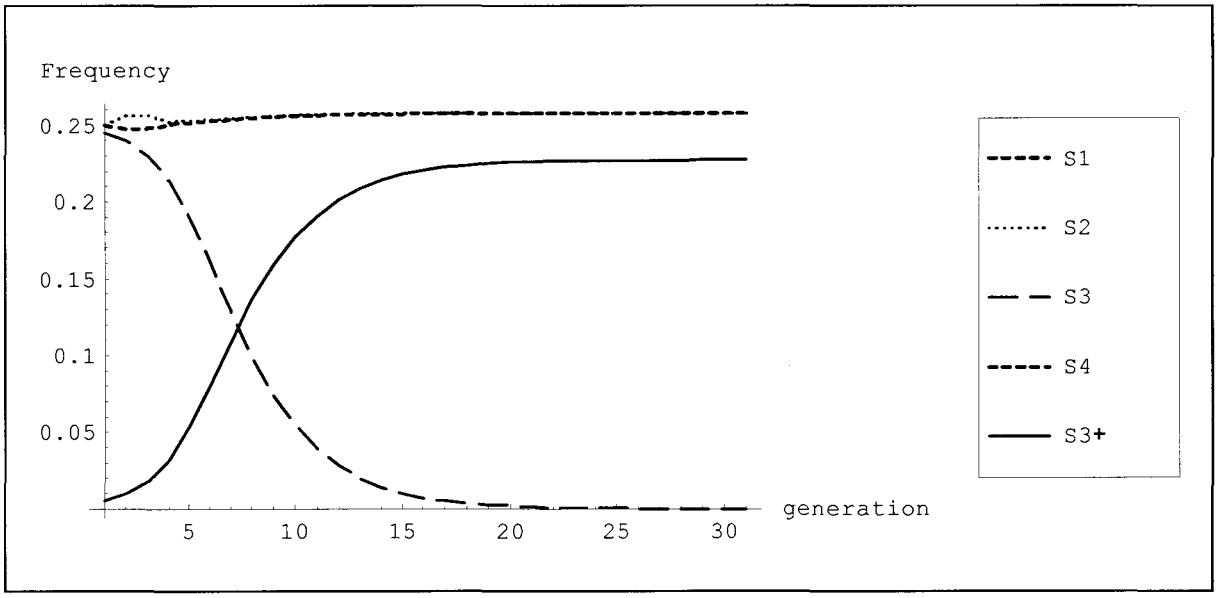


Figure 2a: S-alleles frequencies evolution in a COD system under *Wright* frequency dependent model with a dominant mutant S3+ appearing simultaneously in the pollen and in the pistil

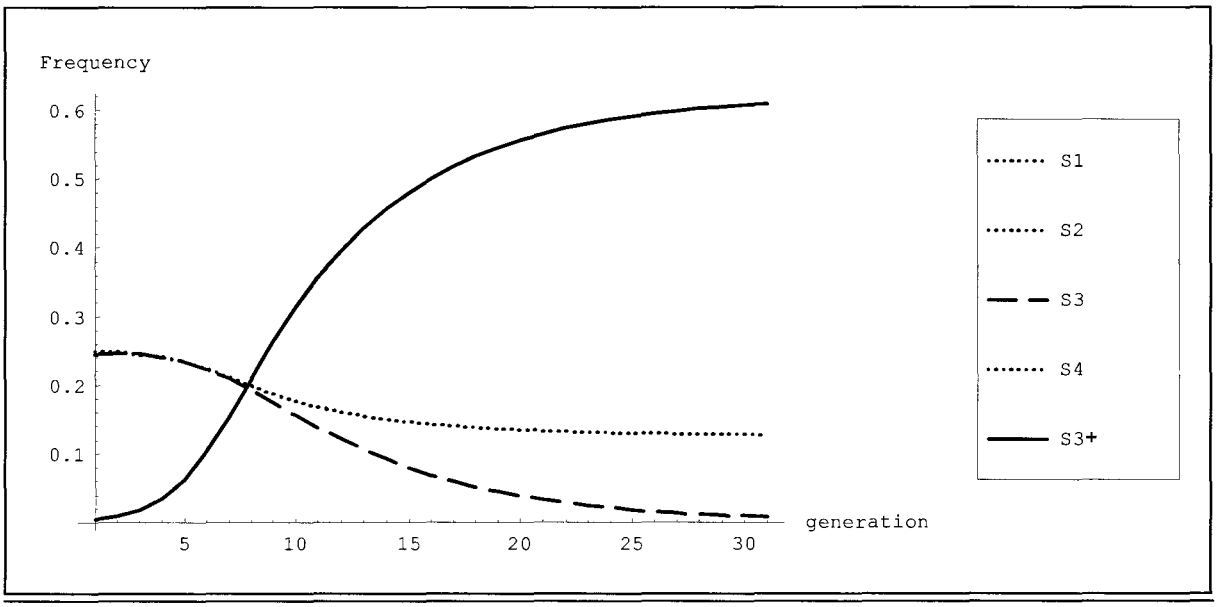


Figure 2b: S-alleles frequencies evolution in a COD system under *Wright* frequency dependent model with a recessive mutant $S3^{-}$ appearing simultaneously in the pollen and in the pistil

mutant (Fig. 2b). However, the frequency of the mutant at equilibrium depended on its dominance level: the recessive mutant S-haplotype S_3^- reached a higher frequency than the dominant S_3^+. This is a common feature of S-haplotypes in sporophytic self-incompatibility when dominance occurs, often referred to as the "*recessive effect*", *i. e.* that recessive S-haplotypes reach higher frequencies at equilibrium than dominant ones (Schierup *et al.*, 1997). These results showed that both recessive and dominant mutant S-haplotypes (S_3^+ or S_3^-) were able to invade the population, thereby eliminating the parental S-haplotype (S_3) with identical specificity. They suggested that the evolution of dominance at the S-locus could be under positive selection. Moreover, as the equilibrium was reached extremely rapidly (the one-generation frequency change after 20 generations was under 10^{-3}), strong positive selection for dominance evolution can be expected .

In the simulations above, *Wright's* model was assumed, and the mutant S-haplotype was dominant in both pollen and pistil. A strikingly different outcome was obtained when the mutant S-haplotype was dominant in the pistil only: the parental S-haplotype (S_3) was maintained and the mutant S-haplotype frequency did not change. This is due to the fact that in *Wright*'s model, the frequency dependent-selection only acted on male function, *i. e.* on the pollen phenotype, whereas all pistils, including those carrying the dominant mutant-haplotype (S_3^+), are assumed to have equal fitness. The dominant mutant S-haplotype (S_3^+) thus behaved as neutral with respect to the ancestral S-haplotype (S_3). Surprisingly, a different result was obtained when a recessive mutant haplotype (S_3^-) expressed in the pistil only was introduced: this mutant haplotype was positively selected even when it only appeared in the pistil. This result could be explained by the indirect effect of homozygote genotypes with two copies of the mutant-haplotype S_3^-. As the mutant haplotype S_3^- was recessive in the pistil, it allowed homozygote genotypes with two copies of the mutant-haplotype S_3^- to be formed, whereas

homozygotes for the ancestral S-haplotypes S_3 were never obtained. This difference produced a small indirect selective advantage to the mutant S-haplotype S_3^-.

Under the *fecundity selection* model, mutants S_3^+ and S_3^-, in pollen and/or pistil always invade. This is because this model of selection is fully symmetrical in determination of male and female fitness. The invasion of the mutant is slightly faster when it is expressed both in pollen and pistil was slightly lower. This suggested that selection for dominance was stronger when acting on male and female fitness simultaneously. The introduction of a big sheltered genetic load in the COD system (d=1) did not qualitatively modify the previous results: co-dominance is also counterselected. In the case of a dominant mutant S-haplotype S_3^+, its equilibrium frequency was higher than the frequencies of the co-dominant S-haplotypes, while it is expected to have a lower frequency without sheltered load (Fig. 2a). This is because, when the mutant enters the population, homozygotes are newly produced for all alleles except the mutant, and therefore sheltered load is expressed. In the case of a recessive mutant S_3^-, the mutant S-haplotype invaded the population but the ancestor S_3 was also maintained. At equilibrium, the mutant-haplotype reached the highest frequency, whereas the parental S-haplotype frequency was lower or equal than that of the other co-dominant S-haplotypes.

We also tested if a mutant becoming codominant could invade in a DOM system with n=3, for both selection models. We found that the mutant haplotype could never invade, whatever was the dominance level of its ancestor: recessive, intermediate, or dominant. We can thus infer that (1) Dominance of S-haplotype evolution could be selected and (2) Dominance is selected against co-dominance. We could then hypothesize that the selection of dominance should lead to hierarchical dominance relationships. We could then ask if the movements of dominance level of S-haplotypes in the dominance ladder were possible.

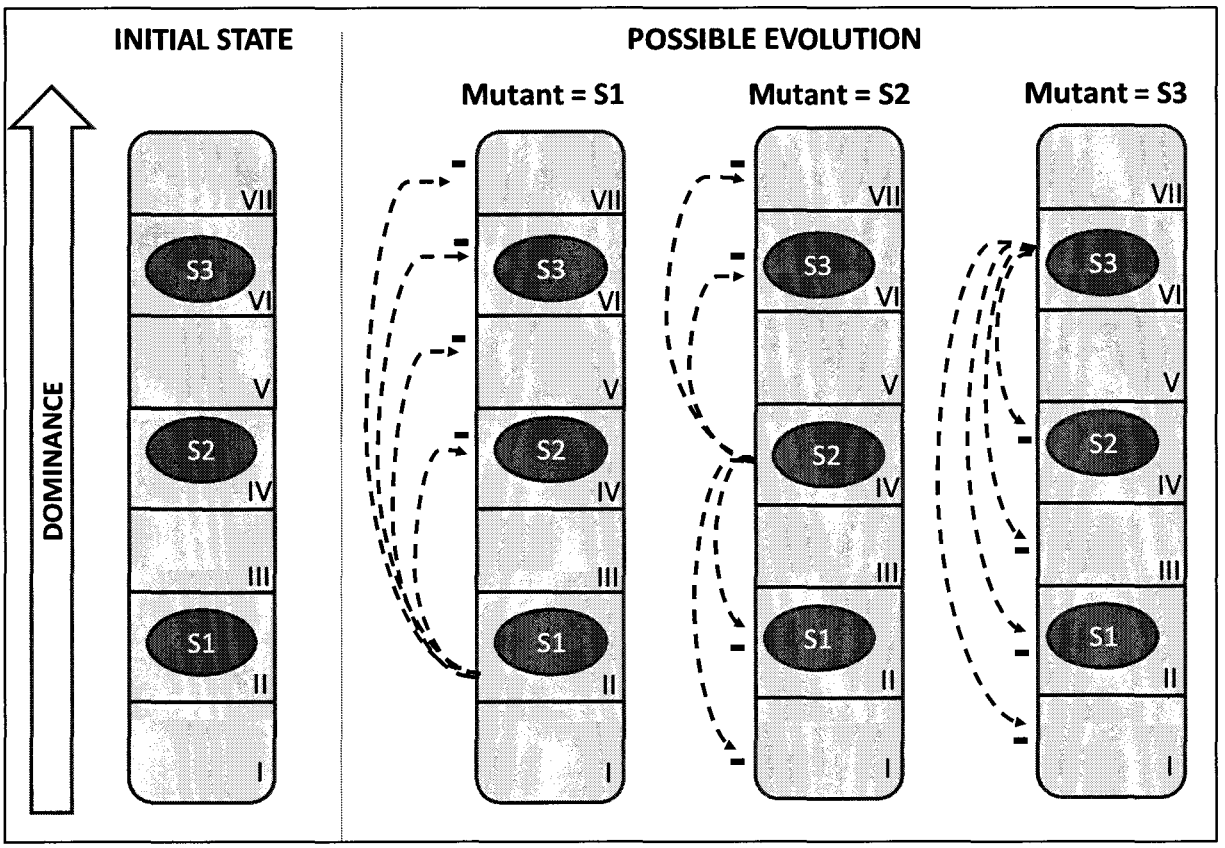


Figure 3a: Possible evolution in the dominance ladder in a DOM system, with no sheltered genetic load (d=0) with mutant S-haplotype appearing simultaneously in pollen and pistil under *Wright* and *Fecundity selection* model

Continuous arrows: Mutant S-haplotype is fixed in the population

Dotted arrows: Mutant S-haplotype is neutral in the population

Movements across a dominance ladder

We looked under which condition a DOM system is fixed, in other words if a given S-allele could move across a dominance ladder. We checked the invasion of mutants climbing up or down the dominance ladder, in pollen and/or pistil, under both selection regimes (Fig.3a-c). When dominance changes either in pollen or in pistil and not both, some genotypes can become self-compatible (Fig. 1). In this part, we only look at the fate of the mutant with no selfing (s=0) and no inbreeding depression (δ=0). The effect of inbreeding depression is investigated next.

Under both frequency-dependent selection regime (*Wright* or *fecundity selection*), when DOM system is assumed as the initial state, no mutants with simultaneous changes of dominance in pollen and pistils under fecundity selection invaded (Fig 3a). In contrast, mutant with dominance changes in either pollen or pistil could invade under the *fecundity selection* and in pollen only under *Wright* model (Fig. 3b). The results showed that invasion of the mutant haplotype occurred when its dominance level created a new dominance class, anywhere in the dominance ladder, *i.e.* when no co-dominance among pairs of haplotypes was introduced (Fig. 3b). Hence, changes in dominance level of S-haplotypes only occurred when (1) a strictly linear pattern of dominance was maintained; and (2) the new dominance level was expressed only in pollen or in pistil. These conditions will inevitably cause contrasted patterns of dominance in pollen and pistil of the successful mutant haplotype, leading to self-compatibility of the resulting genotypes (Fig. 1). The dynamics of invasion of these mutants was very slow when compared to cases of invasion in the COD system, suggesting that in finite populations evolutionary changes across the dominance ladder would be uncommon.

