

# Evolution de l'auto-incompatibilité : Modélisation des conditions de maintien et de diversification en populations finies

## THÈSE

soutenue publiquement le 9 Décembre 2011

pour l'obtention du

**Doctorat de l'Université de Lille 1 – Sciences et  
Technologies**

(spécialité Biologie Évolutive et Écologie)

par

**Camille Gervais**

### Composition du jury

<i>Rapporteurs :</i>	Christine DILLMANN, Professeur	UMR Génétique végétale, Université Paris-Sud
	Emmanuelle PORCHER, Maître de conférence	CERSP, MNHN
<i>Examineurs :</i>	Thomas LENORMAND, Directeur de Recherche CNRS	CEFE/CNRS, Montpellier
	Hélène MORLON, Chargée de recherche CNRS	CMAP, Ecole Polytechnique
	Xavier VEKEMANS, Professeur	GEPV, Université de Lille 1
	Joël CUGUEN, Professeur	GEPV, Université de Lille 1
<i>Directeurs :</i>	Vincent CASTRIC, Chargé de recherche CNRS	GEPV, Université de Lille 1
	Sylvain BILLIARD, Maître de conférence	GEPV, Université de Lille 1





Ris, tout le monde rira avec toi. Pleure, tu seras seul à pleurer.  
*Old Boy* (2003)

## Remerciements

Traditionnellement, on écrit ses remerciements de thèse dans les semaines, voire les mois, qui suivent la soutenance. Pourquoi ? Parce qu'il est tout bonnement impossible de remercier qui que ce soit en fin de thèse, ou même dans les jours qui suivent la délivrance. A ce moment-là, vous en voulez au monde entier. Épuisé, décomposé, vous avez plutôt envie de hurler : "Mais pourquoi Diable personne n'était là pour m'empêcher de m'embarquer dans ce calvaire?". Mais le temps passe, les rancœurs s'atténuent, et un beau jour vous vous dites que finalement ce n'est pas si mal d'en être arrivé là. Et alors seulement, vous pensez à toutes les personnes qui vous ont permis d'accomplir cet ouvrage. Ceux qui vous ont tendu la main, au tout début. Ceux qui vous ont soutenu dans les moments difficiles. Ceux qui vous ont donné des coups de pied au cul quand vous en aviez besoin...

Je tiens donc à remercier :

Sylvain Billiard et Vincent Castric, qui m'ont d'abord initiée à la recherche, durant mes stages de M1 et M2, et qui m'ont accordé suffisamment de confiance pour accepter de m'encadrer durant ces trois années de thèse. Xavier Vekemans également qui, sans m'avoir directement encadrée, a grandement contribué à la supervision de mes travaux.

Les membres de mon jury, Emmanuelle Porcher, Christine Dillmann, Thomas Lenormand, Hélène Morlon et Joël Cuguen, pour le temps consacré à la lecture de mon manuscrit et à l'évaluation de mon travail.

Adrienne Ressayre et Denis Roze pour leur aide précieuse dans l'élaboration des modèles analytiques et de simulation, ainsi que pour la rédaction des articles.

L'équipe "auto-incompatibilité" et le GEPV dans son ensemble qui m'ont accueillie (recueillie ?) depuis mon stage de M1 en 2007. En particulier Pauline, Diala, Romain, Marilyne, Marie, Manue et Mathilde, avec qui j'ai partagé (rayez les mentions inutiles) bières, cigarettes, raclette, cinéma asiatique, bureau, déprime de fin de thèse, coups de gueule...

Un merci tout particulier à la bande des Thésards, JB, Isa, Aude, Meriem, Benon, Camillo et Lucy, ainsi qu'à Fafa et Jeff. Compagnons de galère, compagnons de beuverie, soutiens inébranlables. Ces trois années auraient eu un goût bien amer sans vous. D'ailleurs je ne serais probablement pas allée au bout de ces trois années sans vous. Merci pour tout !

Ma famille, qui supporte depuis plus d'un quart de siècle mon caractère à bouloches. Mes parents, évidemment, sans qui je ne serais pas là aujourd'hui (haha! Je ne pouvais pas ne pas la faire celle-là), Mamie et son fichu caractère, mes frangins, Thomas, Clément et Antoine, mon beauf' Mickael également, Tonton Gilles et Tata Valérie, Georges-Alain et Bernadette, Jean-Pierre et Danièle.

Mes vieux amis, les Bidasses et Sam, pour tous les bons moments passés depuis plus de 10 ans. Ainsi que mes nouveaux amis, Jenny et Olivier, pour tous les bons moments passés depuis moins d'un an.

Mes partenaires et instructeurs de ninjutsu, amis également, Christophe, François, Nicolas, Dominique, Fabien, Tristan (et Caro) et Thibaut, qui m'apportent beaucoup martialement (évidemment), mais aussi personnellement, tant dans la pratique que durant les afters et autres activités annexes. Et aussi simplement parce que rien ne vaut un bon shuto pour se défouler après une journée de boulot.

M. Poteau, mon instit' de CE2, qui m'a donné goût à l'école, la vraie, celle du savoir et des connaissances.

Les Renard, pour leur extrême gentillesse, tout simplement.

Yrsüth pour tout le réconfort qu'elle m'apporte quotidiennement depuis nos trois ans de vie commune.

Andy Lau. Pourquoi? Parce que.

Last but not least, merci à La Chouffe qui m'a permis de tenir le coup jusqu'au bout en me faisant régulièrement oublier mes soucis quant aux mystères de l'évolution des plantes-escargots.



# Table des matières

<b>Introduction</b>	<b>5</b>
1 L'auto-incompatibilité . . . . .	7
1.1 Mécanismes de l'auto-incompatibilité . . . . .	7
1.2 Les questions évolutives liées à l'auto-incompatibilité . . . . .	11
2 Maintien de l'auto-incompatibilité . . . . .	13
2.1 Dépression de consanguinité fixe . . . . .	13
2.2 Dépression de consanguinité dynamique . . . . .	17
2.3 Conclusion partielle . . . . .	18
3 Diversification du système d'auto-incompatibilité . . . . .	19
3.1 Intermédiaire à double spécificité . . . . .	19
3.2 Diversification graduelle . . . . .	21
3.3 Intermédiaire auto-compatible . . . . .	21
3.4 Conclusion partielle . . . . .	24
4 Importance de la dérive . . . . .	25
<b>1 Présentation et comparaison des modèles de simulations</b>	<b>27</b>
1.1 Structure générale commune aux deux modèles . . . . .	27
1.2 Structure des chromosomes . . . . .	28
1.3 Mutation . . . . .	28
1.4 Définition de l'auto-compatibilité . . . . .	29
1.5 Dépression de consanguinité . . . . .	29
1.6 Autofécondation . . . . .	30
1.7 Recombinaison . . . . .	30

<b>2</b>	<b>Maintenance of GSI with evolution of inbreeding depression</b>	<b>31</b>
2.1	Introduction . . . . .	32
2.2	Model . . . . .	33
2.2.1	Mutation . . . . .	34
2.2.2	Reproduction and selection . . . . .	34
2.2.3	Gamete production and recombination . . . . .	35
2.2.4	Simulation methods and parameter values . . . . .	35
2.3	Results . . . . .	37
2.3.1	Dynamics of SI and inbreeding depression . . . . .	37
2.3.2	Conditions for maintenance of SI . . . . .	38
2.3.3	Effect of the rate of deleterious mutation . . . . .	39
2.3.4	Effect of recombination and drift . . . . .	39
2.3.5	Effect of the properties of mutations . . . . .	42
2.4	Discussion . . . . .	44
2.4.1	Conditions for maintenance of SI . . . . .	44
2.4.2	Population size and number of SI alleles . . . . .	44
2.4.3	Recombination rate . . . . .	45
2.4.4	Properties of mutations . . . . .	45
2.4.5	Polymorphism . . . . .	46
2.4.6	Conclusion . . . . .	46
2.5	Appendix . . . . .	47
<b>3</b>	<b>Origin and diversification dynamics of self-incompatibility haplo-</b>	<b>49</b>
	<b>types</b>	
3.1	Introduction . . . . .	50
3.2	Models . . . . .	52
3.2.1	Infinite population model . . . . .	53
3.2.2	Finite populations simulation . . . . .	55
3.3	Results . . . . .	58
3.3.1	Infinite populations . . . . .	58
3.3.2	Finite population simulations . . . . .	60
3.4	Discussion . . . . .	66
3.4.1	Importance of genetic drift and recurrent mutations. . . . .	66



3.4.2	The role of inbreeding depression. . . . .	66
3.4.3	Dynamics of diversification. . . . .	67
3.4.4	S-allele numbers and the shape of phylogenies. . . . .	67
3.4.5	The mutation process. . . . .	69
3.5	Appendix . . . . .	70
3.6	Supplemental . . . . .	74
<b>4</b>	<b>Discussion</b>	<b>77</b>
	<b>Bibliographie</b>	<b>83</b>

*Table des matières*

# Introduction

Le groupe des Angiospermes est le plus répandu parmi les végétaux terrestres. Il présente une grande variété de systèmes de reproduction, allant de la dioécie à l'hermaphrodisme. Parmi ces systèmes, l'hermaphrodisme est le plus fréquent, puisqu'il représente entre 72 et 90% des Angiospermes (Charlesworth, 2002 ; Richards, 1986). On trouve également une importante diversité de systèmes de reproductions au sein même des hermaphrodites en terme de taux d'autofécondation, allant de l'autogamie à l'allogamie stricte, en passant par des systèmes mixtes (Goodwillie et al., 2005). L'autofécondation présente un grand avantage à l'échelle de l'individu, car si on considère la production d'une graine, une plante mère autogame peut lui transmettre deux copies de son génome (par les voies mâle et femelle) alors qu'une plante allogame ne peut en transmettre qu'une (par la voie femelle). C'est ce que l'on appelle l'avantage automatique de l'autofécondation (Fisher, 1941). Elle présente également un important avantage écologique en fournissant aux individus l'assurance d'une reproduction en cas d'absence de partenaire, ce qui permet notamment de coloniser de nouveaux milieux (Pannell and Barrett, 1998). Malgré ces avantages, de nombreuses espèces ont développé des systèmes permettant d'éviter ou de limiter l'autofécondation par un éloignement spatial (monoécie, herchogamie) ou temporel (protandrie, protogynie) des fonctions mâles et femelles, ou par un système de reconnaissance de soi (l'auto-incompatibilité). Parmi ces diverses stratégies, l'auto-incompatibilité est la plus répandue puisqu'on l'observe chez environ la moitié des espèces d'Angiospermes (Igic and Kohn, 2006).

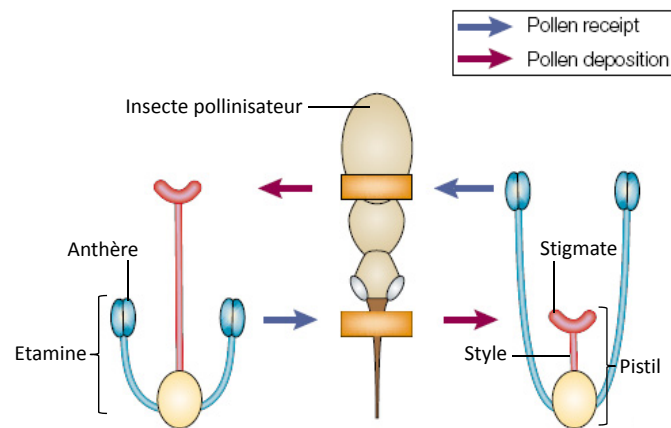


FIGURE 1 – Mécanisme de l’auto-incompatibilité hétéromorphe. Cas de la distylie. Le pollen d’un morphe particulier est déposé sur une partie de l’insecte pollinisateur qui n’entrera en contact qu’avec les stigmates de l’autre morphe. Les fécondations ne peuvent donc avoir lieu qu’entre morphes distincts. D’après Barrett (2002).

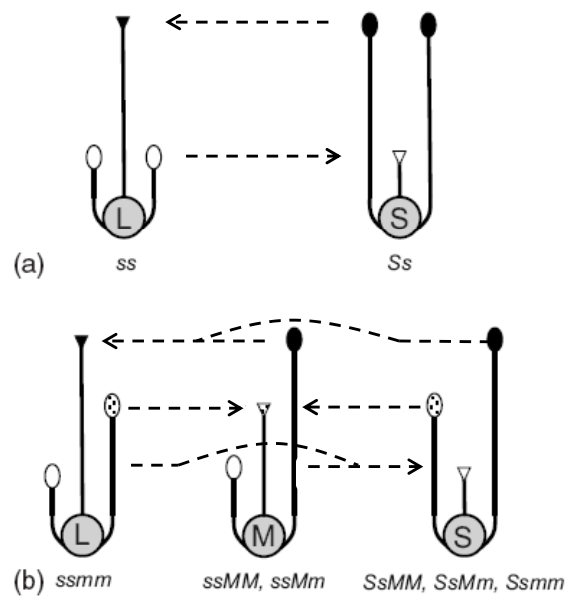


FIGURE 2 – Phénotypes et génotypes associés chez les espèces hétéromorphes distyles (a) et tristyles (b). L, M et S correspondent respectivement aux morphes longistyles, mésostyles et brévistyles. Les flèches indiquent les croisements compatibles, entre anthères et stigmates de même niveau. Les allèles *S* et *M* sont dominants sur les allèles *s* et *m*, respectivement. D’après Barrett and Shore (2008).

# 1 L'auto-incompatibilité

## 1.1 Mécanismes de l'auto-incompatibilité

L'auto-incompatibilité est un mécanisme génétique permettant aux plantes hermaphrodites d'éviter l'autofécondation et de limiter les croisements entre individus apparentés. Elle est basée sur un système de reconnaissance et de rejet du pollen par le pistil lorsque ceux-ci expriment la même spécificité, c'est-à-dire un phénotype d'auto-incompatibilité commun (de Nettancourt, 2001).