Under both frequency-dependent selection regime, introduction of a sheltered genetic load linked to the S-haplotypes had a strong impact on the fate of mutant haplotypes (Fig. 3c with d_1=0.2, d_2=0.6, d_3=0.8). Indeed, invasion of a mutant S-haplotype only occurred when its

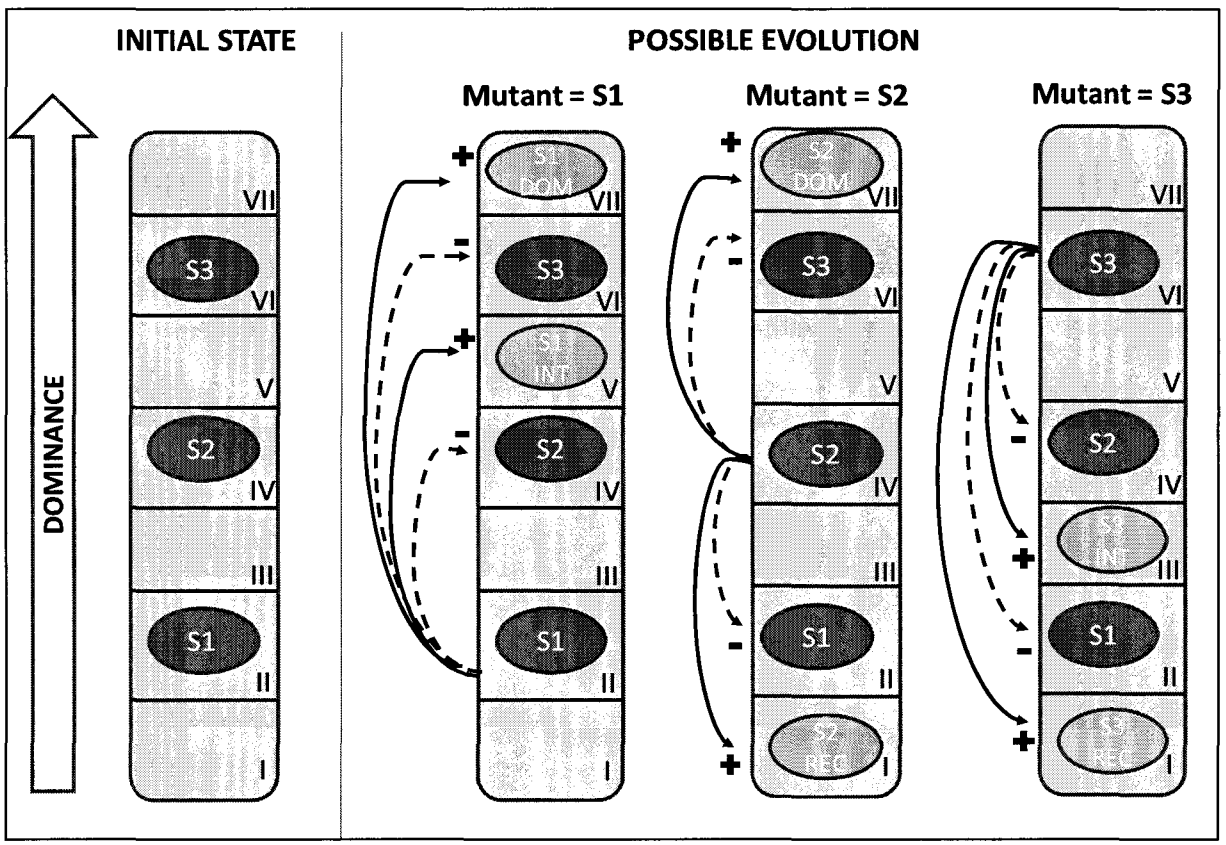


Figure 3b : Possible evolution in the dominance ladder in a DOM system, with no sheltered genetic load (d=0) with mutant S-haplotype appearing in pollen only or in pistil only under *fecundity selection* model or in pollen only under *Wright* model.
Continuous arrows: Mutant S-haplotype is fixed in the population
Dotted arrows: Mutant S-haplotype is neutral in the population

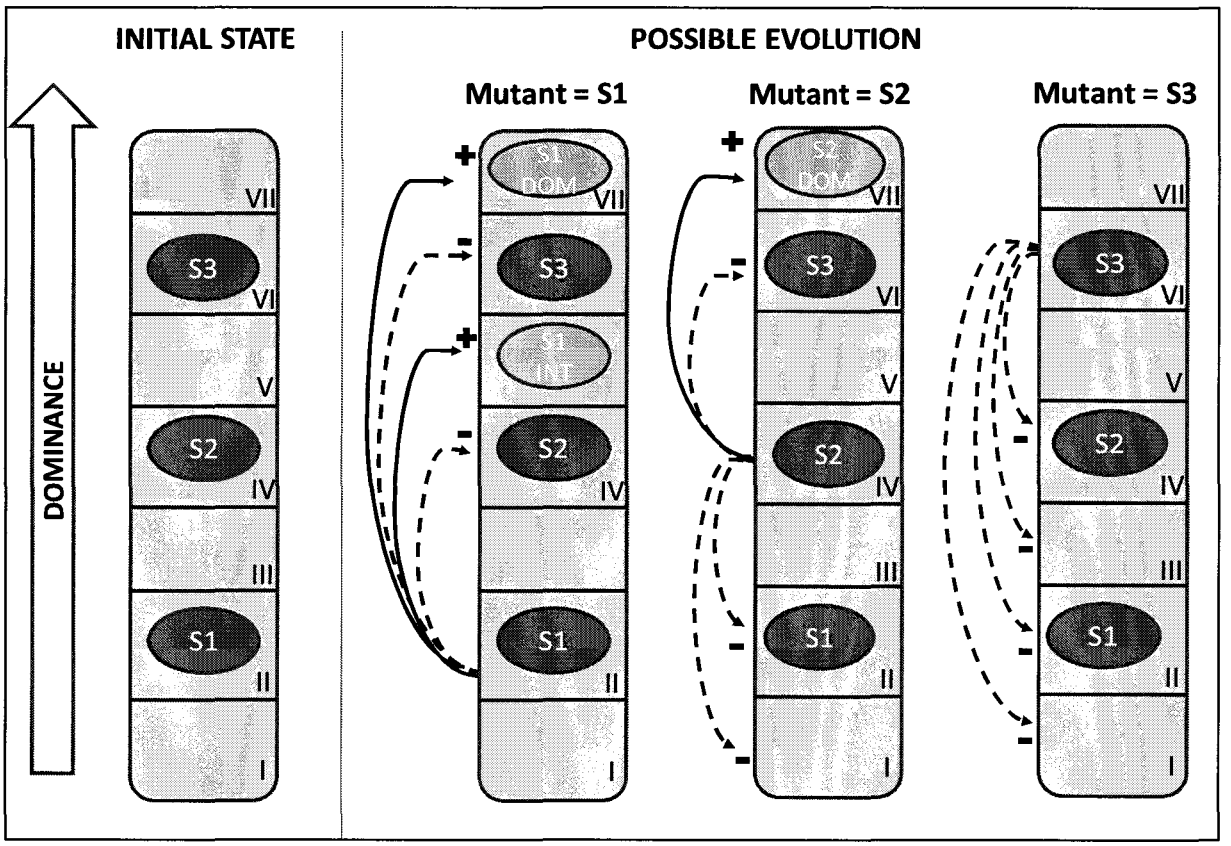


Figure 3c : Possible evolution in the dominance ladder in a DOM system, with equal sheltered genetic load (d=1) with mutant S-haplotype appearing simultaneously in pollen and in pistil under *Wright* or *fecundity selection* model.
Continuous arrows: Mutant S-haplotype is fixed in the population
Dotted arrows: Mutant S-haplotype is neutral in the population

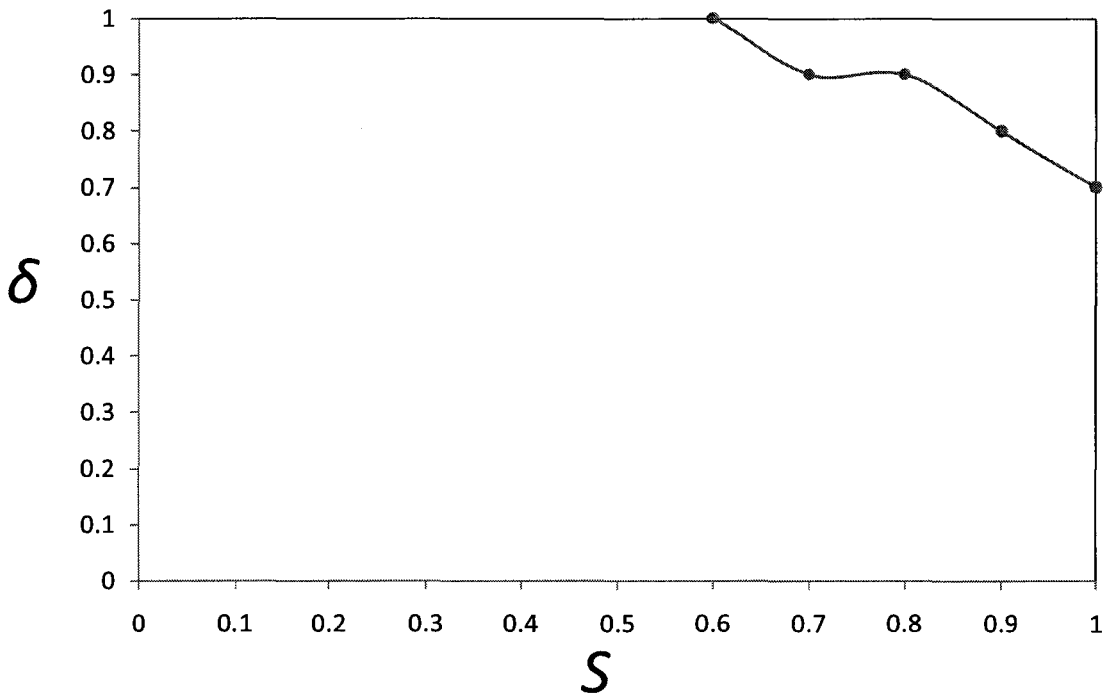


Figure 4: Conditions of invasion of a mutant S-haplotype recessive on all other S-haplotypes in the pollen and dominant in the pistil in a COD system with n=10 S-haplotypes under *fecundity selection* model. Invasion of the mutant S-haplotype is successful in the area under the line.

y-axis: background inbreeding depression (δ); x-axis: selfing rate in the self-compatible genotypes (s).

dominance level was increased as compared to dominance of its parental haplotype. In this case invasion occurred even when the new dominance was expressed in pollen and pistil simultaneously. The selective advantage of higher dominance levels is due to the fact that the most dominant S-haplotype cannot form homozygous genotypes, and thus will never be exposed to the sheltered load.

Coevolution of dominance level in pollen and pistil component genes

We finally investigated the invasion of a mutant in an initial COD system with asymmetric dominance level in pollen and pistil, *i.e.* dominant in pollen and recessive in pistil or *vice-versa*. As shown before, when there is no inbreeding depression, this mutant invades because codominance is counterselected and, in this particular case, because genotypes bearing the mutant alleles are able to mate with their own genotypes and especially with themselves. We therefore looked at the potential effect of the inbreeding expression for different values of s and δ.

Figure 4 shows values of s and δ where the mutant can invade, for number of alleles n =10 in the population. Figure 4 shows that there is a region where a mutant allele with asymmetrical dominance levels in pollen and pistil can invade. This region largely depends on the number of alleles considered, and its size increases when n increases (data not shown). These results show that asymmetrical dominance relationships in pollen and pistil are unlikely when the number of alleles in the population and inbreeding depression are relatively high.

Discussion

The advantage of dominance over co-dominance: the hiding effect

Our results showed that any mutation introducing dominance of an S-haplotype in a COD sporophytic self-incompatibility system was favored. Indeed, in a full co-dominant system, all

S-haplotypes occur in heterozygote genotypes, and each heterozygote can only mate with individuals carrying two different S-haplotypes from itself. Introduction of dominance allows heterozygotes to *hide* one of their S-haplotypes, and thus to have access to a higher proportion of compatible mates than genotypes expressing two co-dominant S-haplotypes. We refer to this phenomenon as the "*hiding effect*". Unintuitively, the *hiding effect* favors invasion of either dominant or recessive mutant S-haplotypes. Nuismer & Otto (2005b) obtained analogous results in hosts-parasites interactions and showed that hierarchical patterns of dominance were favored in the infection genes of the parasites, allowing them to minimize probability of recognition by the host.

We also showed that co-dominance cannot evolve in a population with a linear dominance ladder, whereas some changes in levels of dominance can occur, suggesting that the *hiding effect* constitutes an inherent selective pressure in favor of hierarchical patterns of dominance within species of sporophytic SI. Introducing a sheltered genetic load linked to the S-locus brings about an additional selective advantage for a dominant mutant S-haplotype, but a selective disadvantage for a recessive mutant S-haplotype. However, the qualitative results are not changed when a sheltered load is assumed. Altogether, these results show that co-dominance is expected to be counter-selected, in other words if the rate of appearance of dominance mutants is not too low in natural populations and if there is no other selectivepressures or structural constrtaints on dominance evolution, then we should not observe dominance classes but only dominance ladders.

The upward movement of dominance level: the effect of sheltered load

Our results showed movement up or down in the dominance ladder is possible only for fecundity selection when mutation does not occur in both pollen and pistil, which is mainly due to the fact that genotypes bearing the mutant can then mate with individuals carrying the

same genotypes. It should be noticed here that selfing can be obtained even with a fully functionnal SSI system.

Moreover, the sheltered load causes unidirectional evolution of dominance, favoring mutant S-haplotypes moving upwards in the dominance ladder. Dominance is advantageous in this case because homozygote genotypes are not formed, avoiding the expression of the sheltered load. This trend should be particularly true when assuming that the strength of the sheltered load increases with the dominance level of the S-haplotype (Llaurens *et al.*, in prep.). The system may become locked when the sheltered load is very high in the dominant S-haplotype and low in the recessive one, which suggests that we should observe only few changes in a dominance ladder, for instance between closed species.

Coevolution in pollen and pistil: a negative role of inbreeding depression

The invasion of mutant S-haplotype with opposite dominance relationships in the pollen and in the pistil was found likely in the COD system even in the presence a sheltered genetic load. The hypothesis of the negative selection of the sheltered genetic load was not able to explain maintenance of similar dominance level in the pollen and the pistil. However, the inbreeding depression acting on offspring of self-compatible individuals counter selected the appearance of discordant dominance relationships in pollen and pistil. We noticed that a high level of selfing rate and inbreeding depression is required to avoid the appearance of the discordant mutant, *i. e* of self-compatibility when the number of alleles in the population is low. However the conditions of invasion are less wide when the number of S-haplotype in the population increases. As already highlighted by Porcher & Lande (2005) in GSI, the invasion of self-compatible mutant is likely to occur in a self-incompatible population and is counteracted only for high value of background inbreeding depression and a high number of alleles. It suggested that particular inbreeding depression level could be required to explain

the co-evolution of dominance in pollen gene and the pistil gene, and more generally for the maintenance of the self-incompatibility system.

Evolution of dominance through modifiers

We showed that the dominance level of an allele could evolve independenty of its specificity mainly because of balancing selection resulting from negative frequency-dependent selection inherent to this mating system. Some recent studies brought insights about the molecular mechanism of dominance at the S-locus. Concerning the pollen gene (*SCR* or *Sp11*), some models of dominance mechanism have been by proposed. Shiba *et al.* (2002) showed that in Brassica the dominant/recessive relationship *S*-haplotypes in the determination of pollen phenotype is generated at the level of *SCR* transcription. Fujimoto *et al.* (2006) found that the gene expression of a recessive *SCR* allele could be suppressed by an untranscribed *SCR* allele. The dominant interaction seemed thus independent from the protein SCR itself, but controlled by another factor. Indeed, Shiba *et al.* (2006) showed that 5' promoter sequences of recessive *SCR* allele were specifically methylated in the tapetum before the initiation of *SCR* transcription. It suggested that the tissu.10-specific DNA methylation was involved in determining the dominance interaction among *SCR* alleles.

Hatakeyama *et al.* (2001) suggested that the dominant/recessive relationships between *S*-haplotypes in the stigma are determined by SRK itself, but not as a result of its relative transcription level. Naithani *et al.* (2007) highlighted that in absence of ligand, the receptor kinase of the stigma (SRK) forms dimers which might provide a "primed" condition that allows rapid recruitment and activation of the receptor on ligand binding. They hypothesized that the dominance interactions may be based on the ability of different pairs of SRKs to form heterodimers. It suggested that the dominance ability could depend on the own sequence of the SRK gene or in other word that for the SRK gene dominance seemed to be intrinsically

encoded. The dominance mechanism in pollen and pistil gene seemed thus to be different, however dominance in pollen and pistil were observed consistent, suggesting a co-evolution process due to natural selection instead of biochemical constraints.

Conclusions

This model allowed us to infer general pattern of dominance evolution in a sporophytic self-incompatibility system. Our results demonstrate that the dominance was positively selected against the co-dominance. We also showed a positive selection for hierarchical dominance relationships and a selection for upward dominance level evolution by the sheltered genetic load linked to the S-locus. The co-evolution of dominance levels in the pollen and in the pistil seemed under the control the inbreeding depression, required high cost of inbreeding to be conserved. Finally, we pointed out that the sporophytic self-incompatibility locus is a nice example of locus under balancing selection where were dominance evolution could occur under a natural selection process.

References

Bateman AJ (1952) Self-incompatibility systems in angiosperms. 1. Theory. *Heredity* **6**, 285-310.

Billiard S, Castric V, Vekemans X (2007) A General Model to Explore Complex Dominance Patterns in Plant Sporophytic Self-Incompatibility Systems. *Genetics* **175**, 1351–1369.

Bourguet D, Lenormand T, Guillemaud T, *et al.* (1997) Variation of dominance of newly arisen adaptive genes. *Genetics* **147**, 1225-1234.

Brennan AC, Harris SA, Hiscock SJ (2006) The population genetics of sporophytic self-incompatibility in Senecio squalidus L. (Asteraceae): the number, frequency, and dominance interactions of S alleles across its british range. *Evolution* **60**, 213–224.

Charlesworth D (2006) Balancing Selection and Its Effects on Sequences in Nearby Genome Regions. *Plos genetics* **2**, 375-384.

Fisher RA (1928) The possible modification of the response of the wild type to recurrent mutations. *American Naturalist* **62**, 115-126.

Fujimoto R, Sugimura T, Fukai E, Nishio T (2006) Suppression of gene expression of a recessive *Sp11/SCR* allele by an untranscribed *SP11/SCR* allele in *Brassica* self-incompatibility. *Plant Molecular Biology* **61**, 577-587.

Glémin S, Bataillon T, Ronfort J, Mignot A, Olivieri I (2001) Inbreeding depression in small populations of self-incompatible plants. *Genetics* **159**, 1217-1229.

Haldane JBS (1930) A note on Fisher's theory of the origin of dominance, and on a correlation between dominance and linkage. *American Naturalist* **64**, 87-90.

Hatakeyama K, Takasaki T, Suzuki G, *et al.* (2001) The S receptor kinase gene determines dominance relationships in stigma expression of selfincompatibility in Brassica. *Plant Journal* **26**, 69–76.

Hatakeyama K, Watanabe M, Takasaki T, Ojima K, Hinata K (1998) Dominance relationships between S-alleles in self-incompatible Brassica campestris L. *Heredity* **80**, 241-247.

Kirchner F, Robert A, Colas B (2006) Modelling the dynamics of introduced populations in the narrow-endemic Centaurea corymbosa: a demo-genetic integration. *Journal of Applied Ecology* **43**, 1011-1021.

Mable BK, Robertson AV, Dart S, Di Berardo C, Witham L (2005) Breakdown of self-incompatibility in the perennial Arabidopsis lyrata (Brassicaceae) and its genetic consequences. *Evolution* **59**, 1437-1448.

Mable BK, Schierup MH, Charlesworth D (2003) Estimating the number, frequency, and dominance of S-alleles in a natural population of Arabidopsis lyrata (Brassicaceae) with sporophytic control of self-incompatibility. *Heredity* **90**, 422-431.

Naithani S, Chookajorn T, Ripoll DR, Nasrallah JB (2007) Structural modules for receptor dimerization in the S-locus receptor kinase extracellular domain. *Proceedings of The National Academy of Sciences* **104**, 12211–12216.

Nuismer SL, Otto SP (2005a) Host-parasite interactions and the evolution of gene expression. *Plos Biology* **3**, 1283-1288.

Nuismer SL, Otto SP (2005b) Host–Parasite Interactions and the Evolution of Gene Expression. *Plos Biology* **Volume 3**, 1283-1288.

Otto SP, Bourguet D (1999) Balanced polymorphisms and the evolution of dominance. *American Naturalist* **153**, 561-574.

Porcher E, Lande R (2005) Loss of gametophytic self-incompatibility with evolution of inbreeding depression. *Evolution* **59**, 46–60.

Prigoda NL, Nassuth A, Mable BK (2005) Phenotypic and genotypic expression of self-incompatibility haplotypes in Arabidopsis lyrata suggests unique origin of alleles in different dominance classes. *Molecular Biology and Evolution* **22**, 1609-1620.

Schierup MH, Bechsgaard JS, Nielsen LH, Christiansen FB (2006) Selection at Work in Self-Incompatible Arabidopsis lyrata: Mating Patterns in a Natural Population. *Genetics* **172**, 477–484.

Schierup MH, Vekemans X, Christiansen FB (1997) Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics* **147**, 835-846.

Shiba H, Iwano M, Entani T, *et al.* (2002) The dominance of allelle controlling self-incompatibility in Brassica pollen is regulated at the RNA level. *The plant cell* **14**, 491-504.

Shiba H, Kakizaki T, Iwano M, *et al.* (2006) Dominance relationships between self-incompatibility alleles controlled by DNA methylation. *Nature genetics*.

Stevens JP, Kay QON (1989) The Number, Dominance Relationships and Frequencies of Self-Incompatibility Alleles in a Natural-Population of Sinapis-Arvensis L in South-Wales. *Heredity* **62**, 199-205.

Takayama S, Isogai A (2003) Molecular mechanism of-self-recognition in Brassica self-incompatibility. *Journal of Experimental Botany* **54**, 149-156.

Uyenoyama MK (1997) Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics* **147**, 1389-1400.

Uyenoyama MK (2000) Evolutionary dynamics of self-incompatibility alleles in Brassica. *Genetics* **156**, 351-359.

Veitia R (2005) Phenomenology and mechanistics of dominance. In: *The biology of genetic dominance* (ed. bioscience L).

Vekemans X, Schierup MH, Christiansen FB (1998) Mate Availability and Fecundity Selection in Multi-Allelic Self- Incompatibility Systems in Plants. *Evolution* **52**, 19-29.

Wright S (1929) Fisher's theory of dominance. *American Naturalist* **63**, 274-279.

Wright S (1939) The distribution of self sterility alleles in populations. *Genetics* **24**, 538-552.

Supplemental data

COD system with 4 S-haplotypes

			Fecundity selection					**Wright**				
			Initial frequencies									
			S1	S2	S3	S4	S3 mutant	S1	S2	S3	S4	S3 mutant
			0.25	0.25	0.25	0.25		0.25	0.25	0.25	0.25	
			5	5	5	5	0.01	5	5	5	5	0.01
Inbreeding depression	Sheltered genetic load	Mutant appearance	Equilibrium frequencies									
			S1	S2	S3	S4	S3 mutant	S1	S2	S3	S4	S3 mutant
			Haplotype mutant : specificity 3, dominant on all the other S-haplotype									
		Pistil and pollen	0.26	0.26	0.00	0.26	0.23	0.26	0.26	0.00	0.26	0.23
			6	6	0	6		6	6	0	6	
		Pistil only	0.25	0.25	0.00	0.25	0.22	0.25	0.25	0.25	0.25	0.00
			6	6	0	6		5	5	4	5	
		Pollen only	0.25	0.25	0.00	0.25	0.22	0.25	0.25	0.00	0.25	0.24
			6	6	0	6		5	5	0	5	
			Haplotype mutant : specificity 3, recessive on all the other S-haplotype									
		Pistil and pollen	0.13	0.13	0.00	0.13	0.62	0.13	0.13	0.00	0.13	0.57
	No sheltered load d=0		3	3	0	3		3	3	3	3	
		Pistil only	0.13	0.13	0.00	0.13	0.54	0.25	0.25	0.25	0.25	0.04
			5	5	0	5		4	4	3	4	
		Pollen only	0.13	0.13	0.00	0.13	0.54	0.13	0.13	0.00	0.13	0.48
			5	5	0	5		7	7	0	7	
			Haplotype mutant : specificity 3, recessive in the pollen, dominant in the pistil									
		Pistil and pollen	0.13	0.13	0.00	0.13	0.45					
			8	8	0	8						
			Haplotype mutant : specificity 3, dominant in the pollen, recessive in the pistil									
		Pistil and pollen	0.13	0.13	0.00	0.13	0.45					
No inbreeding depression s=0 and δ=0			8	8	0	8						
			Haplotype mutant : specificity 3, dominant on all the other S-haplotype									
		Pistil and pollen	0.25	0.25	0.00	0.25	0.30					
			3	3	0	3						
		Pistil only	0.25	0.25	0.00	0.25	0.29					
			4	4	0	4						
		Pollen only	0.25	0.25	0.00	0.25	0.29					
			4	4	0	4						
			Haplotype mutant : specificity 3, recessive on all the other S-haplotype									
		Pistil and pollen	0.10	0.10	0.10	0.10	0.26					
	High sheltered load d=1		9	9	9	9						
		Pistil only	0.20	0.20	0.10	0.20	0.25					
			1	1	0	1						
		Pollen only	0.20	0.20	0.10	0.20	0.25					
			1	1	0	1						
			Haplotype mutant : specificity 3, recessive in the pollen, dominant in the pistil									
		Pistil and pollen										
			Haplotype mutant : specificity 3, dominant in the pollen, recessive in the pistil									
		Pistil and pollen										
High inbreeding depression s=0.8 and δ=0.5,	No sheltered load d=0		Haplotype mutant : specificity 3, dominant on all the other S-haplotype									
		Pistil and pollen	0.25	0.25	0.00	0.25	0.23	0.25	0.25	0.00	0.25	0.23
			6	6	0	6		6	6	0	6	
		Pistil only	0.25	0.25	0.00	0.25	0.22	0.25	0.25	0.25	0.25	0.00
			6	6	0	6		5	5	4	5	
		Pollen only	0.25	0.25	0.00	0.25	0.22	0.25	0.25	0.00	0.25	0.24
			6	6	0	6		5	5	0	5	
			Haplotype mutant : specificity 3, recessive on all the other S-haplotype									
		Pistil and pollen	0.13	0.13	0.00	0.13	0.62	0.13	0.13	0.00	0.13	0.57
			3	3	0	3		3	3	3	3	