Lorsque ce mécanisme est associé à des différences morphologiques entre individus, on parle d'auto-incompatibilité hétéromorphe. Les individus diffèrent alors principalement au niveau de la longueur de leurs styles et étamines (Barrett et al., 2000). L'hétérostylie a évolué indépendamment dans au moins 28 familles d'Angiospermes (Barrett and Shore, 2008), ce qui en fait le mécanisme d'auto-incompatibilité le plus fréquent. La plupart des espèces présentant ce type de mécanisme sont distyles, c'est-à-dire qu'on y trouve deux morphes : des individus à long styles et courtes étamines, et des individus à courts styles et longues étamines. Certaines espèces sont tristyles et présentent trois longueurs de styles et étamines : courte, moyenne et longue. On trouve alors trois morphes distincts : les individus portent des étamines de deux longueurs différentes, et le troisième niveau est occupé par le style (par exemple, styles longs et étamines courtes et moyennes). La morphologie des fleurs limite fortement l'auto-pollinisation en raison de la séparation spatiale entre pollen et stigmate, ainsi que la pollinisation entre individus du même morphe. En effet, les insectes pollinisateurs fixent le pollen sur différentes parties de leur corps en fonction de la longueur des étamines, et le déposent donc au niveau des stigmates situés sur des styles de même longueur (Figure 1). De plus, à ce mécanisme morphologique est très souvent associé un système de reconnaissance génétique rendant impossible les fécondations entre individus de même morphe. Cette reconnaissance et les différences morphologiques sont contrôlées par un locus diallélique (S) dans le cas de la distylie, et par deux loci dialléliques liés (S et M) dans le cas de la tristylie (Figure 2, Barrett and Shore, 2008).

Au contraire, lorsque les individus portant des spécificités différentes ne peuvent être différenciés morphologiquement, on parle d'auto-incompatibilité homomorphe. Dans la plupart de ces familles (22 des 23 familles dans lesquelles a été identifiée l'auto-incompatibilité homomorphe; Hiscock, 2000 ; Hiscock and Tabah, 2003), la reconnaissance du pollen par le pistil est contrôlée par un seul locus nommé le locus S. Ce locus

## Introduction

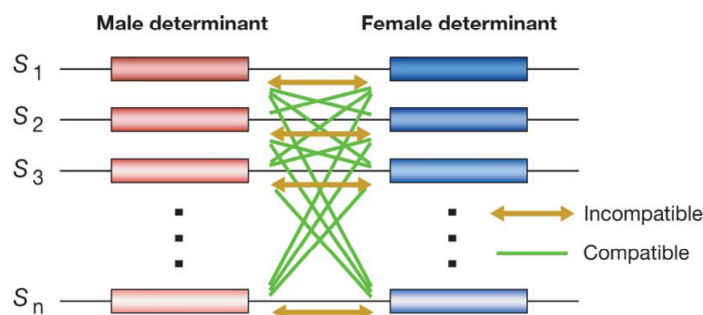


FIGURE 3 – Représentation schématique du locus S dans le cas de l’auto-incompatibilité homomorphe. Le locus S est généralement constitué de deux gènes liés, l’un codant pour la spécificité mâle (gène pollen, à gauche), l’autre pour la spécificité femelle (gène pistil, à droite). Chaque combinaison de spécificités constitue un haplotype  $S_i$ . Les croisements ne sont compatibles (traits verts) que si les gènes pollen et pistil expriment des spécificités différentes, c’est-à-dire s’ils proviennent d’haplotypes différents. D’après Takayama and Isogai (2005).

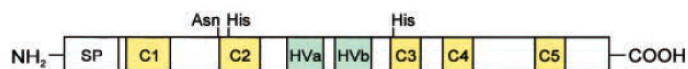


FIGURE 4 – Représentation schématique d’une S-Rnase (gène pistil) chez les Solanaceae. Les différentes régions conservées (en jaune) et hypervariables (en vert) sont représentées. D’après Silva and Goring (2001).

présente typiquement un grand nombre d’allèles<sup>1</sup>, en raison de la sélection fréquence-dépendante négative à laquelle ceux-ci sont soumis (Wright, 1939). En effet, les individus porteurs d’une spécificité rare ont accès à un plus grand nombre de partenaires sexuels que les individus porteurs de spécificités courantes. Le locus S est généralement composé de deux gènes liés, l’un s’exprimant dans le pollen, l’autre dans le pistil (Figure 3, Takayama and Isogai, 2005). Ces gènes et les protéines associées ont été caractérisés chez certaines espèces et ont révélé l’existence de régions conservées supposées contenir les codons impliqués dans le maintien de la structure et de la fonction même des protéines, ainsi que des régions hypervariables, probablement impliquées dans la détermination des spécificités (Figure 4, par exemple Ikeda et al., 2004 ; Ioerger et al., 1991 ; Ishimizu et al., 1998 ; Nishio and Kusaba, 2000). On distingue deux catégories

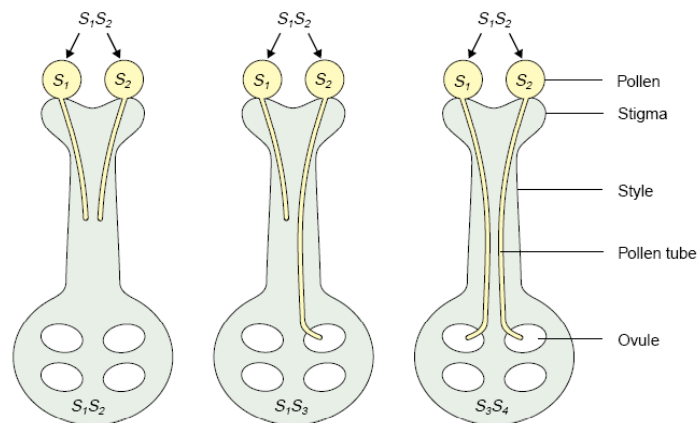
1. Les termes allèle S, spécificité, et haplotype S sont souvent employés indifféremment. Cela n’est pas gênant lorsque l’on considère le locus S comme un unique gène. En revanche, lorsqu’on distingue les deux gènes pollen et pistil, il est important de les différencier. Un haplotype S désigne l’ensemble de ces deux gènes, chacun exprimant une spécificité particulière. Dans le cadre de cette thèse, les termes spécificités et allèles sont employés indifféremment.

d'auto-incompatibilité homomorphe selon la détermination génétique de la spécificité pollen : l'auto-incompatibilité gamétophytique et l'auto-incompatibilité sporophytique.

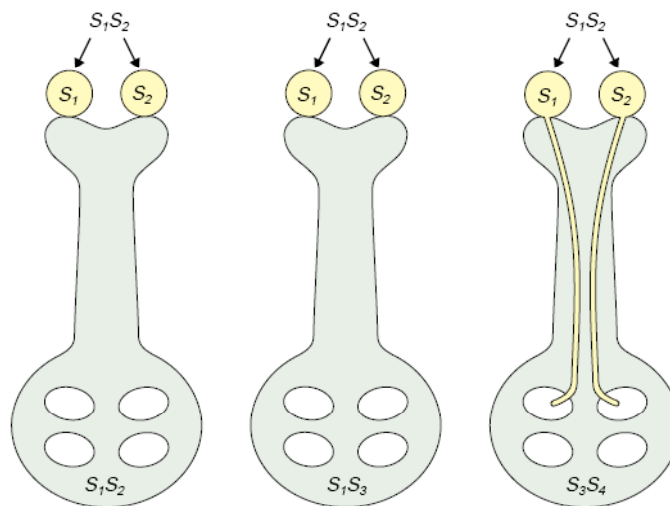
L'auto-incompatibilité gamétophytique est le système d'auto-incompatibilité homomorphe le plus répandu. Elle a été identifiée dans 17 familles d'Angiospermes (Franklin-Tong and Franklin, 2003). Dans ce système, le génotype d'auto-incompatibilité du pollen dépend de son propre génotype haploïde. Un croisement est compatible si la spécificité exprimée dans le pollen est différente des deux spécificités exprimées de façon codominante dans le pistil (Figure 5a). En raison de l'équivalence sélective entre les spécificités, on s'attend à ce que, à l'équilibre, toutes les spécificités soient présentes à la même fréquence sous l'effet de la sélection fréquence-dépendante négative (Wright, 1939).

Deux types de mécanismes de reconnaissance ont été caractérisés au niveau moléculaire chez les espèces gamétophytiques. Chez les Solanaceae, les Rosaceae et les Plantaginaceae, la protéine exprimée dans le pistil est une ribonucléase appelée S-RNase, et le gène déterminant la spécificité du pollen est un gène F-box, appelé SFB chez les Rosaceae et SLF chez les deux autres familles. La réaction d'incompatibilité a lieu dans le style où est exprimée la S-RNase et provoque l'arrêt de la croissance du tube pollinique (Takayama and Isogai, 2005). Chez les Papaveraceae, les gènes pistil et pollen sont respectivement *PrsS* et *PrpS* (Wheeler et al., 2009). *PrsS* est exprimée au niveau du stigmate. La réaction d'incompatibilité a donc lieu à la surface de celui-ci et inhibe la croissance du tube pollinique (Franklin-Tong and Franklin, 2003). Un troisième système d'auto-incompatibilité gamétophytique a été mis en évidence chez les Poaceae. Chez ces espèces, la réaction d'incompatibilité qui a lieu à la surface du stigmate est contrôlée par deux locus indépendants : le locus S et le locus Z (Langridge and Baumann, 2008).

L'auto-incompatibilité sporophytique est plus rare et n'a été observée que dans 6 familles d'Angiospermes (Hiscock and Tabah, 2003). Dans le cas de l'auto-incompatibilité sporophytique, le phénotype d'auto-incompatibilité du pollen dépend du génotype diploïde de la plante paternelle. De plus, il existe des relations de dominance entre les différentes spécificités, en particulier dans le pollen, ce qui rend possible la création d'individus homozygotes pour les allèles récessifs (Figure 5b). Ce mécanisme a principalement été étudié chez les Brassicaceae. Dans cette famille, les spécificités pollen et pistil sont respectivement contrôlées par les gènes SCR (ou SP11) et SRK (Takayama and Isogai, 2005).



(a) Auto-incompatibilité gamétophytique



(b) Auto-incompatibilité sporophytique

**FIGURE 5** – Déterminisme génétique de l’auto-incompatibilité homomorphe. (a) Dans le cas de l’auto-incompatibilité gamétophytique, le phénotype d’incompatibilité du grain de pollen est déterminé par son propre génotype haploïde. Un croisement est compatible si la spécificité exprimée dans le pollen est différente des deux spécificités exprimées dans le pistil. (b) Dans le cas de l’auto-incompatibilité sporophytique, le phénotype d’incompatibilité du pollen est déterminé par le génotype diploïde de la plante paternelle. Dans cet exemple, les allèles S codant pour la spécificité du pollen sont codominants : les individus sont donc compatibles uniquement s’ils n’ont aucun allèle S en commun. D’après Hiscock (2002).



## 1.2 Les questions évolutives liées à l'auto-incompatibilité

### 1.2.1 Le maintien de l'auto-incompatibilité

Malgré les avantages de l'autofécondation en terme de transmission de génome et d'assurance de reproduction, l'auto-incompatibilité est apparue à plusieurs reprises de façon indépendante dans différentes familles (Allen and Hiscock, 2008), sous la forme de différents mécanismes (Takayama and Isogai, 2005), et s'est maintenue chez environ la moitié des Angiospermes. Néanmoins, les événements de transition de l'auto-incompatibilité vers l'auto-compatibilité sont très fréquents. Il a été montré que l'auto-incompatibilité gamétophytique à S-RNase est ancestrale chez la plupart des familles d'Eudicotylédones (Igic and Kohn, 2001 ; Steinbachs and Holsinger, 2002) mais a été perdue indépendamment à de nombreuses reprises au profit de l'auto-compatibilité (par exemple chez *Lycopersicon* Kondo et al., 2002), et ces transitions semblent irréversibles (Igic et al., 2004). Il en est de même pour le système d'auto-incompatibilité sporophytique, ancestral chez les Brassicaceae (Fobis-Loisy et al., 2004 ; Stebbins, 1957). Au sein même de certains genres, on trouve des espèces auto-compatibles et des espèces auto-incompatibles, voire même au sein d'une même espèce au niveau des populations. Par exemple, dans le genre *Arabidopsis*, *A. halleri* est auto-incompatible (Llaurens et al., 2008), *A. thaliana* est auto-compatibile (Shimizu et al., 2008), et *A. lyrata* présente à la fois des populations auto-compatibles et auto-incompatibles (Willi and Maeaettaenen, 2010). Ce type de variabilité inter-espèces et/ou inter-populations s'observe également chez *Capsella* (Foxe et al., 2009) et *Leavenworthia* (Busch, 2005). Ce patron de coexistence de l'auto-incompatibilité (ancestrale et donc maintenue depuis longtemps) avec de l'auto-compatibilité (apparue à de maintes reprises et sans réversion possible) soulève la question des conditions favorisant l'une ou l'autre de ces évolutions. Dans les populations naturelles, la perte de l'auto-incompatibilité est souvent associée à une réduction de la taille des populations (par exemple Busch et al., 2010) entraînant une diminution du nombre de partenaires sexuels, du nombre d'allèles d'auto-incompatibilité et/ou de la dépression de consanguinité, qui sont toutes des conditions a priori favorables à la rupture de l'auto-incompatibilité. D'un point de vue théorique, les conditions de maintien ou de perte de l'auto-incompatibilité ont principalement été étudiées en considérant la balance entre l'avantage automatique de l'auto-fécondation d'une part, et la dépression de consanguinité d'autre part, ainsi que l'effet du nombre d'allèles S présents dans les populations. Les effets de ces différents facteurs sur l'évolution de l'auto-incompatibilité

seront détaillés dans la suite de l'Introduction. D'autres mécanismes ont également été considérés, tels que la limitation en pollen ou l'existence d'un fardeau lié au locus S (Porcher and Lande, 2005).