Pistil only	0.15	0.15	0.00	0.15	0.54	0.24	0.24	0.23	0.24	0.04
Pollen only	0.15	0.15	0.00	0.15	0.54	0.17	0.17	0.00	0.17	0.48

Haplotype mutant : specificity 3, recessive in the pollen, dominant in the pistil

Pistil and pollen	0.23	0.23	0.23	0.13	0.18

Haplotype mutant : specificity 3, dominant in the pollen, recessive in the pistil

Pistil and pollen	0.25	0.25	0.25	0.15	0.11

DOM model with 3 S-haplotypes (S1=class II; S2=classIV; S3=class VI)

Inbreeding depression	Sheltered genetic load		Fecundity selection				Wright			
			S1 MUTANT							
			Initial frequencies							
						mutant				mutant
		S-Haplotype	S1	S2	S3	S1	S1	S2	S3	S1
No inbreeding depression s=0 and δ=0	No sheltered load d=0		0.60	0.22	0.17	0.01	0.60	0.22	0.17	0.01
			Observed frequencies after 20 generations							
			S1 mutant = class VII							
						mutant				mutant
		S-Haplotype	S1	S2	S3	S1	S1	S2	S3	S1
		Pistil and pollen	0.60	0.23	0.17	0.00	0.60	0.23	0.17	0.00
		Pistil only	0.60	0.22	0.17	0.01	0.60	0.23	0.17	0.01
		Pollen only	0.60	0.22	0.17	0.01	0.60	0.22	0.17	0.00
			S1 mutant = class VI							
		Pistil and pollen	0.61	0.22	0.17	0.00	0.61	0.23	0.17	0.00
		Pistil only	0.61	0.22	0.17	0.00	0.60	0.23	0.17	0.01
		Pollen only	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
			S1 mutant =class V							
		Pistil and pollen								
		Pistil only	0.60	0.22	0.17	0.01				
		Pollen only	0.60	0.22	0.17	0.01				
			S1 mutant = class IV							
		Pistil and pollen	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
		Pistil only	0.61	0.22	0.17	0.00	0.60	0.22	0.17	0.00
		Pollen only	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
			S1 mutant = class I							
		Pistil and pollen	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.01
		Pistil only	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.01
		Pollen only	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.01
			S2 MUTANT							
						mutant				mutant
		S-Haplotype	S1	S2	S3	S2	S1	S2	S3	S2
		Initial frequencies	0.61	0.22	0.17	0.01				
			Observed frequencies after 20 generations							
						mutant				mutant
			S1	S2	S3	S2	S1	S2	S3	S2
			S2 mutant=class VII							
		Pistil and pollen	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
		Pistil only	0.61	0.22	0.17	0.01	0.61	0.22	0.17	0.00
		Pollen only	0.61	0.22	0.17	0.01	0.61	0.22	0.17	0.00
			mutant S2= class VI							
		Pistil and pollen	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
		Pistil only	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
		Pollen only	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
			mutant S2 = class II							
		Pistil and pollen	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
		Pistil only	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.01

Pollen only	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
				mutant S2= class I				
Pistil and pollen	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
Pistil only	0.61	0.22	0.17	0.01	0.61	0.22	0.17	0.01
Pollen only	0.61	0.22	0.17	0.01	0.61	0.22	0.17	0.00

				S3 MUTANT				
S-Haplotype	S1	S2	S3	S3 mutant	S1	S2	S3	S3 mutant
Initial frequencies	0.61	0.22	0.16	0.01	0.61	0.22	0.16	0.01
				Observed frequencies after 20 generations				
	S1	S2	S3	S3 mutant	S1	S2	S3	S3 mutant
				mutant S3 = class VII				
Pistil and pollen	0.61	0.22	0.16	0.00	0.61	0.22	0.16	0.00
Pistil only	0.61	0.22	0.16	0.00	0.61	0.22	0.16	0.00
Pollen only	0.61	0.22	0.16	0.00	0.61	0.22	0.16	0.00
				mutant S3 = class IV				
Pistil and pollen	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
Pistil only	0.61	0.22	0.17	0.00	0.61	0.22	0.16	0.01
Pollen only	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
				mutant S3 = class III				
Pistil and pollen								
Pistil only	0.61	0.22	0.16	0.01				
Pollen only	0.61	0.22	0.16	0.01				
				mutant S3 = class II				
Pistil and pollen	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
Pistil only	0.61	0.22	0.17	0.00	0.61	0.23	0.16	0.01
Pollen only	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
				mutant S3= class I				
Pistil and pollen	0.61	0.22	0.17	0.00	0.61	0.22	0.17	0.00
Pistil only	0.61	0.22	0.16	0.01	0.61	0.22	0.17	0.01
Pollen only	0.61	0.22	0.16	0.01	0.61	0.22	0.16	0.00

Sheltered load d1=0.2; d2=0.6; d3=0.8

				Observed frequencies after 20 generations				
				mutant S1 = class VII				
S-Haplotype	S1	S2	S3	mutant S1	S1	S2	S3	mutant S1
Pistil and pollen	0.43	0.26	0.21	0.10	0.43	0.25	0.22	0.10
Pistil only								
Pollen only								
				S1 mutant = class VI				
Pistil and pollen								
Pistil only								
Pollen only								
				S1 mutant =class V				
Pistil and pollen								
Pistil only								
Pollen only								
				S1 mutant = class IV				

Pistil and pollen

Pistil only

Pollen only

S1 mutant = class I

Pistil and pollen

Pistil only

Pollen only

S2 MUTANT

Observed frequencies after 20 generations

S-Haplotype	S1	S2	S3	mutant S2	S1	S2	S3	mutant S2
				S2 mutant=class VII				
Pistil and pollen	0.58	0.09	0.21	0.12	0.57	0.08	0.21	0.14
Pistil only								
Pollen only								
				mutant S2= class VI				
Pistil and pollen								
Pistil only								
Pollen only								
				mutant S2 = class II				
Pistil and pollen								
Pistil only								
Pollen only								
				mutant S2= class I				
Pistil and pollen	0.58	0.22	0.20	0.00	0.57	0.23	0.21	0.00
Pistil only								
Pollen only								

S3 MUTANT

Observed frequencies after 20 generations

S-Haplotype	S1	S2	S3	S3 mutant	S1	S2	S3	S3 mutant
				mutant S3 = class VII				
Pistil and pollen								
Pistil only								
Pollen only								
				mutant S3 = class IV				
Pistil and pollen								
Pistil only								
Pollen only								
				mutant S3 = class III				
Pistil and pollen								
Pistil only								
Pollen only								
				mutant S3 = class II				
Pistil and pollen								
Pistil only								
Pollen only								

				mutant S3= class I				
Pistil and pollen	0.58	0.22	0.20	0.00	0.57	0.23	0.21	0.00
Pistil only								
Pollen only								

Discussion générale

1. Preuves empiriques de la sélection balancée en population naturelle

La sélection balancée est une force de sélection tendant à augmenter la variabilité génétique en opposition à la sélection directionnelle qui favorise la fixation d'allèles, et qui tend donc à diminuer la variabilité génétique. L'action de la sélection balancée est cependant souvent modifiée par d'autres processus évolutifs, tels que la dérive génétique, la migration, ou le déterminisme génétique des phénotypes sélectionnés. La sélection balancée est invoquée pour des traits divers : elle semble être un phénomène important dans la résistance aux pathogènes, que ce soit chez les plantes (Bergelson *et al.*, 2001) ou chez les animaux tels les mammifères (Hedrick, 2002). Cependant, puisque des preuves empiriques détaillées de ce type de sélection restent rares, il nous semblait donc important d'apporter des observations de ses effets pour un système modèle, le système d'auto-incompatibilité chez les *Brassicaceae*.

a. Approches directes

Nous souhaitions mettre en évidence l'effet de la sélection fréquence-dépendante au locus d'auto-incompatibilité en population naturelle. L'idée était de comparer les succès reproducteurs des individus dans une population naturelle, et de vérifier si les individus portant des allèles d'auto-incompatibilité rares avaient un succès reproducteur supérieur aux individus portant des allèles communs. Il serait ainsi possible de comparer les succès reproducteurs mâles, estimés par le nombre de pistils fécondés par une même plante, et les succès maternels qu'on pourrait évaluer par la production de graines de chaque plante. Cependant, le succès reproducteur dans le cas de l'auto-incompatibilité sporophytique dépend du phénotype exprimé par le pollen ou le pistil. Ce phénotype est déterminé par le génotype et les relations de dominance entre allèles du locus S. Il faut donc connaitre à la fois tous les génotypes des individus de la population, ainsi que toutes les relations de dominance entre les allèles présents, afin de déterminer les phénotypes. D'autre part, le succès reproducteur des plantes dépend fortement de la composition génotypique de la population, c'est-à-dire du nombre de partenaires compatibles présents dans la population, de la capacité de dispersion des individus, ainsi que de la répartition spatiale des individus.

Encadré 1 : La disponibilité en partenaire

- Définitions

La disponibilité en partenaire dans une population est définie comme la proportion de plantes compatibles avec un gamète donné choisi aléatoirement dans la population.
Il est possible de la calculer pour chaque individu en faisant le rapport entre le nombre de partenaires compatibles avec un gamète issu de cet individu et le nombre total d'individus dans la population. Ainsi si les phénotypes d'auto-incompatibilité exprimé par le pollen et le pistil sont différents, la disponibilité en partenaire pourra prendre des valeurs différents pour le pollen et le pistil.

-Effet du type de système d'auto-incompatibilité sur la disponibilité en partenaire

Vekemans *et al.*{, 1998 #68} ont étudié théoriquement l'effet du type de système d'auto-incompatibilité (GSI ou SSI) et du schéma de relation de dominance entre allèles sur la disponibilité en partenaire. Dans les systèmes sporophytiques, les systèmes DOM ont une disponibilité en partenaires plus élevé que les systèmes DOMCOD ou COD.

TABLE 3. Mate availability (i.e., the proportion of compatible matings) at deterministic equilibrium for two, three, four, five, six, seven, and 11 alleles, in models with or without fecundity selection (FS). For the GSI model, values of pollen compatibility are also included. Number of alleles = n.

n	GSI (pollen compatibility)	GSI	SSIcod	SSIdom	SSIdomcod (without FS)	SSIdomcod (with FS)
2	—	—	—	0.500	—	—
3	0.333	0.667	—	0.667	0.455	0.461
4	0.500	0.833	0.167	0.750	0.571	0.584
5	0.600	0.900	0.300	0.800	0.646	0.663
6	0.667	0.933	0.400	0.833	0.699	0.716
7	0.714	0.952	0.476	0.857	0.738	0.755
11	0.818	0.982	0.655	0.909	0.828	0.842

Tableau d'après Vekemans *et al.*{, 1998 #68}

Lors de l'étude de la population de Nivelle, nous nous sommes donc attachés à décrire la structure génétique de la population grâce à des marqueurs neutres (chapitre 1, premier article), ainsi qu'à caractériser les génotypes au locus S et les relations de dominance entre allèles S (chapitre 2). Ceci nous a permis de calculer la disponibilité en partenaires pour chaque individu de la population (Encadré 1). La disponibilité en partenaires dans la population de Nivelle était très élevée dans le pollen comme dans le pistil : nous avons trouvé en moyenne 78.4% (variance: 0.009) de partenaires compatibles pour le pollen et 79.3% (variance 0.012) pour le pistil. Cette observation correspond aux attendus théoriques pour un système DOMCOD avec 7 allèles (Encadré 1). En effet, le nombre élevé d'allèles observés à Nivelle (au minimum 8), ainsi que les patrons de dominance entre allèles très hiérarchisés, tendent à minimiser les cas de pollinisation incompatibles, et par conséquent à une disponibilité élevée en partenaires. Cette disponibilité élevée en partenaires est sans doute l'un des facteurs important de la forte diversité en pères observée chez les descendants de la population de Nivelle (chapitre 1, premier article).

Nous avons ensuite tenté de voir si la disponibilité en partenaires avait une influence sur le succès reproducteur paternel. En effet l'analyse de paternité menée dans cette population nous permettait d'estimer le nombre de pistils fécondés par plante. Avec la collaboration d'Etienne Klein (Chargé de recherche ; INRA d'Avignon-Unité de biométrie), nous avons utilisé le modèle décrit par Oddou-Muratorio *et al.* (2005) qui permet de détecter l'influence de fécondité variable sur les patrons de reproduction. Cette méthode permet de prendre en compte à la fois la répartition des individus dans l'espace et les fécondités des individus, qu'elles soient par exemple liées à la taille des individus, ou au phénotype d'auto-incompatibilité dans notre cas. Nous avons donc défini des groupes de plantes ayant la même disponibilité en partenaires et testé l'effet de cette disponibilité sur le succès reproducteur paternel. Aucun effet significatif n'a été détecté (p = 0.21), ne permettant pas de mettre en évidence l'avantage des phénotypes rares dans le succès reproducteur paternel.

La sélection fréquence-dépendante décrite dans le modèle de Wright (1939) (c'est-à-dire agissant par le biais du succès reproducteur du pollen portant des allèles d'auto-incompatibilité rares) n'a donc pas pu être démontré de cette manière dans le cas de la population de Nivelle. Plusieurs raisons peuvent expliquer que cette approche n'a pas permis de détecter la sélection fréquence-dépendante. Tout d'abord, notre échantillonnage semblait trop réduit: nous n'avions pas beaucoup de plante-mère par classe de disponibilité en partenaires, ce qui a diminué considérablement notre pouvoir statistique. D'autre part, notre

population présentait une disponibilité en partenaires très élevée, et relativement peu variables, ce qui signifie que quel que soit leur phénotype d'auto-incompatibilité, toutes les plantes disposaient d'un grand nombre de partenaires compatibles à proximité. Les données de la courbe de dispersion de pollen (chapitre 1, premier article) suggéraient une dispersion du pollen importante dans la population. Cette capacité de dispersion, associée à la forte densité en individus, conférait aux grains de pollen un accès à un grand nombre de partenaires possibles. En conséquence, les différences de succès reproducteur paternels étaient probablement assez faibles dans cette population. Ces hypothèses pourraient expliquer pourquoi aucune action de sélection fréquence-dépendante n'a pu être détectée dans la population de Nivelle. L'application de la même approche avec un échantillonnage plus important dans une population à faible densité pourraient vraisemblablement donner des résultats plus clairs en faveur de la sélection fréquence-dépendante. Il pourrait également être intéressant de mesurer la production de graines pour évaluer le succès reproducteur maternel des individus. Il serait ainsi possible de tester si la sélection fréquence-dépendante agit à la fois par la voie mâle et la voie femelle comme suggéré dans le modèle de *fecundity selection* (Vekemans *et al.*, 1998), notamment dans des populations où la limitation en pollen est forte. Une autre façon de tester l'existence de la sélection fréquence-dépendante serait de créer une population artificielle avec des individus de génotype connus situés à égale distance d'individus cibles, afin d'éliminer les biais introduits par la structure génétique spatiale. Toutefois, cette approche « artificielle » ne permettrait pas une estimation réaliste de l'effet de la sélection fréquence-dépendante en population naturelle.

Les preuves directes de la sélection fréquence-dépendante sont un point crucial de la compréhension de l'intensité et de l'occurrence de ce type de sélection. Les vérifications expérimentales de ce type de sélection restent rares. Des preuves indirectes de la sélection fréquence-dépendante au locus d'auto-incompatibilité peuvent cependant être obtenues.

b. Approches indirectes

De nombreux modèles théoriques ont permis de dégager des résultats attendus de l'effet de la sélection fréquence-dépendante : sur le polymorphisme aux locus soumis à la sélection balancée, sur la répartition de la variabilité génétique dans l'espace, et sur l'évolution des séquences.

La sélection balancée a également un fort effet sur le polymorphisme allélique : le nombre d'allèles maintenu et la distribution des fréquences alléliques pour un locus soumis à sélection balancée sont très fortement différents d'un cas concernant un locus neutre. Les données de polymorphisme au locus S récoltées dans la population de Nivelle montrent une signature claire de la sélection balancée sur le polymorphisme observé au locus S (chapitre 2). Cependant, l'effet de la sélection balancée est également modifié par les relations de dominance et la dérive génétique. La détection de la sélection balancée en population naturelle nécessite donc des approches statistiques appropriées permettant de tester différents modèles alternatifs, et par conséquent de repérer l'effet des différents facteurs en interaction.

La sélection balancée a aussi des conséquences sur les flux de gènes : dans le cas de l'avantage du rare, un allèle migrant sera favorisé lors de son arrivée dans une nouvelle population puisqu'il aura une fréquence très faible. En augmentant la sélection positive pour les allèles migrants, la fréquence-dépendante tend à uniformiser la répartition de la variabilité génétique dans l'espace au locus concerné. Ainsi la structure génétique spatiale pour des locus soumis à la sélection balancée serait moins forte que pour des locus neutres (Schierup *et al.*, 1997).

A l'échelle de la population, les résultats obtenus chez *A. halleri* montrent une différence significative entre le patron de structure génétique spatiale au locus S par rapport aux marqueurs neutres seulement dans une population. Schueler *et al.* (2006) n'ont pas détecté de différence significative dans les patrons d'isolement par la distance pour des marqueurs neutres et le locus d'auto-incompatibilité gamétophytique chez *Prunus avium*. Nos résultats théoriques ont cependant montré que la différence de structure génétique spatiale entre marqueurs neutres et locus S dépend très fortement de la capacité de dispersion des individus, du taux d'immigration, et du nombre d'allèles au locus S (chapitre1, deuxième article). La signature de la sélection balancée sur la structure génétique spatiale intra-population dépend donc de l'écologie de l'espèce considérée, ainsi que de la diversité allélique au locus S, ce qui pourrait expliquer la difficulté de détection de ce phénomène en population naturelle.

A l'échelle inter-populationnelle, l'efficacité supérieure de la migration des allèles d'auto-incompatibilité par rapport aux allèles présents à des locus neutres a été démontrée. Une étude comparative de la différenciation génétique entre populations de *Beta insularis* (*Chenopodiaceae*) a mis en évidence que le F_{ST} pour le locus S était significativement plus faible que pour des locus neutres (Glémin *et al.*, 2005). Chez *A. halleri*, une étude similaire a

été menée sur des populations réparties à l'échelle européenne, et les F_{ST} pour les locus neutres étaient également significativement plus faibles que le F_{ST} au locus S (S. Le Cam, M2R et V. Castric, MCF, Université de Lille 1, données non publiées). Cette répartition de la variabilité génétique au locus S entre populations chez *B. insularis* et *A. halleri* est une signature de la sélection balancée agissant sur les locus d'auto-incompatibilité.

La sélection balancée a des conséquences importantes sur le polymorphisme des séquences des locus concernés. La sélection balancée favorise en effet la divergence des allèles; ainsi les mutations non-synonymes sont soumises à une sélection positive puisqu'elles aboutissent à la création de nouveaux allèles. Ainsi au locus de détermination du sexe chez les Hyménoptères (locus *csd*), décrit comme fortement influencé par la sélection balancée, Hasselmann *et al.* (2004) ont détecté un excès de mutations non-synonymes par rapport aux mutations synonymes. Cette sélection en faveur des mutations non synonymes dans les séquences des locus soumis à sélection balancée a également été observée aux locus impliqués dans l'auto-incompatibilité chez les *Solanaceae* (Clark & Kao, 1991) et les *Brassicaceae* (Charlesworth *et al.*, 2003b). D'autre part, la sélection balancée permet la conservation des allèles sur des périodes de temps importantes. Ainsi, les études théoriques suggèrent que les allèles des locus soumis à la sélection balancée auraient des durées de vie supérieures aux successions d'espèces : un polymorphisme trans-spécifique pourrait donc être observé (Wiuf *et al.*, 2004). Cette signature moléculaire de la sélection balancée a été observée expérimentalement dans le cas de l'auto-incompatibilité chez les *Solanaceae* (Ioerger *et al.*, 1990) et les *Brassicaceae* (Castric & Vekemans, 2007).

La sélection balancée agissant à long terme sur les séquences, elle peut également être détectée sur les taux de mutation synonymes / non-synonymes pour des régions flanquantes du locus d'auto-incompatibilité, qui subissent l'effet de la sélection balancée par entrainement génétique (Kamau & Charlesworth, 2005).

Les méthodes indirectes permettent donc de repérer la signature de la sélection balancée mais d'autres phénomènes démographiques ou génétiques comme les effets d'auto-stop moléculaires peuvent interférer dans la détection de cette signature. La détection de la sélection balancée en population naturelle nécessite donc d'étudier en détails les interférences avec d'autres processus évolutifs. Cette approche permet de déterminer l'importance de ces différentes forces évolutives sur le locus S, et ainsi de mieux comprendre l'évolution des systèmes d'auto-incompatibilité.

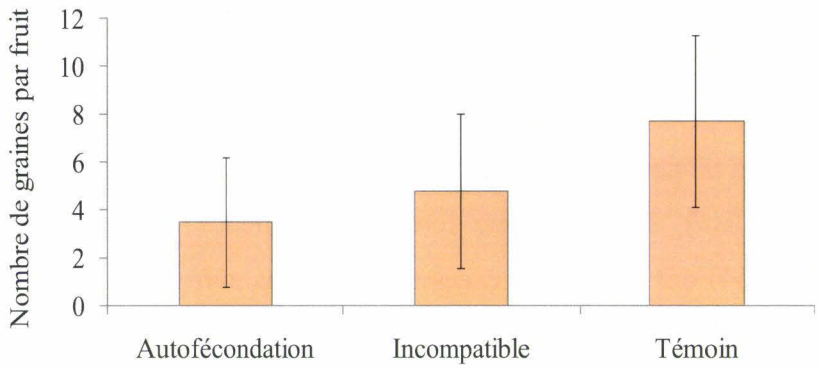


Figure 12 : Nombre de graines par fruits obtenus pour des croisements témoins, incompatibles et autofécondation avec le génotype maternel **(Sh01 ; Sh12)**

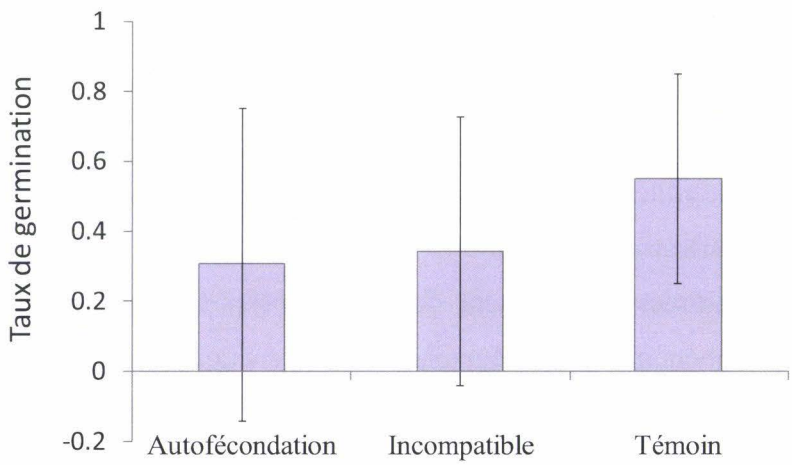


Figure 13: Taux de germination des graines issues de croisements témoins, incompatibles et autofécondation avec le génotype maternel **(Sh01 ; Sh12)**

2. Maintien de l'auto-incompatibilité en population naturelle

Un grand nombre de mécanismes ont été évoqués pour expliquer la diversité des systèmes de reproduction sexuée observée chez les organismes hermaphrodites : Jarne & Charlesworth(1993) évoquent notamment l'assurance reproductive, le coût de la recherche de partenaires, ou la dépression de consanguinité comme des facteurs conditionnant le taux d'auto-fécondation chez les hermaphrodites. Cependant, de nombreuses questions restent ouvertes, comme le paradoxe de l'apparition multiple et indépendante de l'auto-incompatibilité chez les plantes hermaphrodites. En effet, il est difficile d'expliquer la mise en place et le maintien d'un système limitant l'accès aux partenaires sexuels en population naturelle. L'hypothèse généralement admise est la limitation des effets délétères de la dépression de consanguinité par le système d'auto-incompatibilité.

a. Mise en évidence d'un fardeau génétique lié au locus S

Au chapitre 3, nous avons évoqué l'hypothèse de la mise en place d'un fardeau génétique lié au locus d'auto-incompatibilité au cours de l'évolution des allèles d'auto-incompatibilité. Nous avons montré que chez *A. halleri* un fardeau important lié à l'allèle dominant Sh15 pouvait être mis en évidence tandis que le fardeau lié à l'allèle de dominance intermédiaire Sh02 apparaissait plus faible et qu'aucun fardeau n'était détecté pour l'allèle récessif Sh01.

Lors d'une étude préliminaire, nous avions effectué les mêmes manipulations sur 3 individus de génotype (Sh01 ; Sh12) mais avec des effectifs plus faibles. Nous avions trouvé des différences significatives entre les nombres de graines observés après croisements incompatibles et témoins (Fig. 12). La même tendance a été constatée pour le taux de germination. Nous avons donc émis l'hypothèse de l'existence d'un fardeau important lié à l'un des allèles du génotype (Sh01 ; Sh12). Comme nous avons suggéré que l'allèle Sh01 ne présentait pas ou peu de fardeau (chapitre 3), nous supposons donc que le fardeau observé ici serait plutôt lié à l'allèle Sh12. Nous n'avons observé aucune différence significative entre croisements incompatibles et autofécondations (Fig. 13) : l'effet du fardeau lié sur les

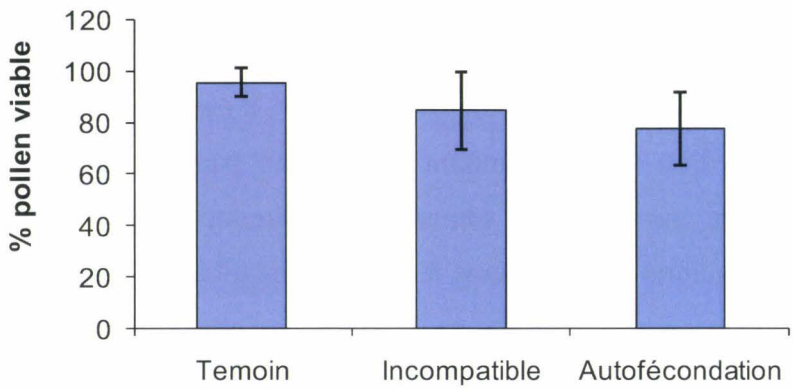


Figure 14: Proportion de pollen viable déterminé par coloration Alexander chez des descendants de croisements témoins, incompatibles et autofécondation avec le génotype maternel **(Sh01 ; Sh12)**

autofécondations semblait prépondérant par rapport à la dépression de consanguinité génomique.

Cependant, les résultats de ségrégation au locus S n'ont pas montré d'écart significatif à la ségrégation mendélienne. L'allèle Sh12 ayant un comportement atypique dans la hiérarchie de dominance et une fréquence relativement élevée par rapport à l'attendu déterministe dans la population de Nivelle (chapitre 2), nous avons émis l'hypothèse que sa ségrégation pouvait être biaisée en population naturelle. Afin de confirmer l'existence du fardeau lié, nous avons poursuivi les mesures de *fitness* sur les descendants de ces croisements.