### 1.2.2 La diversification au locus S

Les espèces auto-incompatibles présentent typiquement un grand nombre de spécificités. Le nombre d'allèles ségrégeant au locus S est souvent de plusieurs dizaines, et peut exceptionnellement aller jusqu'à près de 200 (Lawrence, 2000), ce qui fait du locus S l'un des locus les plus diversifiés du monde vivant. Le maintien de cette diversité s'explique facilement par la sélection fréquence-dépendante négative, les individus portant des allèles rares pouvant accéder à un plus grand nombre de partenaires sexuels que les individus portant des allèles fréquents (Wright, 1939). En conséquence, les allèles rares, et en particulier les nouveaux allèles, sont plus difficilement perdus par dérive

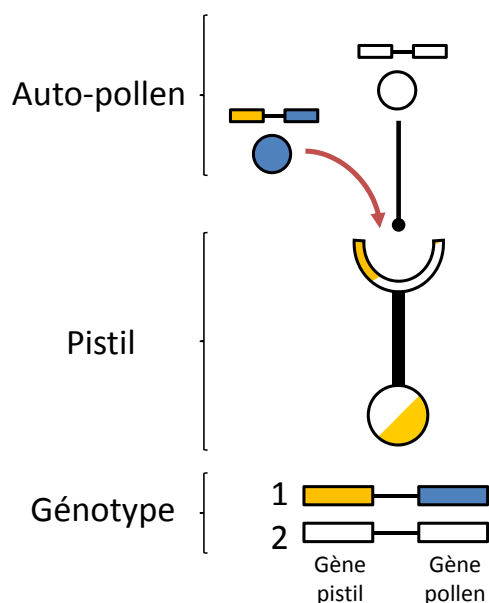


FIGURE 6 – Auto-compatibilité générée par la présence de deux spécificités distinctes aux gènes pollen et pistil. Dans le cas de l'haplotype 2, le phénotype exprimé par le pollen est reconnu par le pistil. Cet haplotype est donc auto-incompatible. Dans le cas de l'haplotype 1, la spécificité exprimée par le pollen (bleue) est différente de celle exprimée au niveau du pistil (jaune). Cet haplotype est donc auto-compatible. L'haplotype 2 exprime la même spécificité (blanche) aux gènes pollen et pistil.

qu'attendu pour des allèles neutres. En revanche, le mécanisme permettant l'évolution de nouveaux allèles est encore inconnu, notamment en raison de la structure bipartite du locus *S*. En effet, pour qu'un haplotype *S* soit fonctionnel, il faut que les gènes pollen et pistil portent la même spécificité. Si une nouvelle spécificité apparaît par mutation au niveau d'un seul des deux gènes, il n'y aura plus de reconnaissance entre les protéines pollen et pistil, entraînant une rupture de l'auto-incompatibilité : les individus portant cet haplotype seront auto-compatibles (Figure 6). L'apparition d'un nouvel haplotype nécessite donc une coévolution stricte entre les deux gènes du locus *S*.

## 2 **Maintien de l'auto-incompatibilité**

L'auto-incompatibilité, et en particulier l'auto-incompatibilité gamétophytique, est très répandue chez les Angiospermes, et ce malgré les avantages immédiats que procure l'autofécondation en terme de transmission de génome et d'assurance reproductive. L'allofécondation, et donc l'auto-incompatibilité, présente également différents avantages. Il a notamment été montré que les plantes allogames présentent généralement une diversité génétique plus élevée et un plus fort taux de recombinaison, ce qui facilite l'effet de la sélection en terme de fixation de mutations avantageuses ou d'élimination de mutations délétères (Wright and Barrett, 2010). Néanmoins, la dépression de consanguinité, c'est-à-dire la réduction de valeur sélective des individus issus d'autofécondation par rapport aux individus issus d'allofécondation, est généralement considéré comme le facteur principal permettant de contrecarrer l'avantage de l'autofécondation et expliquer le succès évolutif de l'auto-incompatibilité. La compréhension des mécanismes de maintien de l'auto-incompatibilité face à l'avantage automatique de l'autofécondation a donné lieu à une riche littérature, et trois modèles principaux ont été considérés successivement.

### 2.1 **Dépression de consanguinité fixe**

Charlesworth and Charlesworth (1979) ont étudié analytiquement les conditions d'invasion d'une population auto-incompatible par un haplotype auto-compatible en fonction de la dépression de consanguinité, du nombre d'haplotypes auto-incompatibles présents dans la population, et du taux d'autofécondation des individus auto-compatibles. Leur modèle considère une population infinie composée de  $n$  haplotypes auto-incompatibles (notés  $S_i$ ,  $1 \leq i \leq n$ ) en même fréquence, dans laquelle est introduit un haplotype

## Introduction

autocompatible (noté  $S_f$ ) en faible fréquence. Le pollen  $S_f$  est compatible avec tous les génotypes présents dans la population : les hétérozygotes  $S_iS_j$  ( $1 \leq i, j \leq n$ ) auto-incompatibles, les hétérozygotes  $S_iS_f$  partiellement auto-compatibles, et les homozygotes  $S_fS_f$  auto-compatibles qui peuvent être fécondés par n'importe quel pollen. Le taux d'autofécondation des homozygotes  $S_fS_f$  est donc supérieur au taux d'autofécondation des hétérozygotes  $S_iS_f$ . La dépression de consanguinité est fixe et représente la réduction de viabilité des graines issues d'autofécondation par rapport aux graines issues d'allofécondation.

Les auteurs ont montré que, lorsqu'on considère un nombre infini d'haplotypes auto-incompatibles et que le taux d'autofécondation des homozygotes  $S_fS_f$  est nul, la dépression de consanguinité doit être supérieure à  $2/3$  pour empêcher l'invasion de la population par  $S_f$ . Lorsqu'on augmente le taux d'autofécondation, cette valeur critique de dépression de consanguinité diminue mais reste néanmoins très élevée ( $\geq 0.61$ ). Les auteurs ont également montré que le nombre d'haplotypes  $S_i$  a une grande influence sur le maintien de l'auto-incompatibilité. Avec une dépression de consanguinité maximale (égale à 1, c'est-à-dire que tous les individus issus d'autofécondation meurent), il faut un minimum de 5 haplotypes  $S_i$  pour empêcher l'invasion de la population par  $S_f$ . Ce nombre augmente très rapidement quand on diminue la dépression de consanguinité ou le taux d'autofécondation.

Les auteurs ont également considéré le cas d'haplotypes auto-compatibles non fonctionnels uniquement au niveau de l'un des deux gènes (pollen ou pistil), l'autre gène portant la même spécificité que l'haplotype ancestral  $S_i$  présent dans la population. Un haplotype  $S'_1$  inactif dans le pollen permet de féconder tous les génotypes présents dans la population, mais les individus portant ce même haplotype ne peuvent être fécondés par du pollen portant la spécificité  $S_1$ . A l'inverse, l'haplotype  $S''_1$  inactif dans le style ne rejette aucune spécificité, mais ne permet pas de féconder les individus portant l'haplotype  $S_1$ .

Ce modèle a été repris et amélioré par Uyenoyama et al. (2001), qui ont notamment modélisé plus explicitement l'autofécondation. En effet, Charlesworth and Charlesworth (1979) considéraient deux paramètres distincts : le taux d'autofécondation des homozygotes auto-compatibles ( $S'_1S'_1$  par exemple) et le taux d'autofécondation plus faible des hétérozygotes partiellement auto-compatibles ( $S_iS_f$ ). Or, ces différents taux d'autofécondation dépendent directement des génotypes des individus et de la proportion de leur propre pollen (auto-pollen) qu'ils reçoivent. Uyenoyama et al. (2001) ont donc

considéré un unique paramètre qui est le taux d'autopollinisation. Les résultats de Uyenoyama et al. (2001) sont très semblables à ceux de Charlesworth and Charlesworth (1979). Dans le cas d'un haplotype non-fonctionnel au niveau du gène pollen, les résultats sont identiques à ceux observés dans le cas de l'haplotype  $S_f$ , si ce n'est que l'haplotype  $S'_1$  entraîne quasi-systématiquement l'exclusion de l'haplotype  $S_1$ , avec qui il partage la même spécificité pistil (Figure 7). Cette exclusion est due à la reconnaissance asymétrique entre les deux haplotypes,  $S'_1$  pouvant féconder tous les génotypes présents dans la population alors que  $S_1$  ne peut accéder aux individus portant les haplotypes  $S_1$  et/ou  $S'_1$ . Dans le cas d'un haplotype non-fonctionnel au niveau du gène pistil, la dépression de consanguinité critique permettant son exclusion ne dépend pas du nombre d'haplotypes fonctionnels  $S_i$ , mais uniquement du taux d'autofécondation (Figure 8). La dépression de consanguinité critique est toujours égale à  $2/3$  lorsque le taux d'autofécondation est nul, et diminue lorsque celui-ci augmente.

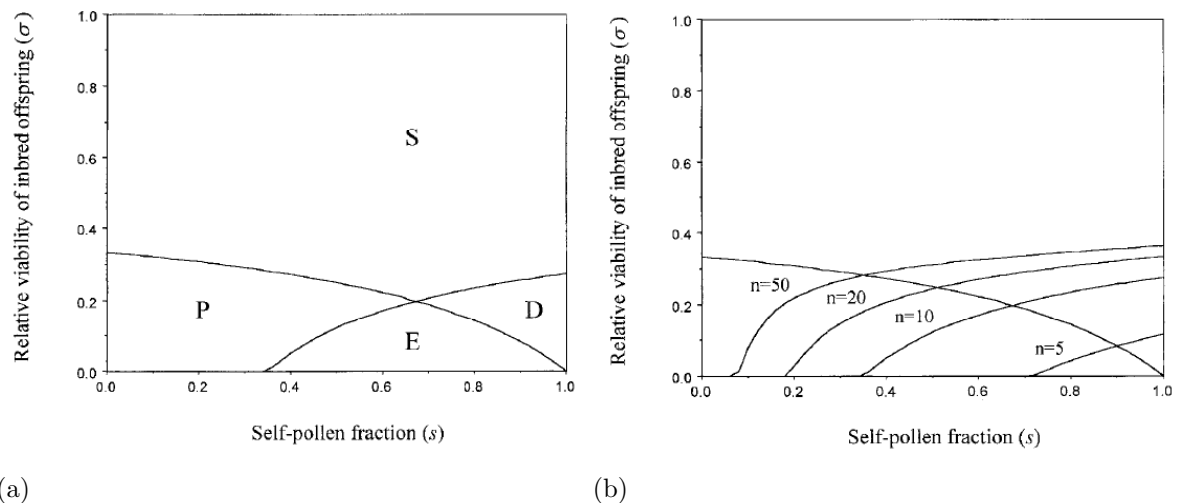


FIGURE 7 – Différents devenir possibles d'un mutant au gène pollen, en fonction de la survie relative des descendants d'autofécondation et du taux d'autopollinisation. (a) Nombre initial d'haplotypes  $S$  fonctionnels  $n = 10$ . Région S : le mutant envahit lorsqu'il est rare et se fixe, le système d'auto-incompatibilité est donc perdu. Régions D et E : le mutant est éliminé lorsqu'il est rare, le système est donc maintenu sans modification. Région P : le mutant augmente en fréquence lorsqu'il est rare, et diminue en fréquence lorsqu'il est fréquent. La population est donc maintenue dans un état de polymorphisme avec le mutant pollen et  $n - 1$  haplotypes fonctionnels, l'haplotype fonctionnel  $n$  dont dérive le mutant étant éliminé. (b) Effet du nombre initial d'haplotypes fonctionnels ( $n$ ). Plus  $n$  est grand, plus les conditions de maintien de l'auto-incompatibilité sont larges et plus les conditions de polymorphisme sont restreintes. D'après Uyenoyama et al. (2001).

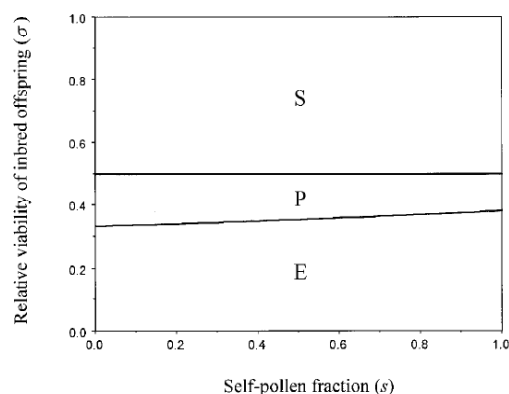


FIGURE 8 – Différents devenir possibles d’un mutant au gène pistil, en fonction de la survie relative des descendants d’autofécondation et du taux d’autopollinisation. Région S : le mutant envahit quand il est rare et se fixe, le système d’auto-incompatibilité est alors perdu. Région E : le mutant est éliminé quand il est rare, le système est alors maintenu sans modification. Région P : le mutant envahit quand il est rare et diminue en fréquence quand il est fréquent, la population contient alors à la fois le mutant pistil et les haplotypes fonctionnels. D’après Uyenoyama et al. (2001).

Le maintien de l’auto-incompatibilité nécessite donc une forte dépression de consanguinité ( $> 2/3$ ), et est facilité par un fort taux d’autofécondation et un nombre élevé d’haplotypes auto-incompatibles. Le principal avantage des auto-compatibles réside dans leur capacité à pouvoir s’autoféconder. Cet avantage disparaît si la valeur sélective des individus issus d’autofécondation est trop faible, et particulièrement lorsque le taux d’autofécondation est élevé. Outre l’autofécondation, les haplotypes auto-compatibles permettent de féconder tous les génotypes présents dans la population. Dans une population à l’équilibre, tous les haplotypes auto-incompatibles sont présents en même fréquence de l’ordre de  $1/n$  (si on néglige la faible fréquence de l’haplotypes auto-compatible). L’avantage d’un haplotype auto-compatible par rapport aux auto-incompatibles (compatibles avec tous les haplotypes sauf eux-mêmes) est donc de ce même ordre  $1/n$ . Cet avantage diminue donc lorsque le nombre d’haplotypes auto-incompatibles dans la population augmente. Ceci n’est pas valable dans le cas d’un haplotype non fonctionnel dans le pistil, car il est également reconnu par l’un des haplotypes auto-incompatibles présents dans la population ( $S_1''$  est reconnu par  $S_1$ ), ce qui explique que les conditions d’invasion d’un tel haplotype ne dépendent pas de  $n$ .

## 2.2 Dépression de consanguinité dynamique

La dépression de consanguinité est principalement due à des mutations récessives délétères qui s'expriment à l'état homozygote (Charlesworth and Willis, 2009). L'auto-incompatibilité empêche l'autofécondation et réduit donc fortement l'homozygotie. Les mutations causant la dépression de consanguinité sont donc peu exprimées et peuvent s'accumuler le long du génome. On s'attend donc à observer une forte dépression de consanguinité chez les espèces auto-incompatibles. En revanche, l'introduction d'un haplotype auto-compatible permet aux individus de s'autoféconder et augmente ainsi l'homozygotie. Les mutations délétères sont alors exprimées et peuvent être purgées, ce qui peut mener à une diminution de la dépression de consanguinité. Il est donc important de considérer la dynamique de la dépression de consanguinité lorsque l'on s'intéresse au maintien de l'auto-incompatibilité.

Porcher and Lande (2005) ont modifié le modèle de Charlesworth and Charlesworth (1979) en y introduisant de la dépression de consanguinité dynamique. Leur modèle considère une infinité de loci indépendants et non liés au locus S affectés par des mutations létales fortement récessives. Chaque mutation a lieu sur un nouveau locus, de façon à ce que les homozygotes létaux n'apparaissent que par autofécondation. La dépression de consanguinité est déterminée par la distribution du nombre de mutations à l'état hétérozygote par individu. De même que Charlesworth and Charlesworth (1979), leur modèle considère une population infinie composée de  $n$  haplotypes auto-incompatibles en mêmes fréquences et d'un haplotype auto-compatible initialement rare. Lorsque le nombre d'haplotypes auto-incompatibles est infini et que le taux d'autofécondation est nul, la population se comporte exactement comme attendu par Charlesworth and Charlesworth (1979), c'est-à-dire que l'haplotype auto-compatible envahit la population quand la dépression de consanguinité est inférieure à  $2/3$ . En revanche, contrairement au modèle précédent, l'augmentation du taux d'autofécondation a pour effet d'augmenter la dépression de consanguinité critique nécessaire au maintien de l'auto-incompatibilité (Figure 9), ce qui s'explique par une purge des mutations délétères plus importante en raison de l'augmentation de l'homozygotie. La diminution du nombre d'haplotypes auto-incompatibles a pour effet de considérablement réduire les conditions de maintien de l'auto-incompatibilité.

Les auteurs ont également considéré le cas de mutations délétères récessives complètement liées au locus S et allèle S-spécifiques, ce que l'on appelle le fardeau lié.

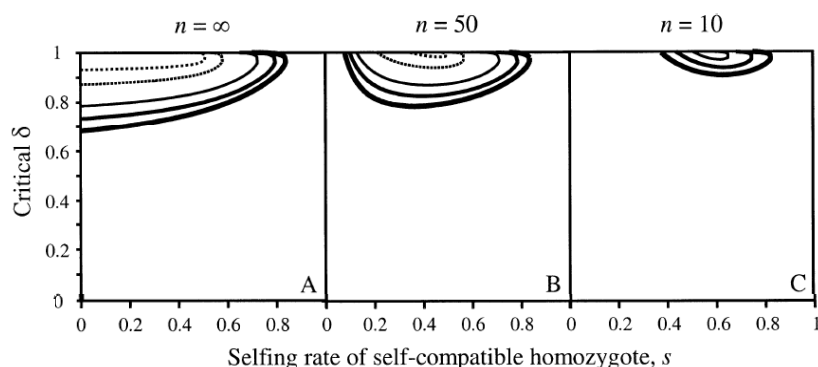


FIGURE 9 – Dépression de consanguinité critique ( $\delta$ ) au-dessus de laquelle un mutant auto-compatible ne peut envahir une population auto-incompatible, en fonction du taux d'autofécondation  $s$  et du nombre d'allèles d'auto-incompatibilité fonctionnels  $n$ . Les différentes courbes représentent différents conditions de limitation en pollen depuis l'absence de limitation (en bas, trait gras) à une forte limitation (en haut, trait pointillé). D'après Porcher and Lande (2005).

En effet, du fait de l'auto-incompatibilité, les allèles S ne se retrouvent jamais à l'état homozygote, de même que les mutations qui y sont associées. Ces mutations ne peuvent donc être purgées (dans le cas de mutations délétères récessives) et auront donc tendance à s'accumuler. Ce fardeau lié ne s'exprimant que chez les homozygotes entraîne une diminution de la valeur sélective chez les individus homozygotes auto-compatibles et facilite donc le maintien de l'auto-incompatibilité.

En accord avec les modèles précédents (Charlesworth and Charlesworth, 1979 ; Uyenoyama et al., 2001), Porcher and Lande (2005) ont montré que le maintien de l'auto-incompatibilité nécessitait une forte dépression de consanguinité et un nombre élevé d'haplotypes S. Néanmoins, ces conditions sont plus restreintes que celles déterminées par les modèles considérant une dépression de consanguinité constante, en raison de la purge des mutations délétères favorisée par la présence d'un d'haplotype auto-compatible. La purge étant d'autant plus efficace que le taux d'autofécondation est élevé, le maintien de l'auto-incompatibilité est favorisé par des taux d'autofécondation intermédiaires.

## 2.3 Conclusion partielle

Dans l'état actuel de nos connaissances, il apparaît que le maintien de l'auto-incompatibilité ne peut se réaliser que dans des conditions très restreintes, ce qui semble



en désaccord avec le maintien de l'auto-incompatibilité dans 50% des espèces d'Angiospermes.

## 3 Diversification du système d'auto-incompatibilité

Trois modèles ont été proposés afin d'expliquer la diversification au locus S. Matton et al. (1999) et Chookajorn et al. (2004) se sont basés sur des observations expérimentales pour proposer deux modèles verbaux permettant d'expliquer l'évolution de nouveaux haplotypes S sans rupture de l'auto-incompatibilité. Le seul modèle de génétique des populations a été proposé par Uyenoyama et al. (2001) et fait quant à lui intervenir un intermédiaire auto-compatible présentant une nouvelle spécificité sur un seul des deux gènes, suivi d'une mutation compensatrice sur l'autre gène restaurant l'auto-incompatibilité.

### 3.1 Intermédiaire à double spécificité

Le modèle de Matton et al. (1999) est basé sur des observations expérimentales de mutagenèse chez *Solanum chacoense*. Les S-RNases  $S_{11}$  et  $S_{13}$  de *Solanum chacoense* diffèrent l'une de l'autre par uniquement 10 acides aminés (Sabaellil et al., 1994), dont quatre sont situés dans les régions hypervariables. Ces quatre acides aminés suffisent à déterminer la spécificité. En effet, les S-RNases  $S_{11}$  chez lesquelles ces quatre acides aminés ont été remplacés par ceux de  $S_{13}$  rejettent le pollen  $S_{13}$  (Matton et al., 1997). En revanche, celles chez lesquelles seulement trois de ces acides aminés ont été substitués présentent une double spécificité et rejettent à la fois les pollens  $S_{11}$  et  $S_{13}$  (Figure 10).

Cette observation a mené les auteurs à proposer un modèle de diversification au locus S via un intermédiaire présentant une double spécificité (Figure 11). Ce modèle fait intervenir trois pas de mutations sur le même haplotype. Notons respectivement les gènes pistil et pollen  $A$  et  $B$ , et considérons un haplotype ancestral  $A_1B_1$  qui exprime la spécificité 1 aux deux gènes. Un premier pas de mutations au sein du gène pistil  $A_1$  permet au stigmate de reconnaître le pollen ancestral  $B_1$  ainsi qu'un nouveau type de pollen  $B_2$  encore absent de la population. A ce stade, la double-spécificité n'est que potentielle. L'haplotype intermédiaire noté  $A_{1,2}B_1$  se comporte comme un variant neutre de l'haplotype ancestral et ne présente donc aucun désavantage sélectif. Puis le gène pollen mute également pour aboutir à  $B_2$ . Ce nouvel haplotype  $A_{1,2}B_2$  est donc reconnu

## Introduction

par lui-même, mais également par l'haplotype  $A_{1,2}B_1$  dont il est issu. Enfin, le gène pistil mute à nouveau jusqu'à ne plus reconnaître le pollen ancestral  $B_1$  et on obtient bien un nouvel haplotype fonctionnel  $A_2B_2$ .

Ce scénario a été critiqué car l'haplotype  $A_{1,2}B_2$ , reconnu par deux haplotypes différent, présente un important désavantage sélectif par rapport à l'haplotype ancestral et aura donc tendance à être exclus de la population (Uyenoyama and Newbigin, 2000). De plus, le nombre de mutations successives nécessaires sur un même haplotype (au moins trois) est trop important pour expliquer le nombre élevé de spécificités observé dans les populations naturelles (Charlesworth, 2000).

	HVa region	HVb region	SI phenotype
<b>S<sub>11</sub> wt</b>	KPKLTYNYFSDKML	IDQASARKDQP	<b>S<sub>11</sub></b>
<b>S<sub>11</sub> HVapb mutant</b>	.....N...F.....	.....L.....	<b>S<sub>11</sub> and S<sub>13</sub></b>
<b>S<sub>11</sub> HVab mutant</b>	.....N..KF.....	.....L.....	<b>S<sub>13</sub></b>
<b>S<sub>13</sub> wt</b>	.....N..KF.....	.....L.....	<b>S<sub>13</sub></b>

FIGURE 10 – Différences entre les séquences protéiques des allèles  $S_{11}$  et  $S_{13}$  (wt) du gène pistil chez *Solanum charcoense* et les séquences obtenues par transgénèse à partir de  $S_{11}$ . Les phénotypes d'incompatibilité associés à chacune des séquences, c'est-à-dire leur capacité à reconnaître et à rejeter certains pollens, sont indiqués dans la colonne de droite. 3 substitutions d'acides aminés confèrent à  $S_{11}$  une double spécificité. Une quatrième substitution confère à l'allèle mutant la spécificité  $S_{13}$ . D'après Matton et al. (1999).

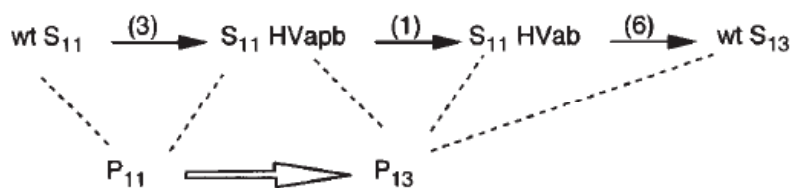


FIGURE 11 – Séquence évolutive proposée par Matton et al. (1999) pour l'apparition d'un nouvel haplotype S. En haut, mutations successives dans les régions hypervariables du gène pistil entraînant responsable de l'évolution de  $S_{11}$  vers  $S_{13}$  (voir Figure 10). Le nombre de substitutions est indiqué entre parenthèses. La flèche blanche représente les mutations hypothétiques entraînant la transition de l'allèle  $P_{11}$  du gène pollen en allèle  $P_{13}$ . En pointillés sont représentées les interactions de rejet des allèles  $P_{11}$  et  $P_{13}$  par les différents allèles au gène pistil.

### 3.2 Diversification graduelle

Chookajorn et al. (2004) se sont basés sur des expériences d'échanges de domaines protéiques et de mutagenèse chez *Brassica oleracea* pour proposer un modèle d'évolution de nouvelles spécificité faisant intervenir des mutations successives quasi-neutres menant progressivement à la scission d'une classe allélique en deux classes distinctes. Les auteurs ont mis en évidence que le remplacement de cinq acides aminés particuliers de la protéine SCR<sub>6</sub> par quatre de SCR<sub>13</sub> suffisait à conférer à SCR<sub>6</sub> une spécificité S<sub>13</sub>, tandis que l'expérience inverse, c'est-à-dire le remplacement chez SCR<sub>13</sub> de ces mêmes quatre acides aminés par les cinq de SCR<sub>6</sub>, ne changeait pas la spécificité de SCR<sub>13</sub>.

Les auteurs ont également mis en évidence que la protéine SCR pouvait tolérer un grand nombre de mutations sans que l'interaction SRK-SCR ne soit rompue, mais que ces mutations affectaient néanmoins l'intensité de cette interaction (Figure 12). Seules cinq mutations particulières affectant la surface de la protéine ont entraîné une rupture du système d'auto-incompatibilité. Cette observation a conduit les auteurs à proposer un nouveau modèle verbal selon lequel un nouvel haplotype S peut apparaître lorsque chaque classe allélique présente une importante variabilité neutre dans les séquences des gènes pollen et pistil (Figure 13). Si par hasard, dans un sous-ensemble d'une classe allélique, l'interaction entre les protéines pollen et pistil est plus forte qu'avec le reste de la classe allélique, la sélection naturelle aurait tendance à renforcer les interactions au sein des sous-classes, tout en réduisant les interactions entre sous classes. A terme, ce processus mènerait à une absence totale d'interaction entre sous-classes et donc à la scission de la classe allélique ancestrale en deux classes distinctes.

Mais ce modèle est très imprécis quant aux forces sélectives supposées agir sur l'intensité des interactions entre et au sein des différentes classes alléliques, et ne tient pas compte des forces sélectives connues pour agir sur l'évolution de l'auto-incompatibilité, notamment la dépression de consanguinité et la sélection fréquence-dépendante.