Nous avons donc gardé les descendants des croisements jusqu'à la floraison dans le but d'étudier l'influence des mutations délétères au cours du cycle de vie. Nous avons choisi d'étudier le succès reproducteur mâle, car il semble être une des composantes majeures de la *fitness*. Sur 117 individus issus des croisements contrôlés, nous avons ainsi effectué des mesures de viabilité du pollen par coloration Alexander (Alexander, 1969), ainsi que des comptages du nombre de grains de pollen par anthère grâce à un compteur à particules CASY.TT®. Nous n'avons pas trouvé de différence significative du nombre de grains de pollen par anthère, mais nous avons constaté que la proportion de pollen viable était plus faible chez les descendants issus d'autofécondation par rapport aux croisements incompatibles, eux-mêmes légèrement plus faibles que les croisements témoins (Fig. 14). Cette tendance n'est pas significative, mais ces résultats suggèrent tout de même que les croisements incompatibles donnent des descendants exprimant des effets délétères tout au long du cycle de vie. L'étude du génotype (Sh01 ; Sh12) concernant un faible effectif, nos conclusions restent limitées mais cette étude souligne l'intérêt d'une évaluation la plus complète possible de la *fitness* pour appréhender les effets du fardeau génétique lié au locus S.

Il serait donc intéressant de répéter ce type d'approche en effectuant un suivi à long terme des descendants des différents types de croisements avec des mesures des composantes de la fitness, telles que le nombre et la taille des feuilles, le nombre de hampes florales, le taux de mise à fleur et de mise à graines ainsi que des mesures de quantité et de qualité du pollen. Cependant, nos observations suggèrent que certains effets des mutations délétères pourraient s'exprimer à des stades précoces du cycle de vie, il serait donc particulièrement important d'évaluer le taux d'avortement des ovules, ou de mesurer la ségrégation au locus S avant et après la germination des graines. L'extraction d'ADN étant une manipulation destructive, il faudrait réaliser plusieurs jeux de croisements où les individus seraient prélevés à différents stades.

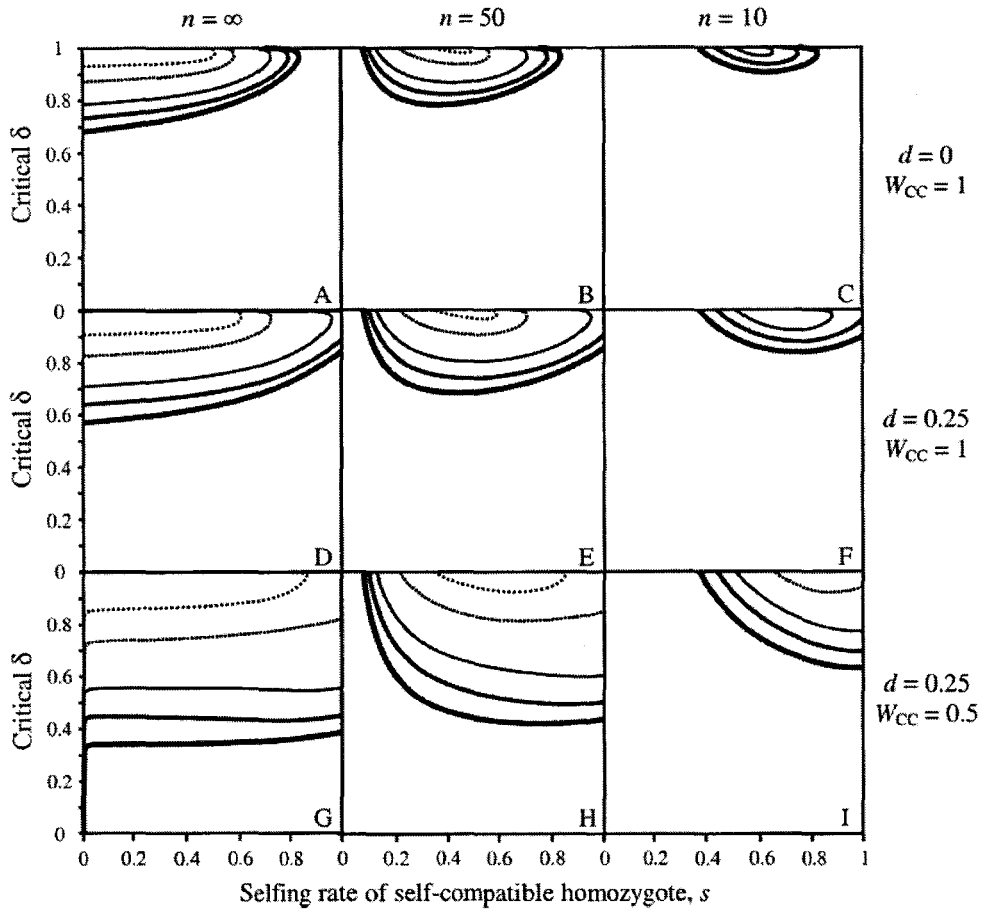


FIG. 1. Conditions for invasion of a self-incompatible population by a nonfunctional S-allele (model 1). The critical inbreeding depression due to lethals, δ, above which invasion cannot occur, is plotted against the selfing rate, s, of the homozygous self-compatible genotype CC, under various conditions of pollen limitation (amount of outcross pollen landing on the stigma per ovule P_o: coarse dotted line = 0.5, fine dotted line = 1, regular line = 2, thicker line = 3, thickest line = ∞), number of S-alleles (n = ∞, 50, and 10), background inbreeding depression (d = 0 and 0.25), and genetic load linked to the self-compatible allele ($1 - W_{CC}$ = 0 and 0.5). The selfing rates of the homozygous CC genotype, s, and of the heterozygous I_1C genotypes, s', are given in Table 1 as a function of the amount of self and outcross pollen, P_s and P_o, landing on the stigma.

Figure 15: Conditions d'invasion d'un allèle S auto-compatible, d'après Porcher & Lande (2004).

Dépression de consanguinité critique (δ) au dessus de laquelle l'auto-incompatibilité est stable en fonction du taux d'autofécondation des homozygotes auto-compatible (s), du niveau de fardeau génomique (d) et du fardeau génétique lié au locus S (1-W_{CC}) et de la limitation en pollen (ligne épaisse : 50% du pollen déposé sur les pistils provenant d'autres plantes ; ligne fine : infinité de pollen extérieur).

Nous avons interprété les résultats de ségrégation au locus S comme une signature d'un fardeau lié aux allèles S. Cependant, pour confirmer cette interprétation, il faudrait tester les mêmes allèles (Sh01, Sh02, Sh15 et Sh12) dans des génotypes différents, afin de distinguer les biais de ségrégation entre allèles des effets d'un fardeau associé.

D'autre part, nous avons suggéré dans le chapitre 3 que le poids du fardeau semble d'autant plus important qu'il est porté par un allèle dominant. Pour étayer ce résultat, il serait nécessaire de répéter la même manipulation sur d'autres allèles de classes dominantes, intermédiaires, et récessives, pour pouvoir conclure que le niveau de dominance de l'allèle conditionne le poids de son fardeau associé et non son identité. En effet, Stone (2004) suggérait que le fardeau accumulé pour un allèle pourrait dépendre de l'âge de cette allèle, c'est-à-dire de la longueur de sa branche dans la phylogénie d'allèle. Il serait particulièrement remarquable de distinguer l'effet de la dominance et de l'histoire des allèles du locus S sur le fardeau génétique accumulé.

Les allèles du locus S pouvant être transpécifiques, il est possible que les régions flanquantes du locus d'auto-incompatibilité puissent conserver des mutations délétères pendant des temps évolutifs supérieurs à l'échelle de l'espèce. Ainsi tout comme observé chez le chromosome Y des mammifères (Charlesworth & Charlesworth, 2000), ces régions génomiques pourraient subir une dégénérescence au cours de leur histoire évolutive, aboutissant à un verrouillage du système d'auto-incompatibilité.

b. Conséquences sur les hypothèses du maintien de l'autoincompatibilité

Le maintien de l'autoincompatibilité a généralement été expliqué par l'effet de la dépression de consanguinité génomique, et plus récemment par l'expression du fardeau lié au locus S. Cependant, le modèle de Porcher (2004) qui étudie le comportement d'allèles auto-compatible dans des populations d'espèces ayant un système d'auto-incompatibilité gamétophytique, a montré que les conditions de stabilité de l'auto-incompatibilité étaient restreintes : le modèle souligne qu'une dépression de consanguinité forte et un fardeau génétique lié important sont nécessaires pour expliquer le maintien de l'auto-incompatibilité (Fig. 15).

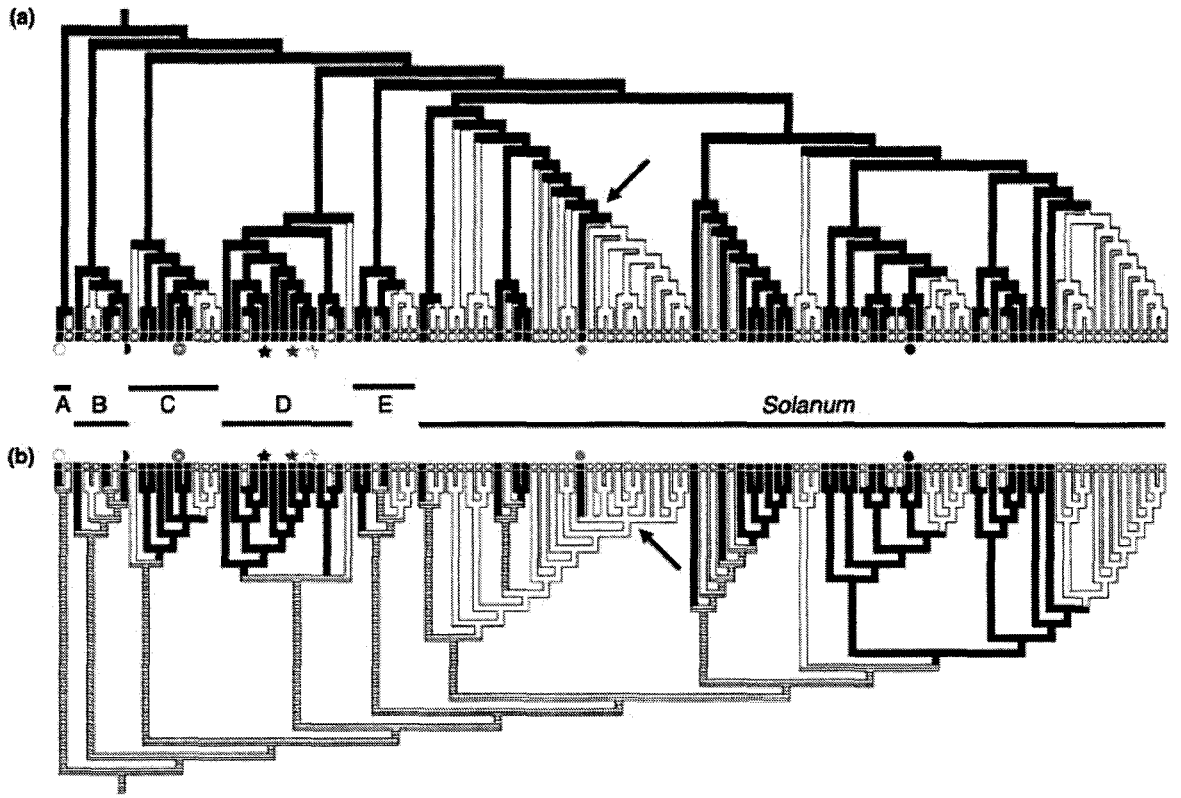


Fig. 3 Solanaceae species cladogram, SI lineages in black, SC in white, uncertain character state hatched. Bottom: Unweighted analysis (1 : 1; gain and loss of SI equally weighted) produces a minimum of 2 reversals. Species labelled with circles and stars appear in Fig. 2 and show *trans*-generic S-allele sharing. Top: Parsimony analysis with 5 : 1 weight applied to favor loss of SI. This is the minimum weighting needed to connect, with uninterrupted ancestry of SI, all SI taxa showing *trans*-specific S-polymorphism. This weighting produces no inferred reversals from SC to SI. Parsimony analyses were conducted in MacClade 4.0 (Maddison & Maddison, 2000). See text for results of maximum likelihood (ML) analyses. Black arrows show an important polytomy and its conservative resolution used in the ML analyses presented. Labeled groups follow Olmstead *et al.* (1999) and are encoded as follows: A, subfam. Petunioideae; B, subfam. Nicotianoideae; C–E, subfam. Solanoideae, with C, (tribes Lycieae + Hyoscyameae); D, (tribe Datureae + genus *Jaltomata*); E, tribe Capsiceae; *Solanum* (genus).

Figure 16 : Cladogramme des espèces de la famille des Solanacées, d'après Igic *et al.* (2004). Les espèces auto-incompatibles sont représentées en noir et auto-compatible en blanc. L'hypothèse d'une auto-incompatibilité ancestrale dans cette famille semble la plus parcimonieuse. De nombreux événements indépendants de rupture du système d'auto-incompatibilité sont observés.

L'auto-incompatibilité est cependant un phénomène très répandu chez les Angiospermes avec plus de 50 familles botaniques concernées et il semble difficile d'expliquer un si grand nombre de cas d'auto-incompatibilité par des conditions de dépression de consanguinité si drastiques. D'autre part, notre étude empirique sur la dépression de consanguinité et le fardeau génétique lié au locus S chez *A. halleri* suggère une dépression moyenne et un fardeau moyen pour les allèles dominants voire nul pour les allèles récessifs (chapitre 3). Or la population dont sont issues les plantes étudiées semble pourvue d'un système d'auto-incompatibilité fonctionnel comme l'on suggéré les estimations du taux d'allo-fécondation (t_m=0.98, chapitre 1, premier article) et les auto-pollinisations contrôlées (3% d'autofécondation donnant un fruit, chapitre 2). Il est donc probable que le niveau de dépression de consanguinité et de fardeau génétique observé dans la population de Nivelle ne soit donc pas forcément suffisant pour expliquer le maintien de l'auto-incompatibilité dans cette population et à plus large échelle dans cette espèce.