### 3.3 Intermédiaire auto-compatible

Le seul modèle de génétique des populations a été proposé par Uyenoyama et al. (2001) et reprend celui de Charlesworth and Charlesworth (1979) en faisant intervenir un haplotype intermédiaire auto-compatible exprimant une nouvelle spécificité à un seul des deux gènes du locus S. Les gènes pistil et pollen sont respectivement notés *A* et *B*. Ce modèle considère une population initiale composée de *n* haplotypes auto-incompatibles

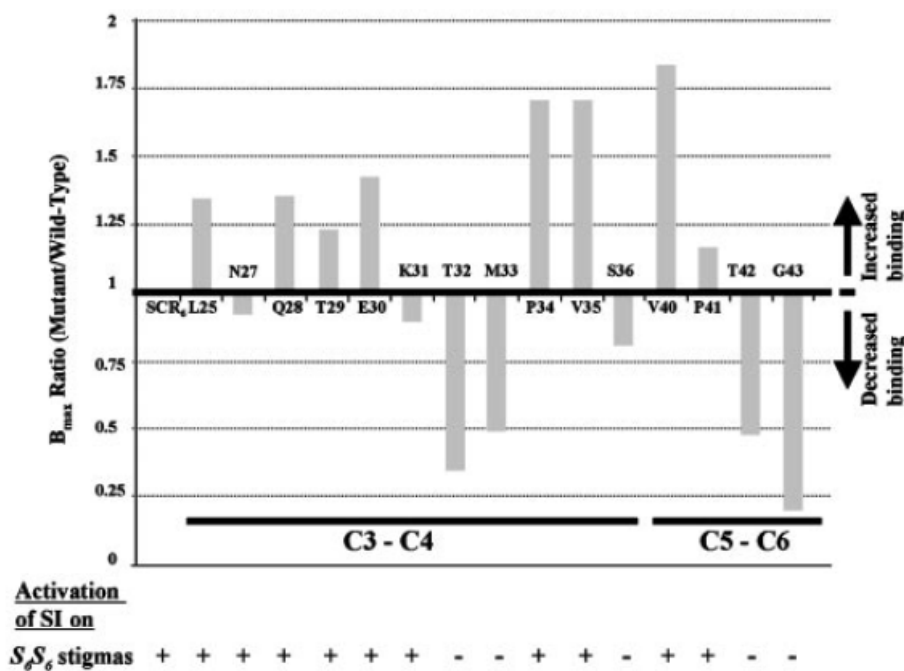


FIGURE 12 – Rapport entre la liaison des protéines SCR<sub>6</sub> mutantes avec SRK<sub>6</sub> et la liaison des protéines SCR<sub>6</sub> non modifiées avec SRK<sub>6</sub>, chez *Brassica oleracea*. Pour chaque acide aminé muté, ce rapport permet de savoir si la mutation augmente ou diminue la liaison entre les deux protéines. Les régions auxquelles appartiennent les acides aminés mutés (entre les régions C3 et C4 ou entre les régions C5 et C6) sont indiquées au bas du graphique. Seules les mutations affectant cinq codons particuliers entraînent l’absence de reconnaissance entre les deux protéines, c’est-à-dire la rupture de l’auto-incompatibilité (indiquée par un “-” sous le graphique). D’après Chookajorn et al. (2004).

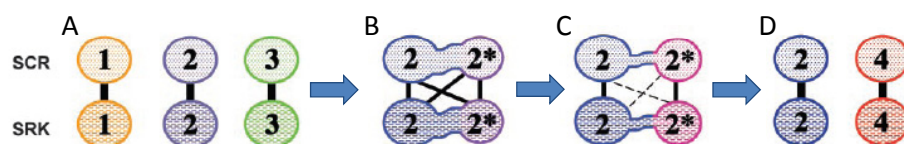


FIGURE 13 – Séquence évolutive proposée par Chookajorn et al. (2004) pour l’apparition d’un nouvel haplotype S. Chaque cercle correspond à une spécificité donnée aux gènes pistil (SRK) et pollen (SCR). Les segments reliant ces cercles indiquent la reconnaissance entre les deux protéines. Les segments en pointillés indiquent une plus faible reconnaissance (A) Etat initial avec trois haplotypes fonctionnels, présentant de la variabilité neutre au sein des classes alléliques. (B) Apparition d’un sous-ensemble de l’haplotype 2, noté 2\*. (C) La reconnaissance entre SCR et SRK devient progressivement plus forte au sein de la sous-classe 2\* qu’entre les sous classes 2 et 2\*. (D) Il n’y a plus de reconnaissance entre les protéines SCR et SRK des deux sous classes. Cette évolution aboutit donc à une nouvelle classe distincte 4, c’est-à-dire un nouvel haplotype fonctionnel.

### 3. Diversification du système d'auto-incompatibilité

$A_i B_i$  ( $1 \leq i \leq n$ ), notés  $S_i$ . Une première mutation affecte l'un des deux gènes de l'haplotype  $S_n$ , produisant ainsi soit un haplotype  $A_{n+1} B_n$  (noté  $S_a$ ) si la mutation affecte le pistil, soit un haplotype  $A_n B_{n+1}$  (noté  $S_b$ ) si la mutation affecte le pollen (Figure 14). Ces deux haplotypes expriment une nouvelle spécificité à un seul des deux gènes et sont donc auto-compatibles. Puis une seconde mutation dite compensatrice affecte le second gène, produisant dans les deux cas un nouvel haplotype auto-incompatible  $A_{n+1} B_{n+1}$  (noté  $S_{n+1}$ ).

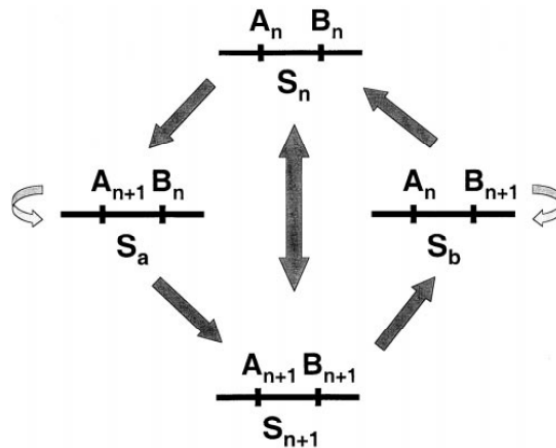


FIGURE 14 – Génotypes et phénotypes des différents haplotypes impliqués dans l'évolution d'un nouvel haplotype  $S$  d'après Uyenoyama et al. (2001). Les haplotypes qui expriment la même spécificité aux gènes pistil ( $A$ ) et pollen ( $B$ ) sont auto-incompatibles ( $S_n$  et  $S_{n+1}$ ). Au contraire, les haplotypes exprimant une spécificité différente aux deux gènes sont auto-compatibles ( $S_a$  et  $S_b$ ). Les flèches indiquent les sens de pollinisation compatible, c'est-à-dire que le pollen portant l'haplotype à la base de la flèche n'est pas reconnu par le pistil portant l'haplotype à la pointe de la flèche.

Les auteurs ont étudié le comportement des deux types de mutants auto-compatibles introduits en faible fréquence en fonction des différents paramètres du modèle (la dépression de consanguinité, le taux d'autofécondation et le nombre d'haplotypes  $S_i$  dans la population) et se sont particulièrement intéressés aux cas où ces haplotypes intermédiaires étaient maintenus en polymorphisme avec les haplotypes auto-incompatibles, puisque la diversification du système d'auto-incompatibilité nécessite à la fois le maintien des anciens haplotypes auto-incompatibles et celui de l'haplotype intermédiaire qui mènera à l'évolution d'un nouvel haplotype fonctionnel. Puis les auteurs ont étudié le devenir de l'haplotype  $S_{n+1}$  introduit en faible fréquence dans ces populations polymorphes à l'équilibre.

Dans le cas où la première mutation a lieu dans le gène pistil, l'haplotype  $S_a$  est maintenu en polymorphisme avec les haplotypes  $S_i$  (y compris l'haplotype ancestral  $S_n$ ) lorsque la dépression de consanguinité est globalement comprise entre 1/2 et 2/3, la limite supérieure diminuant légèrement quand le taux d'autofécondation augmente (Figure 8). Comme déjà observé par Charlesworth and Charlesworth (1979), ces conditions ne dépendent pas du nombre d'haplotypes  $S_i$  présents dans la population. Dans les conditions permettant le maintien de l'haplotype  $S_a$  en polymorphisme avec les  $n$  haplotypes  $S_i$ , l'haplotype  $S_{n+1}$  ne peut pas augmenter en fréquence et est toujours exclus de la population en raison de la reconnaissance asymétrique entre  $S_{n+1}$  et  $S_a$  ( $S_{n+1}$  est rejeté par  $S_a$ , mais l'inverse n'est pas vrai). La diversification du système d'auto-incompatibilité est donc impossible lorsqu'on considère une première mutation dans le gène pistil.

Dans le cas où la première mutation a lieu dans le gène pollen, l'haplotype  $S_b$  se maintient en polymorphisme avec les haplotypes auto-incompatibles  $S_i$  lorsque la dépression de consanguinité est élevée et que le taux d'autofécondation est faible (Figure 7). Ces conditions sont d'autant plus restreintes que le nombre d'haplotypes  $S_i$  dans la population est élevé. Comme déjà observé par Charlesworth and Charlesworth (1979), l'introduction d'un haplotype mutant au niveau du pollen conduit à l'exclusion de l'haplotype ancestral. Cette population polymorphe est donc composée de l'haplotype  $S_b$  et de  $n - 1$  haplotypes  $S_i$ . L'introduction de l'haplotype  $S_{n+1}$  dans cette population entraîne toujours l'exclusion de l'haplotype intermédiaire  $S_b$ , ce qui mène à une population totalement auto-incompatible composée de  $n - 1$  haplotypes  $S_i$  et du nouvel haplotype  $S_{n+1}$ .

Le modèle d'Uyenoyama et al. (2001) suggère donc que l'apparition de nouveaux haplotypes S ne peut se faire que via une première mutation dans le gène pollen, dans une gamme de paramètres caractérisée par une forte dépression de consanguinité et un faible taux d'autofécondation. En revanche, il ne permet pas d'expliquer la diversification du système d'auto-incompatibilité, car, l'haplotype ancestral étant systématiquement exclu, le nombre d'haplotypes dans la population reste constant.

### 3.4 Conclusion partielle

Jusqu'à aujourd'hui, aucun modèle n'a permis d'expliquer l'émergence de nouveaux allèles S. Pourtant, en raison du nombre élevé d'allèles S observé en populations naturelles, on s'attend à ce que les conditions permettant la diversification du système d'auto-incompatibilité soient relativement larges.

## 4 Importance de la dérive

Nous avons vu que tous les modèles concordent sur le fait qu'un haplotype auto-compatible peut envahir une population auto-incompatible sous de très larges conditions. Le mystère du maintien de l'auto-incompatibilité chez plus de la moitié des espèces d'Angiospermes n'est donc toujours pas résolu. De même, aucun modèle n'a jusqu'à présent permis d'expliquer la diversification au locus S. Mais tous les modèles précédemment décrits sont déterministes et considèrent des populations infinies. Or, on sait que la stochasticité et la dérive génétique due à la taille finie des populations sont des facteurs importants à prendre en compte en évolution, par exemple lorsqu'on considère la question évolutive majeure que représente l'évolution et le maintien du sexe. Différents modèles ont en effet montré que la stochasticité, par un effet indirect, est un mécanisme majeur dans l'évolution du sexe et de la recombinaison (Barton and Otto, 2005 ; Otto and Lenormand, 2002 ; Roze and Barton, 2006). La stochasticité a pour effet de créer des déséquilibres de liaison négatifs entre allèles, ce qui réduit l'efficacité de la sélection, que ces allèles soient avantageux ou délétères. En effet, si l'on considère une population asexuée, ces allèles seront en concurrence dans le cas d'allèles avantageux, ou seront difficilement éliminés en raison de l'absence de génotype extrême dans le cas d'allèles délétères. Le sexe, par l'effet de la ségrégation et de la recombinaison, permet de créer des combinaisons favorables de ces allèles. Un autre exemple typique de l'effet de la dérive est le cas de la dispersion (Ronce, 2007). La dérive a pour effet d'augmenter l'apparentement génétique au sein des populations et favorise donc la dispersion qui permet d'éviter d'une part la dépression de consanguinité, et d'autre part la compétition entre individus apparentés.

A ma connaissance, aucune étude n'a jusqu'à présent étudié l'effet de la dérive sur l'évolution des systèmes de reproduction en terme de taux d'autofécondation. En revanche, on sait que la dérive joue un rôle important dans l'évolution de la dépression de consanguinité, elle-même corrélée à l'évolution des taux d'autofécondations (Charlesworth and Charlesworth, 1987). La dépression de consanguinité est principalement due à des mutations délétères récessives qui ne s'expriment donc qu'à l'état homozygote. La dérive a pour effet d'augmenter l'homozygotie au sein des populations, permettant donc l'expression et la purge de ces mutations délétères. En conséquence, la dépression de consanguinité est d'autant plus faible que la taille des populations est réduite (Bataillon and Kirkpatrick, 2000 ; Glémin, 2003). Glémin et al. (2001) se sont

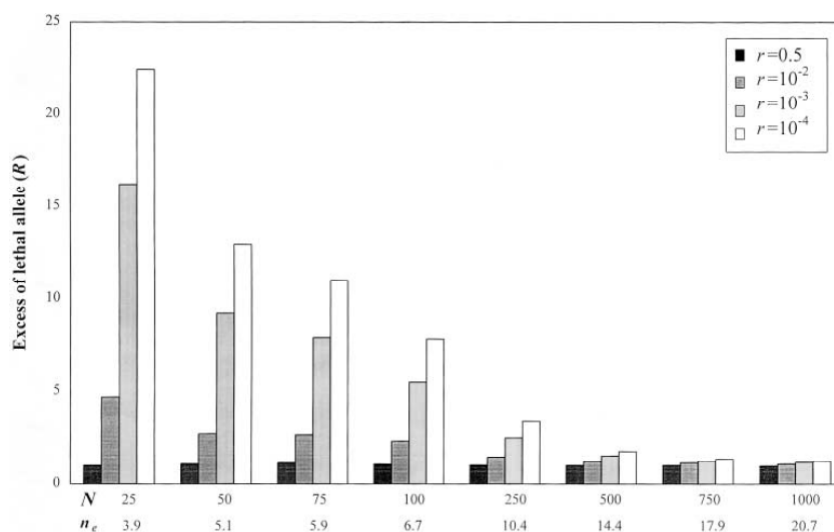


FIGURE 15 – Rapport ( $R$ ) de la fréquence moyenne des allèles létaux récessifs partiellement liés au locus S sur leur fréquence moyenne en population panmictique (en l’absence d’auto-incompatibilité).  $R$  est exprimé en fonction de la taille de la population ( $N$ ) et du nombre d’allèles S maintenus dans la population ( $n_e$ ), et pour quatre taux de recombinaison différents ( $r$ ). D’après Glémin et al. (2001).

intéressés à l’évolution de la dépression de consanguinité dans les petites populations auto-incompatibles. Ils ont montré que le système d’auto-incompatibilité a pour effet d’augmenter la dépression de consanguinité par rapport à une population panmictique, ce qui s’explique par l’absence d’autofécondation et donc le nombre réduit d’homozygotes dans ces populations, ce qui limite fortement la purge des mutations délétères. Cet effet d’augmentation de la dépression de consanguinité est particulièrement fort pour de très petites tailles de populations et lorsque les mutations sont létales (Figure 15).

Dans le cadre de ma thèse, je me suis intéressée à l’évolution du système d’auto-incompatibilité gamétophytique en population finie, en utilisant principalement une approche de simulations individus-centrées. Deux modèles ont été développés (Chapitre 1) afin d’étudier d’une part le maintien de l’auto-incompatibilité (Chapitre 2) et d’autre part l’apparition de nouveaux haplotypes S (Chapitre 3).



# Chapitre 1

## Présentation et comparaison des modèles de simulations

Afin d'étudier le maintien et la diversification du système d'auto-incompatibilité, j'ai construit deux modèles de simulations. Ces deux modèles sont basés sur le même principe général mais diffèrent sur certaines hypothèses. Le but de ce chapitre n'est pas d'expliquer en détail chacun de ces modèles (ceci sera fait dans les sections "Model" des chapitres 2 et 3), mais de mettre en évidence leur points communs et surtout leur différences afin d'éviter toute confusion lors de la lecture de ce manuscrit. Nous appellerons "modèle M" (pour maintien) et "modèle D" (pour diversification) les modèles utilisés respectivement dans les chapitres 2 et 3.

### 1.1 Structure générale commune aux deux modèles

On considère une population de taille finie composée de  $N$  individus diploïdes. Le cycle de vie se déroule en deux étapes : une étape de mutation, puis une étape de reproduction et sélection. Les individus produisent une infinité de pollen. Il n'y a donc pas de limitation pollinique. Tous les individus reçoivent une proportion  $\alpha$  d'auto-pollen. On considère un système d'auto-incompatibilité gamétophytique, c'est-à-dire qu'un croisement est compatible si la spécificité exprimée dans le pollen est différente des deux spécificités exprimées dans le pistil. Tous les individus meurent après la production des graines.

## 1.2 Structure des chromosomes

### Modèle M

Les chromosomes sont composés d'une infinité de loci sous sélection purifiante et du locus S, situé au milieu du chromosome. Le locus S est considéré comme un unique gène, c'est-à-dire que les gènes pollen et pistil ne sont pas représentés.  $k + 1$  allèles ségrègent au locus S :  $k$  allèles auto-incompatibles ( $S_i, 1 \leq i \leq k$ ) et 1 allèle auto-compatible ( $S_C$ ). Le locus S et les loci sous sélection purifiante sont partiellement liés.

### Modèle D

Les chromosomes sont constitués du seul locus S, représenté par ses deux gènes, pollen et pistil, strictement liés.  $k$  allèles ségrègent à chacun de ces deux gènes. Les haplotypes sont notés  $S_{ij}$  ( $1 \leq i, j \leq k$ ), les indices  $i$  et  $j$  représentant respectivement les spécificités exprimées au gène pistil et au gène pollen.

## 1.3 Mutation

### Modèle M

Deux types de mutations distinctes affectent les individus : les mutations au locus S et les mutations délétères. A chaque génération et pour chaque chromosome, le locus S mute avec une probabilité  $V$  et le nombre de mutations délétères est tiré dans une loi de Poisson de paramètre  $U$ . Les mutations au locus S entraînent soit la perte, soit la restauration de l'auto-incompatibilité. C'est-à-dire que les allèles  $S_i$  mutent obligatoirement en allèle  $S_C$ , tandis que l'allèle  $S_C$  mute en l'un des  $k$  allèles  $S_i$  possibles tiré aléatoirement. Un allèle auto-incompatible ne mute donc jamais en un autre allèle auto-incompatible. Les positions des mutations délétères sont tirées aléatoirement entre 0 et  $2L$ ,  $2L$  étant la longueur du chromosome.

### Modèle D

A chaque génération, le nombre total de mutations est tiré dans une loi de Poisson de paramètre  $4N\mu$ ,  $\mu$  étant le taux de mutation par gène par génération.  $U$  chromosomes sont tirés aléatoirement avec remise et subissent une mutation au gène pollen ou au gène

pistil, avec une probabilité de  $1/2$ . La spécificité du gène en question mute en l'une des  $k - 1$  autres spécificités possibles.

## 1.4 Définition de l'auto-compatibilité

### Modèle M

L'allèle  $S_C$  est auto-compatibile. Il ne reconnaît et n'est reconnu par aucun génotype présent dans la population.

### Modèle D

Un haplotype  $S_{ij}$  est auto-compatibile si  $i \neq j$ , c'est-à-dire qu'il exprime des spécificités différentes aux gènes pollen et pistil. Cet haplotype est reconnu et rejeté par tout individu portant un haplotype  $S_{jl}$ , et entraîne le rejet de tout haplotype  $S_{li}$ , quelque soit  $l$ .

## 1.5 Dépression de consanguinité

### Modèle M

La dépression de consanguinité  $\delta$  est une variable du modèle. Elle est mesurée par  $\delta = 1 - W_s/W_o$ , où  $W_s$  et  $W_o$  sont respectivement les valeurs sélectives moyennes des individus issus d'autofécondation et d'allofécondation. La valeur sélective d'un individu dépend du nombre de mutations qu'il porte à l'état homozygote et à l'état hétérozygote, ainsi que des propriétés des mutations délétères en terme de coefficients de sélection et de dominance ( $s$  et  $h$ , respectivement).

### Modèle D

La dépression de consanguinité  $\delta$  est un paramètre du modèle. Les graines issues d'autofécondation ont une probabilité  $\delta$  de mourir et de ne pas participer à la génération suivante.

## 1.6 Autofécondation

Le taux d'autofécondation varie selon les individus. Dans les deux modèles, il dépend :

- du taux d'auto-pollinisation  $\alpha$ ,
- du génotype d'auto-incompatibilité de l'individu considéré (un individu peut être auto-incompatible, partiellement auto-compatible ou totalement auto-compatible, s'il porte respectivement 0, 1 ou 2 haplotypes auto-compatibles),
- des relations de compatibilité de cet individu avec les autres individus de la population,
- de la fréquence relative des différents génotypes dans le pool pollinique.

Dans le cas du modèle M, la fréquence relative des différents génotypes est déterminée par la valeur sélective de la plante paternelle.

## 1.7 Recombinaison

### Modèle M

La probabilité qu'un évènement de recombinaison ait lieu entre deux mutations dépend de la longueur du chromosome ( $2L$ ) et de la longueur physique entre les deux mutations ( $x$ ). Plus  $L$  est grand, plus le taux de recombinaison est élevé.

### Modèle D

Les gènes pollen et pistil sont strictement liés. Il n'y a pas de recombinaison.

## Chapitre 2

# Maintenance of gametophytic self-incompatibility in a finite population with evolution of inbreeding depression

**Contribution des auteurs.** J'ai développé le modèle de simulations et analysé ses résultats, et j'ai rédigé l'article. Adrienne Ressayre a écrit et analysé les équations de l'annexe. Denis Roze a participé au développement du programme de simulations. Diala Abu Awad a participé au développement d'une première version du programme de simulations et a contribué à la formalisation de la problématique. Vincent Castric et Sylvain Billiard ont supervisé le travail.

## 2.1 Introduction

Self-incompatibility (SI) is a genetic system which enables hermaphroditic individuals to avoid self-fertilization and mating with close relatives by recognition and rejection of pollen by pistils if they express cognate specificities. In many species, SI specificity is controlled by a single multiallelic locus, the S-locus. SI is widespread, found in more than 100 families of angiosperms (Igic et al., 2008), despite the automatic selection advantage of self-fertilization (Fisher, 1941). Indeed, on average, a selfing individual transmits three copies of its genome (being both mother and father to its own seeds and father to another individual's seeds), whereas an outcrossing individual transmits only two copies of its genome (being mother to its own seeds and father to other individual's seeds). Maintenance of SI thus requires strong evolutionary benefits to counteract the automatic advantage of selfing. Among them, the increase of genetic diversity and the avoidance of inbreeding depression are believed to be major advantages (Charlesworth and Charlesworth, 1987).

It has been shown that inbreeding depression has to be very high to allow the maintenance of SI (Charlesworth and Charlesworth, 1979 ; Porcher and Lande, 2005). Indeed, Charlesworth and Charlesworth (1979) showed that a self-compatible (SC) mutant is expected to invade a SI population if inbreeding depression is less than  $2/3$ . This threshold has been determined using a model with fixed inbreeding depression, yet inbreeding depression is expected to be affected by the presence of SC alleles. Indeed, individuals bearing SC alleles can self-fertilize, which may allow the purging of deleterious mutations and thus decreases inbreeding depression. Consequently, the  $2/3$  inbreeding depression threshold is probably conservative, and the conditions for the maintenance of SI may be expected to be even more restrictive. Porcher and Lande (2005) studied the joint evolution of SI and inbreeding depression using an infinite population model. They considered an infinity of unlinked loci under purifying selection, deleterious mutations being nearly recessive lethals. Homozygote individuals thus could only be formed by selfing. They showed that invasion of a SI population by a SC mutant is easier than expected by classical models such as Charlesworth and Charlesworth (1979), especially for high selfing rate due to the purging of deleterious mutations, hence confirming that jointly considering the evolution of SI and inbreeding depression was important.

Population size is known to have a strong effect on inbreeding depression which is expected to be lower in small populations. Indeed, drift in small populations increases homozygosity and thus enhances purge against recessive deleterious mutations

(Bataillon and Kirkpatrick, 2000 ; Glémin, 2003). Glémin et al. (2001) investigated the effect of SI on inbreeding depression using a finite population two-locus model, the S-locus and one diallelic locus subject to recurrent deleterious mutations, these two loci being partially linked. They considered a SI population without any SC allele, so there was no selfing. They showed that inbreeding depression is higher in a SI population than in a panmictic population, especially in small populations and when there is little recombination between the two loci.

Population size can also have an effect on the evolution of SI. Indeed, the number of S-alleles is expected to vary according to population size (Yokoyama and Hetherington, 1982), and the number of S-alleles in the population is a key factor for the maintenance of SI, as it was shown that maintenance of SI is easier for high numbers of S-alleles (Charlesworth and Charlesworth, 1979 ; Gervais et al., 2011 ; Porcher and Lande, 2005).

Up to now, no model took into account simultaneously the joint evolution of SI and inbreeding depression, and the effect of population size and linkage. We thus combined characteristics of previous models and built a finite population model with an infinity of partially linked loci under purifying selection. We also modeled the number of S-alleles as a variable rather than a fixed parameter of our model. This allowed us to investigate how strongly the conditions for maintenance of SI in terms of self-pollination rate and inbreeding depression are qualitatively affected by this more realistic setting. We then studied how these conclusions were affected by variations of population size and recombination rate, and by the selective effect and dominance coefficient of the deleterious mutations.

## 2.2 Model

We used a modified version of the model described in Roze (2009). Briefly, this model assumes a finite population of  $N$  diploid individuals. Individuals are modeled as two chromosomes comprising an infinite number of diallelic loci under purifying selection. We modified the model by implementing a S-locus (located at the center of the chromosome). We assume that  $k + 1$  alleles can segregate at the S-locus :  $k$  self-incompatible alleles  $S_i$  ( $1 \leq i \leq k$ ) and a self-compatible allele  $S_C$ . The life cycle occurs in three successive steps : mutation, reproduction/selection, and meiosis.

### 2.2.1 Mutation

Each generation, for each chromosome, the S-locus mutates with probability  $V$ . This mutation leads to either loss or restoration of self-incompatibility, *i.e.* allele  $S_i$  mutates to  $S_C$ , whereas  $S_C$  mutates to one of the  $k$  possible  $S_i$  drawn at random. For each chromosome, the number of deleterious mutations introduced each generation is drawn from a Poisson distribution with parameter  $U$ . Positions of mutations are drawn at random between 0 and  $2L$ ,  $2L$  being the length of the chromosome. The number of deleterious mutations is thus effectively infinite. All deleterious mutations have the same selection and dominance coefficients ( $s$  and  $h$  respectively).

### 2.2.2 Reproduction and selection

We assume a species with gametophytic self-incompatibility (GSI), *i.e.* a mate is compatible if the specificity expressed by pollen is different from the two specificities codominantly expressed in style. In GSI, specificities are subject to negative frequency-dependent selection because pollen with rare specificity can fertilize more individuals than pollen with common specificity (Wright, 1939). A plant with genotype  $S_i S_j$  ( $i \neq j$ ) is self-incompatible and can be fertilized by pollen of genotype  $S_l$  ( $l \neq i$  and  $l \neq j$ ) or  $S_C$ . A plant with genotype  $S_i S_C$  is partially self-compatible (through its  $S_C$  pollen) and can be fertilized by any pollen whose genotype is different from  $S_i$ . A plant  $S_C S_C$  is fully self-compatible and can be fertilized by all pollen.

Selection occurs at two levels : individuals reproduce or not according to their fitness, which depends upon the number of deleterious mutations in their genome (purifying selection), and according to their compatibility with other individuals of the population, which depends upon the S-alleles they carry (balancing selection). Specifically, a first individual is randomly drawn as the mother with probability  $1/N$ . This plant  $n$  has a probability  $W_n/W_{max}$  to be fertilized, with  $W_n$  its own fitness and  $W_{max}$  the fitness of the fittest individual in the population. Individual fitness is calculated as  $W_n = (1 - hs)^{N_{he}}(1 - s)^{N_{ho}}$ , where  $N_{he}$  and  $N_{ho}$  are respectively the number of mutations in the heterozygous and homozygous state in the focal individual  $n$ . The maternal gamete is obtained by recombination between the two chromosomes (see below). If the plant bears at least one self-compatible allele  $S_C$ , it does self with probability  $a$  which depends on the proportion  $\alpha$  of self-pollen received by the plant and on the fitness of the individuals of the population assumed to contribute to the load of pollen deposited on the stigma



in proportion of their fitness ( $W_m$ ). The selfing rate of the individual  $n$  is thus :

$$a = \frac{\alpha \gamma_{nn} W_n}{\alpha \gamma_{nn} W_n + (1 - \alpha) \sum_{m, m \neq n} \frac{\gamma_{mn} W_m}{N-1}}, \quad (2.1)$$

where  $\gamma_{mn}$  is the number of S-alleles carried by the individual  $m$  which are compatible with the plant  $n$  ( $\gamma_{mn} = 0, 1$  or  $2$ ). Selfing rate  $a$  is only measured for self-compatible individuals ( $\gamma_{nn} \neq 0$ ) so that self-incompatible individuals always reproduce by outcrossing even when  $\alpha = 1$ .