L'auto-incompatibilité semble ancestrale dans plusieurs familles : chez les Solanacées, Igic *et al.* (2004) suggère que l'hypothèse de l'auto-incompatibilité comme état ancestral est la plus probable (Fig. 16). Chez les *Brassicaceae*, il semble également que l'auto-incompatibilité soit l'état ancestral et qu'elle ait été perdue récemment chez *A. thaliana* par exemple (Bechsgaard *et al.*, 2006). Le système d'auto-incompatibilité semble donc pouvoir être supprimé au cours des lignées évolutives. La rupture de l'auto-incompatibilité pourrait être favorisée par l'avantage du à l'autofécondation, qui permet une transmission doublée des gènes. L'autofécondation présente également des avantages due à l'assurance reproductive en particulier dans des petites populations à faible disponibilité en partenaire. Le niveau de dépression de consanguinité porté par les populations naturelles pourrait également déterminer l'invasion de l'auto-compatibilité. Des conditions démographiques, écologiques et génétiques particulières pourraient donc permettre la rupture de l'auto-incompatibilité. Cependant ces conditions restent encore aujourd'hui à définir si bien que la question du maintien ou de la perte de l'auto-incompatibilité reste ouverte.

3. Génétique et évolution de la dominance au locus d'auto-incompatibilité

Lors des croisements contrôlés effectués pour déterminer les relations entre allèles S (chapitre 2), nous avons observé des cas de variation de phénotype d'auto-incompatibilité exprimé par des individus différents ayant un même génotype au locus S. Notamment dans les génotypes [Sh01 ; Sh02] et [Sh04 ; Sh12], dans le pollen, les allèles Sh01 et Sh04 étaient en général récessifs par rapport à Sh02 et Sh12, respectivement, mais pour certains individus les deux allèles semblaient être codominant dans le pollen.

C'est à partir de cette observation que nous nous sommes intéressés à l'évolution de la dominance au locus d'auto-incompatibilité. Ainsi que nous l'avons évoqué dans le chapitre 4, les mécanismes moléculaires de la dominance au locus S chez les *Brassicaceae* sont en cours de caractérisation. Pour le gène du pollen (*SCR* or *Sp11*), quelques études ont permis de comprendre une partie du mécanisme de dominance impliqué. Shiba *et al.* (2002) ont montré que chez Brassica *sp.*, les relations de dominance sont générées au niveau de la transcription du gène *SCR*. Fujimoto *et al.* (2006) ont démontré que l'expression d'un allèle *SCR* récessif pouvait être réprimée par un haplotype dominant dont la région codante de *SCR* n'était pas transcrite. La dominance serait donc indépendante de la protéine SCR elle-même et donc serait contrôlée par un autre facteur associé au locus S. En effet, Shiba *et al.* (2006) ont montré que les séquences promotrices situées dans la région 5' d'un allèle *SCR* récessif étaient méthylées spécifiquement dans les cellules du *tapetum* de l'anthère avant l'initiation de la transcription de *SCR*. Ces résultats suggèrent donc que la méthylation de l'ADN pourrait être responsable de l'expression de la dominance entre allèles du gène *SCR*.

Hatakeyama *et al.* (2001) ont suggéré que les relations de dominance entre allèles S dans le pistil étaient déterminées par le récepteur SRK lui-même, et non pas un résultat de son niveau de transcription relatif. Naithani *et al.* (2007) ont ainsi mis en évidence qu'en absence de ligand, la protéine récepteur kinase (SRK) formait des dimères, ce qui pourrait être une condition pré-requise à l'activation du récepteur pour la fixation du ligand. Ils ont donc fait l'hypothèse que la dominance reposerait sur la capacité des différentes paires d'allèle *SRK* à former des hétérodimères. Cette capacité dépendrait donc de la propre séquence du gène *SRK* ce qui revient à dire que la dominance au gène *SRK* serait intrinsèque.

Ces études récentes évoquent donc des mécanismes de dominance différents pour les gènes pollen et pistil. Ceci pourrait avoir des conséquences importantes sur l'évolution de la dominance pour ces deux gènes.

Le mécanisme de la dominance a fait l'objet d'un long débat en biologie évolutive, qui a démarré entre R. A. Fisher (Fisher, 1928a; Fisher, 1928b) et S. Wright (Wright, 1929; Wright, 1934). D'un côté, R. A. Fisher pensait que la dominance pouvait évoluer par sélection

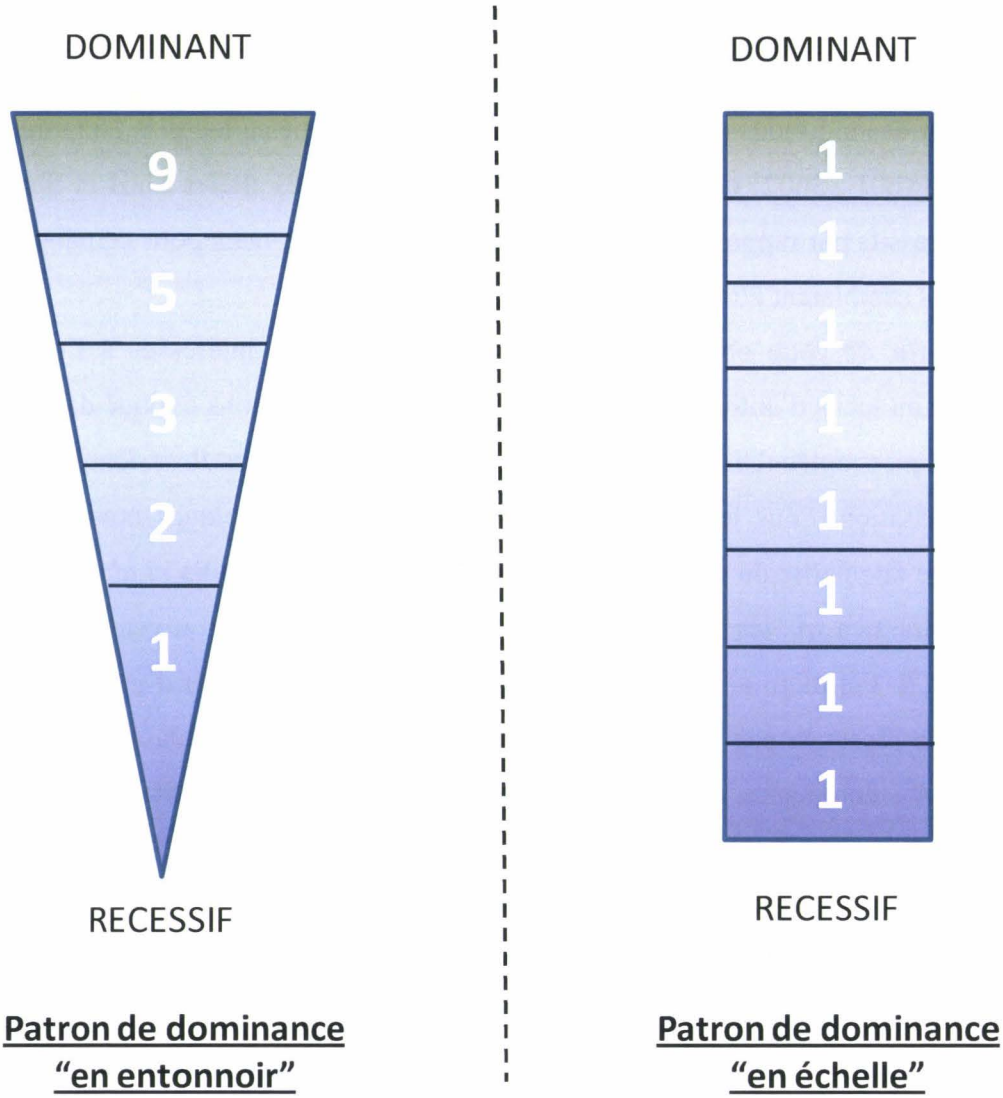


Figure 17: Exemples de patrons de dominance « en entonnoir » ou « en échelle » au locus S, le nombre d'allèles présents dans chaque classe de dominance est indiqué en blanc.

naturelle sur un locus "modifieur". A l'opposé, S. Wright supposait que la dominance était imposée par les propriétés physiologiques de la molécule exprimée. La théorie de Fisher est fondée sur l'existence d'une sélection naturelle agissant sur les hétérozygotes: cette hypothèse est apparue peu réaliste au vu de la taille de population nécessaire pour que ce type de sélection puisse être efficace (Veitia, 2006). La théorie de Wright a donc été retenue comme modèle le plus réaliste et l'exemple de la dominance des enzymes impliquées dans des cascades de réaction est venue confirmer cette vision (Kacser & Burns, 1981). Dans le cas particulier de locus soumis à une sélection balancée, l'hétérozygotie étant très élevée, l'évolution de la dominance proposée par Fisher serait néanmoins possible (Otto & Bourguet, 1999). Le locus d'auto-incompatibilité, comme d'autres locus soumis à la sélection balancée, est donc un cas privilégié pour l'étude de l'évolution de la dominance.

Dans le chapitre 4, nous avons modélisé un système où la dominance pouvait évoluer indépendamment de la spécificité d'auto-incompatibilité. Nous avons montré que, comme attendu, le locus d'auto-incompatibilité étant soumis à la sélection balancée, l'évolution de la dominance de type Fisher était possible. Nous avons vu qu'il existait alors une tendance à la hiérarchisation de la dominance avec une fixation de la position des différents allèles S. Cependant, il est possible que la dominance soit intrinsèque à la nature biochimique des allèles. Dans ce cas, l'évolution de la dominance ne serait pas possible, chaque allèle aurait alors une dominance fixée : c'est la formalisation qui était proposée dans le modèle de Billiard *et al.* (2007), où les nouveaux allèles étaient introduits avec une spécificité donnée et un niveau de dominance fixé. Ce modèle suggère qu'à l'équilibre le nombre d'allèles par classe de dominance serait d'autant plus grand que le niveau de dominance serait élevé (Fig. 17). Ce modèle en « entonnoir » s'oppose donc au patron en « échelle » prédit par le modèle d'évolution de la dominance (chapitre 4). Les données moléculaires restent parcellaires mais semblent tout de même concordantes avec ces prédictions. En effet, pour le gène exprimé dans le pistil *SRK*, la dominance semble intrinsèque aux allèles et le patron de relation de dominance se rapproche d'une forme en « entonnoir » tandis que pour le gène pollen *SCR* la dominance serait indépendante de l'allèle et le patron de dominance aurait une forme en « échelle ». La différence dans les patrons de relation de dominance observés dans le pollen et le pistil pourrait également être expliquée par un investissement maternel fort dans la reproduction : si la reproduction entre apparentés a un coût plus fort du point de vue maternel que paternel, la codominance, qui permet une reconnaissance plus large des pollens par le pistil, pourrait être plus avantageuse dans le pistil que dans le pollen. Ce type d'évolution de la dominance différentielle a été modélisé dans le cas de relations hôte-parasite où la

dominance semblait être favorisée dans le gène de virulence du parasite tandis que la codominance était favorisée dans le gène de résistance de l'hôte (Nuismer & Otto, 2005).
De nombreux points restent à éclaircir du point de vue moléculaire pour permettre une compréhension générale de l'évolution de la dominance au locus S, mais cette première approche par modélisation a mis en évidence l'importance de ce phénomène sur l'évolution du locus S.

Conclusions

Pour comprendre les forces évolutives mises en jeu au locus d'auto-incompatibilité et leurs effets respectifs, je me suis attachée à prendre en compte les différentes échelles mises en jeu : celle du gène pour identifier les allèles d'auto-incompatibilité, du chromosome pour prendre en compte les mutations délétères liées, du génome pour prendre en compte la dépression de consanguinité de consanguinité, de l'individu pour déterminer le phénotype d'auto-incompatibilité, et de la population pour prendre en compte la sélection fréquence dépendante.

J'ai ainsi caractérisé empiriquement certains processus évolutifs déjà reconnus comme ayant une influence sur la dynamique des allèles d'auto-incompatibilité sporophytique, tels que la sélection fréquence-dépendante, les relations de dominance entre allèles, ou la dérive géentique (chapitres 1 et 2). J'ai également mis en évidence l'influence de deux autres phénomènes pouvant jouer un rôle très important dans les systèmes d'auto-incompatibilité sporophytique : le fardeau génétique lié au locus S (chapitre 3), ainsi que l'évolution de la dominance (chapitre 4).

Cependant, de nombreuses pistes restent encore à explorer.

Tout d'abord, l'analyse empirique des patrons de dominance est un élément clef pour la compréhension des systèmes d'auto-incompatibilité sporophytique, et ce type de données reste rare dans la littérature. Il serait donc important de poursuivre l'effort de caractérisation des relations de dominance chez *A. halleri* ainsi que dans d'autres espèces. L'identification des mécanismes moléculaires déterminant les relations de dominance entre allèles permettraient également une meilleure compréhension du phénomène d'évolution de la dominance.

Ensuite, il serait nécessaire de poursuivre l'identification de l'effet du fardeau génétique lié, notamment en étudiant des allèles S de plusieurs niveau de dominance pour confirmer notre hypothèse de lien entre la puissance du fardeau et la dominance de l'allèles S associé. Il serait également important, pour prouver l'existence du fardeau lié, d'identifier la nature et la localisation des mutations délétères liées. Le projet (porté par V. Castric) de séquençage complet du locus S pour cinq haplotypes d'*A. halleri* pourrait permettre d'identifier des gènes porteurs de ces mutations.

De la même manière que dans les systèmes GSI (Porcher & Lande, 2005), il nous semble également important d'étudier théoriquement la probabilité d'invasion d'un mutant autocompatible dans une population d'espèce à SSI, en prenant notamment en compte la dépression de consanguinité génomique et du fardeau génétique lié.

Enfin, notre étude de l'évolution de la dominance au locus S constitue une toute première approche mais ce phénomène étant apparu comme important, il est nécessaire de poursuivre sa caractérisation de manière plus approfondie. Pour identifier l'importance réelle de ce phénomène, il faudrait prendre en compte la dérive génétique par des simulations en population finie.

Références bibliographiques

Alberto (2004) Spatial genetic structure, neighbourhood size and clonal subrange in seagrass (*Cymodosa nodosa*) populations.

Alexander MP (1969) Differential staining of aborted and nonaborted pollen. *Stain technology* **44**, 117-122.

Bateman AJ (1952) Self-incompatibility systems in angiosperms. 1. Theory. *Heredity* **6**, 285-310.

Bechsgaard JS, Castric V, Charlesworth D, Vekemans X, Schierup MH (2006) The transition to self-compatibility in Arabidopsis thaliana and evolution within S-haplotypes over 10 Myr. *Molecular Biology and Evolution* **23**, 1741-1750.

Bergelson J, Kreitman M, Stahl EA, Tian D (2001) Evolutionary dynamics of plant R-genes. *Science* **292**, 2281–2285.

Billiard S, Castric V, Vekemans X (2007) A General Model to Explore Complex Dominance Patterns in Plant Sporophytic Self-Incompatibility Systems. *Genetics* **175**, 1351–1369.

Brennan AC, Harris SA, Hiscock SJ (2006) The population genetics of sporophytic self-incompatibility in Senecio squalidus L. (Asteraceae): the number, frequency, and dominance interactions of S alleles across its british range. *Evolution* **60**, 213–224.

Casselman AL, Vrebalov J, Conner JA, *et al.* (2000) Determining the Physical Limits of the Brassica S Locus by Recombinational Analysis. *The Plant Cell* **12**, 23–33.

Castric V, Vekemans X (2004) Plant self incompatibility in natural populations : a critical assessment of recent theoritical and empirical advance. *Molecular Ecology* **13**, 2873–2889.

Castric V, Vekemans X (2007) Evolution under strong balancing selection: how many codons determine specificity at the female self-incompatibility gene SRK in Brassicaceae? *BMC Evolutionary Biology* **in press**.

Charlesworth B, Charlesworth D (2000) The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society* **355**.

Charlesworth D, Bartolome C, Schierup MH, Mable BK (2003a) Haplotype structure of the stigmatic self-incompatibility gene in natural populations of Arabidopsis lyrata. *Molecular Biology and Evolution* **20**, 1741-1753.

Charlesworth D, Mable BK, Schierup MH, Bartolome C, Awadalla P (2003b) Diversity and linkage of genes in the self-incompatibility gene family in Arabidopsis lyrata. *Genetics* **164**, 1519-1535.

Charlesworth D, Vekemans X, Castric V, Glemin S (2005) Plant self-incompatibility systems: a molecular evolutionary perspective. *New Phytologist* **168**, 61-69.

Chookajorn T, Kachroo A, Ripoll DR, Clark AG, Nasrallah JB (2004) Specificity determinants and diversification of the Brassica self-incompatibility pollen ligand. *Proceedings of the National Academy of Science* **101**, 911–917.

Clark AG, Kao TH (1991) Excess Nonsynonymous Substitution at Shared Polymorphic Sites among Self-Incompatibility Alleles of Solanaceae. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 9823-9827.

Clauss MJ, Koch MA (2006) Poorly known relatives of Arabidopsis thaliana. *Trends in Plant Science* **11**, 449-459.

de Nettancourt D (2001) *Incompatibility and Incongruity in Wild and Cultivated Plants* Springer-Verlag, Berlin.

Fisher RA (1928a) The possible modification of the response of the wild type to recurrent mutations. *American Naturalist* **62**, 115-126.

Fisher RA (1928b) Two further notes on the origin of dominance. *American Naturalist* **62**, 571-574.

Fisher RA (1930) *The genetical theory of natural selection* Clarendon Press, Oxford.

Fujimoto R, Sugimura T, Fukai E, Nishio T (2006) Suppression of gene expression of a recessive *Sp11/SCR* allele by an untranscribed *SP11/SCR* allele in *Brassica* self-incompatibility. *Plant Molecular Biology* **61**, 577-587.

Glémin S, Gaude T, Guillemin ML, *et al.* (2005) Balancing Selection in the Wild: Testing Population Genetics Theory of Self-Incompatibility in the Rare Species *Brassica insularis*. *Genetics* **171**, 279-289.

Hasselmann M, Beye M (2004) Signatures of selection among sex-determining alleles of the honey bee. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 4888-4893.

Hatakeyama K, Takasaki T, Suzuki G, *et al.* (2001) The S receptor kinase gene determines dominance relationships in stigma expression of selfincompatibility in Brassica. *Plant Journal* **26**, 69–76.

Hedrick PW (2002) Pathogen resistance and genetic variation at MHC loci. *Evolution* **56**, 1902-1908.

Hiscock SJ, McInnis SM (2003) Pollen recognition and rejection during the sporophytic self-incompatibility response : Brassica and beyond. *Trends in plant science* **8**, 606-613.

Igic B, Bohs L, Kohn JR (2004) Historical inferences of the self-incompatibility locus. *New Phytologist* **161**, 97-105.

Ioerger TR, Clark AG, Kao T-H (1990) Polymorphism at the selfincompatibility locus in Solanaceae predates speciation. *Proceeding of the National Academy of Science of the United States of America* **87**, 9732–9735.

Iwanaga A, Sasaki A (2004) Evolution of hierachical cytoplasmic inheritance in the plamodal slime mold Physarum polycephalum. *Evolution* **58**, 710–722.

Jarne P, Charlesworth D (1993) The evolution of selfing rate in functionally hermaphrodite plants and animals. *Annual Review of Ecology, Evolution, and Systematics* **24**, 441-466.

Kacser H, Burns JA (1981) The molecular basis of dominance. *Genetics* **97**, 639-666.

Kamau E, Charlesworth D (2005) Balancing Selection and Low Recombination Affect Diversity near the Self-Incompatibility Loci of the Plant Arabidopsis lyrata. *Current Biology* **15**, 1773–1778.

Kawabe A, Hansson B, Forrest A, Hagenblad J, Charlesworth D (2006) Comparative gene mapping in Arabidopsis lyrata chromosomes 6 and 7 and A. thaliana chromosome IV: evolutionary history, rearrangements and local recombination rates. *Genetical Research Cambridge* **88**, 45-56.

Kronstad JW, Staben C (1997) Mating type in filamentous fungi. *Annual review of genetics* **31**, 245–276.

Naithani S, Chookajorn T, Ripoll DR, Nasrallah JB (2007) Structural modules for receptor dimerization in the S-locus receptor kinase extracellular domain. *Proceedings of The National Academy of Sciences* **104**, 12211–12216.

Nou IS, Watanabe M, Isogai A, Hinata K (1993) Comparison of S-alleles and S-glycoproteins between two wild populations of *Brassica campestris* in Turkey and Japan. *Sexual Plant Reproduction* **6**, 79–86.

Nuismer SL, Otto SP (2005) Host-parasite interactions and the evolution of gene expression. *Plos Biology* **3**, 1283-1288.

Ockendon DJ (2000) The S-allele collection of Brassica oleracea. *Acta Horticulturae* **539**, 25–30.

Oddou-Muratorio S, Klein EK, Austerlitz F (2005) Pollen flow in the wildservice tree, Sorbus torminalis (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent-offspring analysis. *Molecular Ecology* **14**, 4441-4452.

Otto SP, Bourguet D (1999) Balanced polymorphisms and the evolution of dominance. *American Naturalist* **153**, 561-574.

Porcher E, Lande R (2005) The evolution of self-fertilization and inbreeding depression under pollen discounting and pollen limitation. *Journal of Evolutionary Biology* **18**, 497-508.

Prigoda NL, Nassuth A, Mable BK (2005) Phenotypic and genotypic expression of self-incompatibility haplotypes in Arabidopsis lyrata suggests unique origin of alleles in different dominance classes. *Molecular Biology and Evolution* **22**, 1609-1620.

Richards A (1997) *Plant breeding systems, 2nd ed.* Chapman and Hall,, Cambridge, UK.

Sakamoto K, Kusaba M, Nishio T (1998) Polymorphism of the S -locus glycoprotein gene (SLG) and the S -locus related gene (SLR1) in Raphanus sativus L. and self-incompatible ornamental plants in the Brassicaceae. *Molecular and General Genetics* **258**, 397-403.

Schierup MH, Bechsgaard JS, Nielsen LH, Christiansen FB (2006) Selection at Work in Self-Incompatible Arabidopsis lyrata: Mating Patterns in a Natural Population. *Genetics* **172**, 477–484.

Schierup MH, Vekemans X, Christiansen FB (1997) Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics* **147**, 835-846.

Schueler S, Tusch A, Scholz F (2006) Comparative analysis of the within-population genetic structure in wild cherry (Prunus avium L.) at the self-incompatibility locus and nuclear microsatellites. *Molecular Ecology* **15**, 3231-3243.

Shiba H, Iwano M, Entani T, *et al.* (2002) The dominance of allelle controlling self-incompatibility in Brassica pollen is regulated at the RNA level. *The plant cell* **14**, 491-504.

Shiba H, Kakizaki T, Iwano M, *et al.* (2006) Dominance relationships between self-incompatibility alleles controlled by DNA methylation. *Nature genetics*.

Stone JL (2004) Sheltered load associated with S-alleles in *Solanum carolinense*. *Heredity* **92**, 335-342.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815.

Uyenoyama MK (1997) Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics* **147**, 1389-1400.

Uyenoyama MK (2000) Evolutionary dynamics of self-incompatibility alleles in Brassica. *Genetics* **156**, 351-359.

Uyenoyama MK (2005) Evolution under tight linkage to mating type. *New Phytologist* **165**, 63–70.

Veitia RA (2006) *The Biology of Genetic Dominance* Landes Bioscience.

Vekemans X, Schierup MH, Christiansen FB (1998) Mate Availability and Fecundity Selection in Multi-Allelic Self- Incompatibility Systems in Plants. *Evolution* **52**, 19-29.

Vekemans X, Slatkin M (1994) Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* **137**, 1157-1165.

Wagenius S, Lonsdorf E, Neuhauser C (2007) Patch aging and the S-Allee effect: breeding system effects on the demographic response of plants to habitat fragmentation. *The american naturalist* **169**, 383-397.

Wiuf C, Zhao K, Innan H, Nordborg M (2004) The probability and chromosomal extent of trans-specific polymorphism. *Genetics* **168**, 2363–2372.

Wright S (1929) Fisher's theory of dominance. *American Naturalist* **63**, 274-279.

Wright S (1934) Physiological and evolutionary theories of dominance. *American Naturalist* **63**, 25-53.

Wright S (1939) The distribution of self sterility alleles in populations. *Genetics* **24**, 538-552.

Résumé : L'auto-incompatibilité est un système empêchant l'autofécondation et la reproduction entre apparentés : chez les espèces auto-incompatibles, chaque individus porte une spécificité déterminée par partir du locus d'auto-incompatibilité (locus S), les pistils ne peuvent être fécondée qu'avec du pollen portant une spécificité différente de la leur. Le locus S est donc soumis à une sélection fréquence-dépendante, puisque les pollens portant une spécificité rare ont un nombre de partenaires plus élevés dans la population. Chez les *Brassicaceae*, le système d'auto-incompatibilité est « sporophytique », c'est-à-dire que les spécificités portées par le pollen et le pistil sont codées par le génotype au locus S des parents. Des relations de dominance entre allèles du locus S déterminent donc les spécificités exprimées dans le pollen et pistil. D'autre part, le locus S ayant une fort hétérozygotie et étant situé dans une région à faible taux de recombinaison, il peut être associé à des mutations délétères. Au cours de cette thèse, nous nous sommes donc intéressés aux forces évolutives agissant au locus d'auto-incompatibilité chez *Arabidopsis halleri* en suivant quatre axes principaux (1) Quels sont les patrons de reproduction en population naturelle ? (2) Quelle est l'influence de la sélection fréquence-dépendante, des relations de dominance et de la dérive sur le polymorphisme au locus S ? (3) Existe-il un fardeau génétique lié au locus S ? Dépend t il du niveau de dominance de l'allèle S associé ? (4) La dominance peut elle évoluer au locus S ?

Abstract : Self-incompatibility is a system avoiding selfing and reproduction among relatives : in SI species, each individual carries a specificity encoded by the self-incompatibility locus (S-locus) and pistil can only be fertilized by pollen carrying a different specificity. The S-locus is under frequency-dependent selection because pollen carrying a rare specificity have a greater number of compatible partners in a population. In Brassicaceae, the SI is sporophytic, *i. e.* specificity carried by the pollen and the pistil are encoded by the diploid genome of the parent. Dominance relationships among S-alleles thus determined the specificities expressed in the pollen and in the pistil. Since the S-locus has a high heterozygosity and a low recombination rate, it could be associated with some deleterious mutation. In this PhD thesis, we aimed at characterizing the evolutionary process acting at the S-locus in *Arabidopsis halleri*, following 4 main questions: (1) What are the mating pattern in natural population (2) What is the influence of frequency-dependant selection, dominance relationships and genetic drift on the S-locus? (3) Is there a sheltered genetic load associated with the S-locus? Does this genetic load depend on the dominance of the linked S-allele? (4) Is dominance at the s-locus likely to evolve?