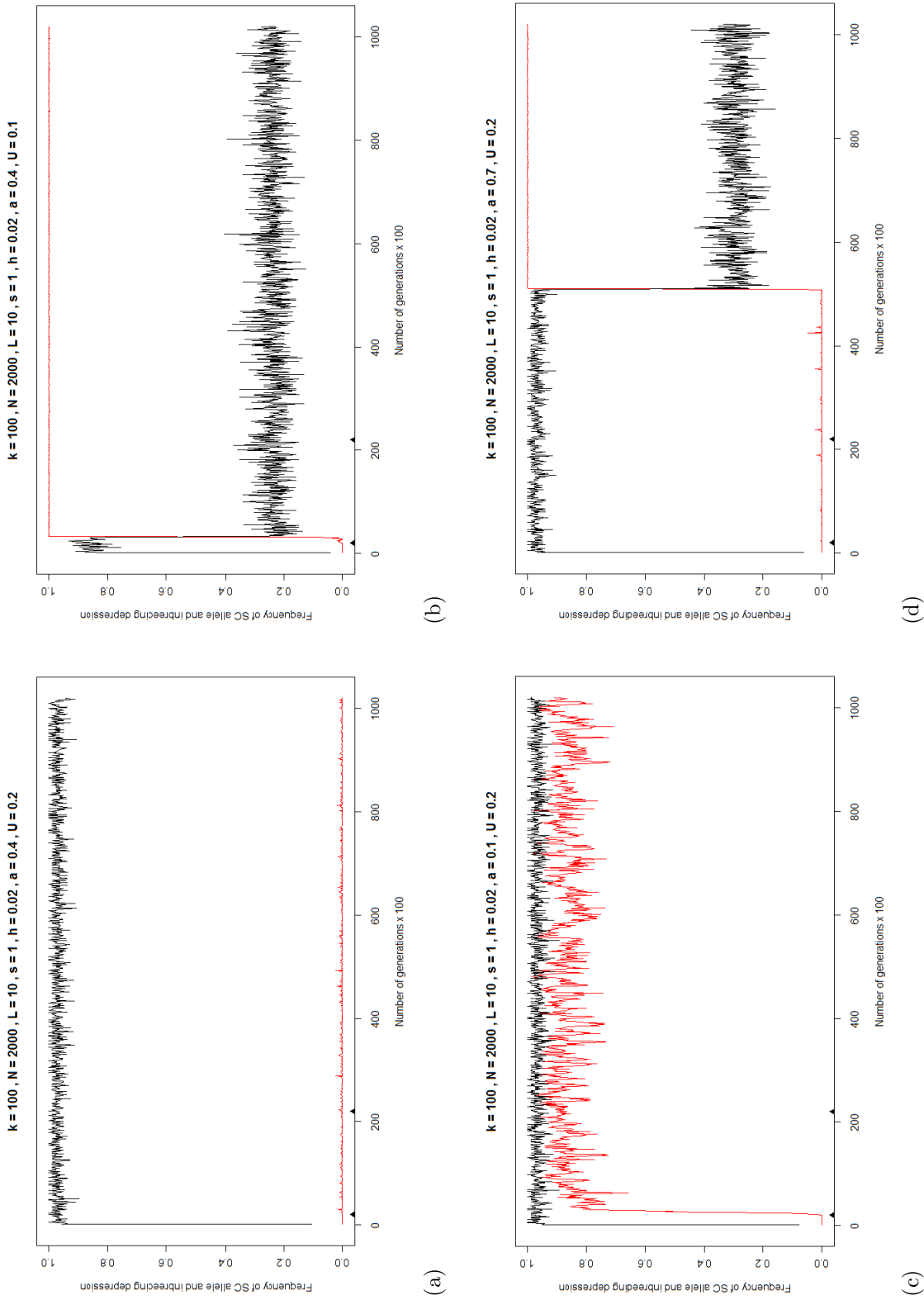
If the plant does self, a paternal gamete is created by recombination (see below) so that it bears the self-compatible allele  $S_C$ . If the plant does not self, a recombined chromosome from another individual  $m$  is randomly drawn as the paternal gamete with a probability  $1/2(N-1)$ . It fertilizes the mother with a probability  $W_m/W_{max}$ , and only if it is compatible with the two S-alleles of the mother.

### 2.2.3 Gamete production and recombination

Recombined gametes are obtained by walking along the parental genome. For optimal computation, two distinct procedures are used to generate crossing-overs. If the number of deleterious mutations in the individual is larger than the map length ( $L$ ), the number of crossing-overs is drawn from a Poisson distribution with parameter  $L$  and their positions are drawn at random in a uniform distribution. If the number of mutations is lower than the map length, the probability that a crossing-over occurs between two mutations is given by  $(1 - e^{-2Lx})/2$  where  $x$  is the physical length between the two mutations. Hence,  $L$  can be considered a proxy for the rate of recombination, larger  $L$  corresponding to higher recombination rate.

### 2.2.4 Simulation methods and parameter values

Individuals are initially free of deleterious mutations and fully self-incompatible. Their two S-alleles are drawn at random among  $k = 100$  S-alleles possibly segregating in the population, and all individuals are heterozygous at the S-locus. During the first 2,000 generations, deleterious mutations are introduced in the population but there is no mutation at the S-locus. During the next 100,000 generations, self-compatible alleles are also introduced in the population with mutation rate at the S-locus  $V = 10^{-3}$ . Every 100 generations, the population is sampled and different variables are recorded :



**FIGURE 2.1** – Dynamics of SI and inbreeding depression over time with  $N = 2000, L = 10, s = 1$  and  $h = 0, 02$ . Frequency of SC allele is plotted in red and inbreeding depression is plotted in black. The two marks on the abscisse indicate the generation at which SC allele was introduced (at 2000 generations) and the generation from which the variables were monitored (at 22000 generations). a)  $a = 0, 4$  and  $U = 0, 2$ . Case of maintenance of SI. b)  $a = 0, 4$  and  $U = 0, 1$ . Case of loss of SI. c)  $a = 0, 1$  and  $U = 0, 2$ . Case of stable polymorphism. d)  $a = 0, 7$  and  $U = 0, 2$ . Case of "false polymorphism" This case was detected as polymorphism but corresponds to a sudden loss of SI by invasion of SC allele after having been maintained for a while.

the number of SI alleles present in the population and the frequency of the SC allele, the effective selfing rate (*i.e.* the number of self-fertilization events divided by  $N$ ) and the level of inbreeding depression. Inbreeding depression is measured as  $\delta = 1 - W_s/W_o$ , where  $W_s$  and  $W_o$  are the mean fitness of selfed and outcrossed progenies, respectively. These two fitnesses are calculated by creating 100 selfed and 100 outcrossed offspring from the sampled population, without taking into account the compatibility between S-alleles of the parents. For each simulation run, the measured variables were averaged over the last 800 samples. The maintenance of SI was determined based on the mean frequency of SC alleles. If it was larger than 0.98, SI was considered lost. If it was lower than 0.02, SI was considered maintained. If the frequency was between 0.02 and 0.98, the system was considered to be in an intermediate state that could correspond either to a stable polymorphism with both SI and SC alleles, or a "false polymorphism" due to a non constant state over the time the population was monitored (last 80,000 generations; see results below).

In order to investigate more specifically the effect of recombination and drift, we performed simulations with three recombination rates ( $L = 0.1, 1$  and  $10$ ) and three population sizes ( $N = 500, 2000$  and  $5000$ ) in the case of nearly recessive lethal mutations ( $s = 1$  and  $h = 0.02$ ). These values of  $s$  and  $h$  are similar to those used in Porcher and Lande (2005). In order to investigate the effect of the properties of mutations, we performed simulations with four selection coefficients ( $s = 0.01, 0.1, 0.5$  and  $1$ ) and four dominance coefficients ( $h = 0, 0.02, 0.1$  and  $0.5$ ) in the case of a population of  $N = 2,000$  individuals and with a high recombination rate ( $L = 10$ ), where mutations can be considered as approximately unlinked (Roze, 2009).

## 2.3 Results

### 2.3.1 Dynamics of SI and inbreeding depression

We observed three contrasted evolutionary outcomes on the basis of the mean frequency of the SC allele (Figure 2.1). SI was either maintained when the SC allele showed no increase in frequency (Figure 2.1a). SI was lost when the SC allele invaded the population (Figure 2.1b), or SI was maintained in a polymorphic state with intermediate mean frequencies of the SC allele. Intermediate frequencies reflected either stable polymorphism with both SI and SC alleles (Figure 2.1c) or "false polymorphism" (Figure 2.1d).

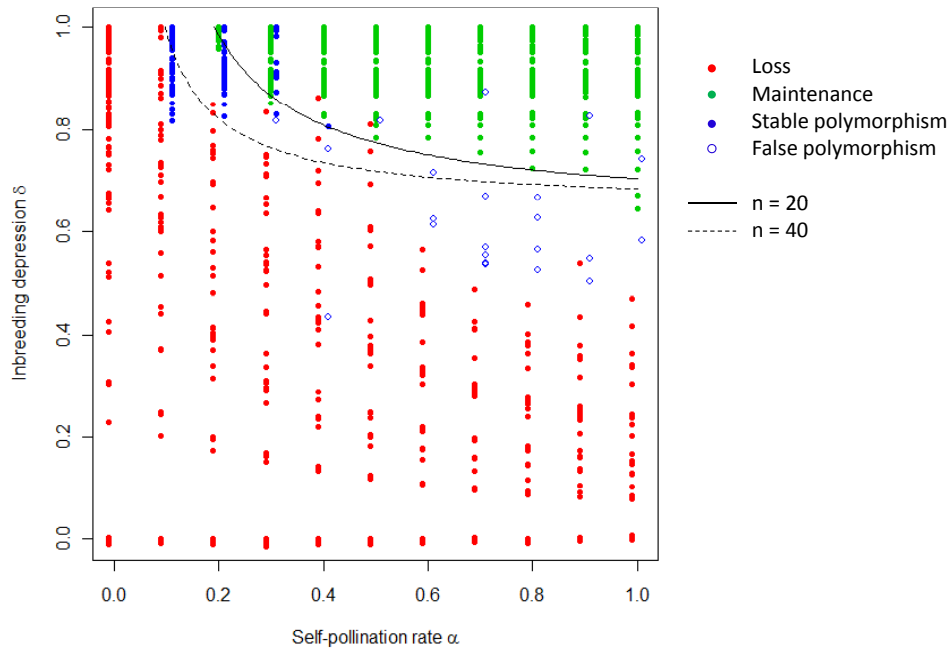


FIGURE 2.2 – Different evolutionary outcomes according to the rate of self pollen  $\alpha$  and the observed inbreeding depression  $\delta$ , for all the parameter values we investigated. Each point refers to one simulation for one set of parameters. Green points refer to cases when SI was maintained, *i.e.* the mean frequency of SC allele was lower than 0.02. Red points refer to cases when SI was lost, *i.e.* the mean frequency of SC allele was higher than 0.98. Blue points and empty circles refer to "intermediate" cases, *i.e.* the mean frequency of SC allele was between 0.02 and 0.98. Blue points refer to stable polymorphism whereas empty blue circles refer to « false polymorphism », when maintenance was followed by a sudden loss of SI. Points are slightly staggered to improve the reading of the figure. Solid and dashed lines represent the critical inbreeding depression above which SC allele is excluded, with the number of SI alleles  $n = 20$  and  $n = 40$  respectively (from 2.11 in appendix).

In this last case, the SI population was maintained and then suddenly invaded by the SC allele during the period when parameters were monitored (last 80,000 generations). In line with Porcher and Lande (2005), the increase in frequency of the SC allele was almost always accompanied by an important decrease of inbreeding depression, especially for high self-pollination rates (Figures 2.1 and 2.2).

### 2.3.2 Conditions for maintenance of SI

The three evolutionary outcomes described above correspond to three distinct parameter regions classically observed in models for the maintenance of SI

(Charlesworth and Charlesworth, 1979 ; Porcher and Lande, 2005). As shown on Figure 2.2, parameter spaces leading to either maintenance, loss or polymorphism were largely distinct. Although results for all tested parameter values are plotted on a same graph, these three regions are well defined and only rarely overlapped. Stable polymorphism was only observed for a low proportion of self-pollen ( $0.1 \leq \alpha \leq 0.3$ ), together with high inbreeding depression ( $\delta > 0.8$ ), which is consistent with Porcher and Lande (2005). As also predicted from infinite population models (Charlesworth and Charlesworth, 1979 ; Porcher and Lande, 2005 ; Uyenoyama et al., 2001), the evolution of a SI population toward either maintenance or loss of SI strongly depended on the self-pollination rate and the maintenance of the system was always observed together with high inbreeding depression ( $\delta > 2/3$ ). The critical inbreeding depression above which the SC allele was excluded (*i.e.* the lower limit of the maintenance region) was lower when the self-pollination rate  $\alpha$  was higher, which is consistent with infinite population model with constant inbreeding depression and fixed number of S-alleles (Charlesworth and Charlesworth, 1979 ; Uyenoyama et al., 2001). However, the critical inbreeding depression necessary to maintain SI observed in our simulations was higher than those expected by previous models. Indeed, the maximum number of S-alleles maintained in our simulations was 36.5, whereas the critical inbreeding depression we observed was consistent with those expected for about 20 S-alleles by analytical model (appendix).

### 2.3.3 Effect of the rate of deleterious mutation

As shown on Figures 2.3 and 2.4, the rate of deleterious mutation  $U$  has a strong effect on whether SI can be maintained or not. The higher  $U$ , the higher inbreeding depression is and the more likely SI is to be maintained. Interestingly, the effect of  $U$  on both inbreeding depression and the maintenance of SI was non-monotonous, being stronger for low values of  $U$  and becoming weaker as  $U$  increases.

### 2.3.4 Effect of recombination and drift

To investigate the effect of recombination and drift, we focused on the case of nearly recessive lethal mutations which are the most susceptible to undergo purging (Glémin, 2003 ; Lande and Schemske, 1985). Figure 2.3 shows that although inbreeding depression always remained very high when SI was maintained,  $\delta$  tended to be lower for lower recombination rate. Indeed,  $\delta$  never exceeded 0.94 with  $L = 1$  whereas it reached 1 for  $L = 10$ .

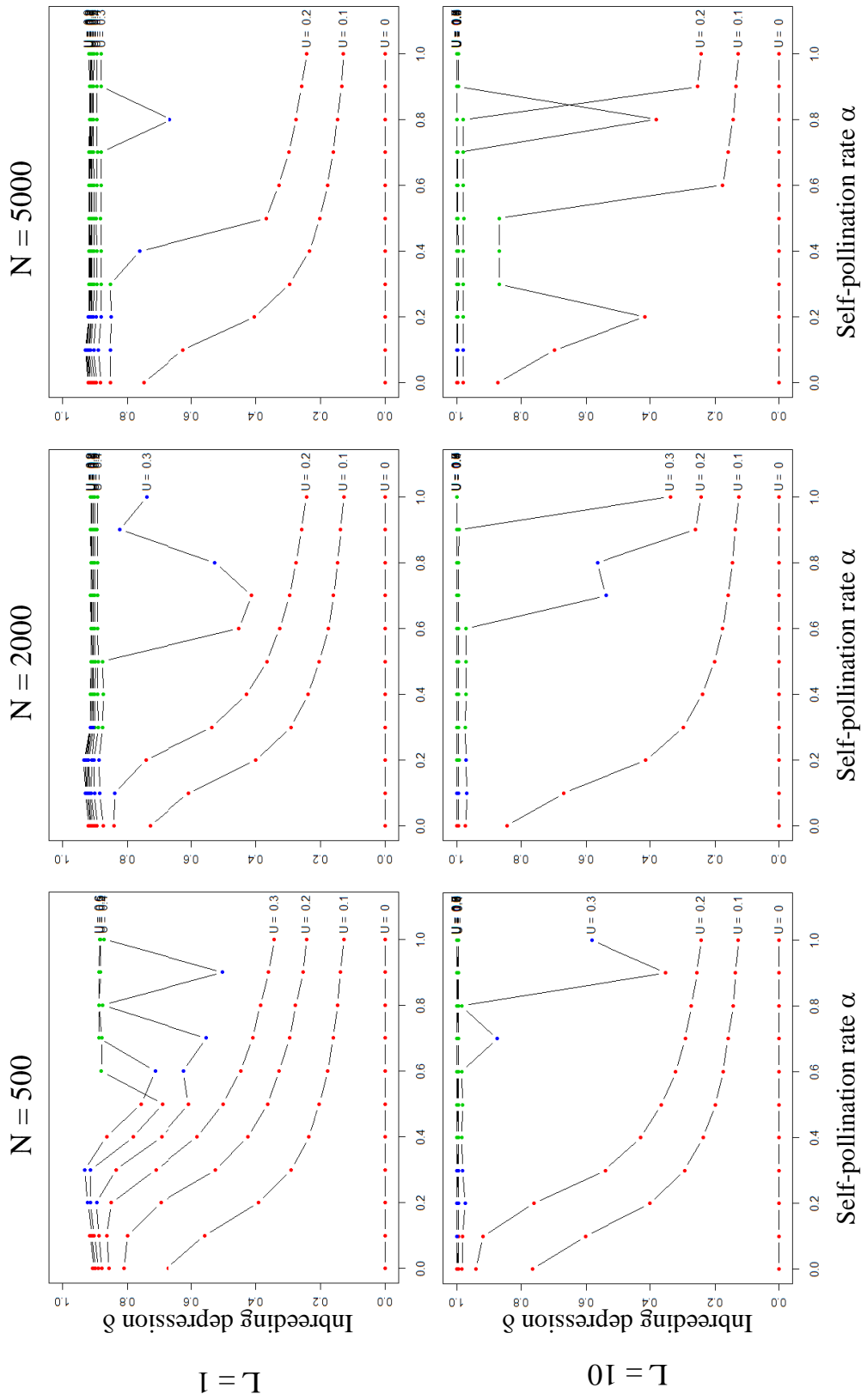
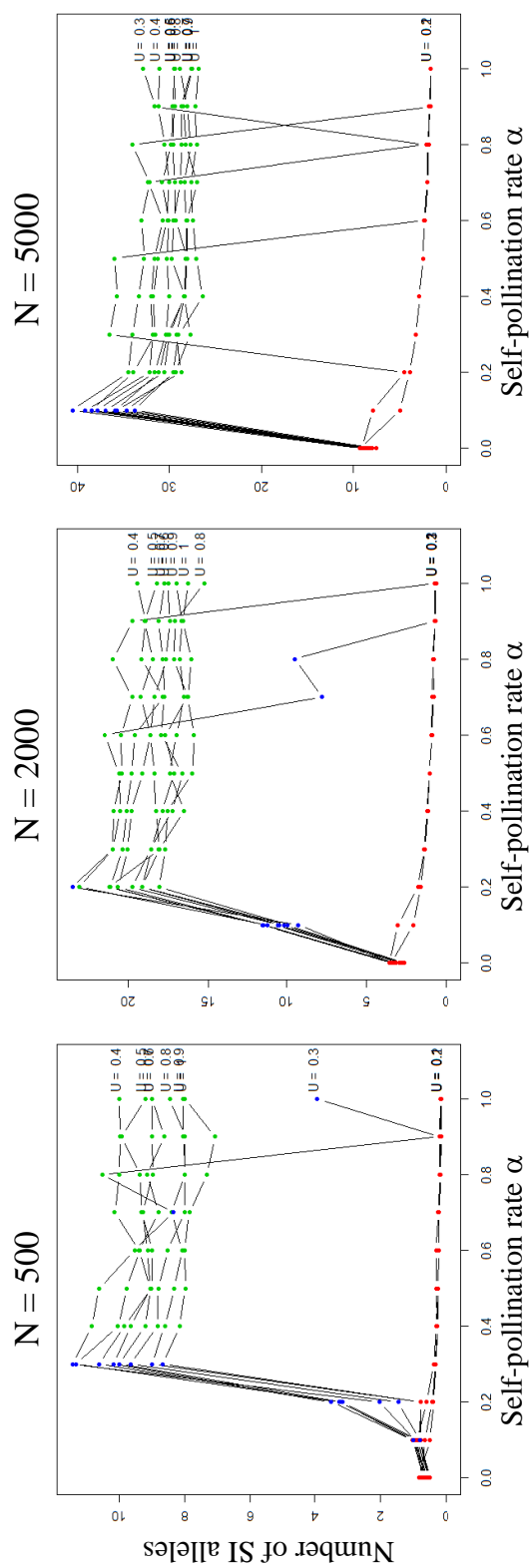


FIGURE 2.3 – Evolutionary outcome and inbreeding depression for different population sizes and recombination rates in case of nearly recessive lethal mutations ( $h = 0, 02$  and  $s = 1$ ). The colors of the points indicate the evolutionary outcome (maintenance, loss or "intermediate" cases, see legend of Figure 2.2). Blue points refer to both stable and "false" polymorphism. Points referring to a same rate of deleterious mutations  $U$  are joined by a black line.



**FIGURE 2.4** – Number of SI alleles according to the rate of self-pollen in case of nearly recessive lethal mutations and for  $N = 2000$  and  $L = 10$ . The colors of the points indicate the evolutionary outcome (see legend of Figure 2.2). Blue points refer to both stable and "false" polymorphism. Points referring to the same rate of deleterious mutations are joined by a black line.

In line with this result, SI was more easily maintained with higher recombination rate and population size. Indeed, for lower  $L$  and  $N$ , the critical deleterious mutation rate  $U$  required for SI to be maintained was higher. Yokoyama and Hetherington (1982) showed that the expected number of SI alleles is higher in large populations. In accordance with this prediction, we observed that the mean numbers of SI alleles for  $N = 500$ ,  $N = 2000$  and  $N = 5000$  were respectively 8.95, 19.51 and 30.07 when  $L = 10$  (Figure 2.4). Interestingly, the number of SI alleles was also lower for lower recombination rate. Indeed, when  $L = 1$ , the mean numbers of SI alleles for  $N = 500$ ,  $N = 2000$  and  $N = 5000$  are respectively 6.84, 12.06 and 18.57 (results not shown), suggesting that linkage among deleterious mutation and/or between deleterious mutation and the S-locus had a strong effect on the dynamics of SI. As also apparent from Figure 2.4, the number of SI alleles decreased when  $\alpha$  increased. A possible explanation for such a relationship between the expected number of SI alleles in the population and the self-pollination rate may rely on the decrease of mating partners for self-incompatible individuals as the frequency of self-compatible individuals and self-pollination rate increase. When  $\alpha$  is high, the outcrossing advantage for SI individuals is reduced since they have a limited access to the ovules produced by SC individuals. At the extreme, when  $\alpha = 1$ , a fraction of the population is not accessible for the SI individuals. In short, as  $\alpha$  and the frequency of SC individuals increase in the population, the effective size of the SI populations decreases and consequently the expected number of SI alleles maintained in the population also decreases as explained by the balance between mutation, negative frequency-dependent selection and drift.

### 2.3.5 Effect of the properties of mutations

Consistent with the fact that high inbreeding depression is required for the maintenance of SI, Figure 2.5 shows that SI is more easily maintained with strongly deleterious recessive mutations. Indeed, with  $s = 0.5$  and  $h = 0$ , SI is always maintained when  $a < 0.1$ , and the inbreeding depression is very high ( $\delta > 0.97$ ). For codominant mutations ( $h = 0.5$ ), SI is always lost and there is no inbreeding depression ( $\delta < 10^{-2}$ ).



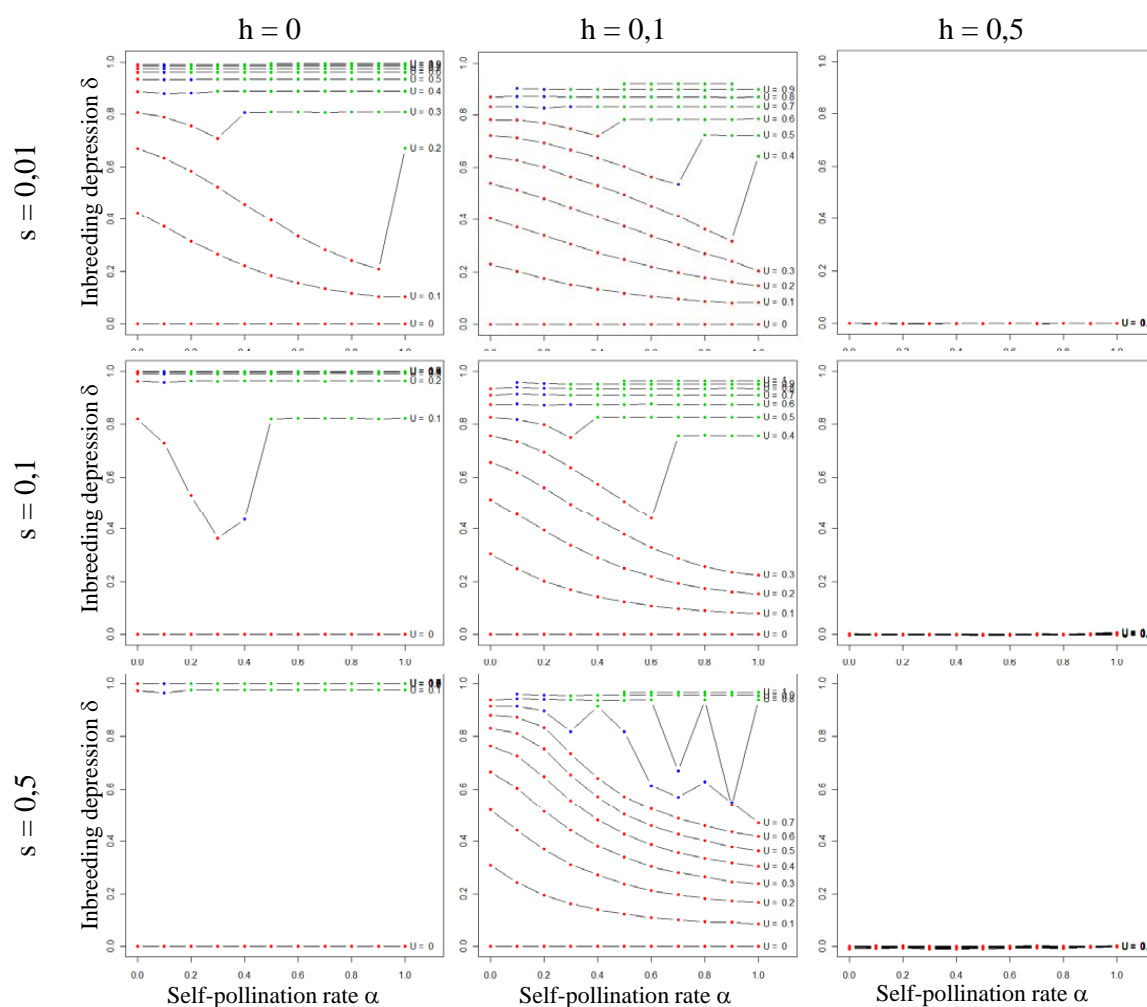


FIGURE 2.5 – Evolutionary outcome and inbreeding depression for different dominance ( $h$ ) and selection coefficients ( $s$ ) with  $N = 2000$  and  $L = 10$ . The colors of the points indicate the evolutionary outcome (maintenance, loss or "intermediate" state, see legend of Figure 2.2). Blue points refer to both stable and "false" polymorphism. Points referring to a same rate of deleterious mutations  $U$  are joined by a black line.

## 2.4 Discussion

### 2.4.1 Conditions for maintenance of SI

SI is widespread among angiosperms, especially among species with gametophytic SI. However, all models for the maintenance of SI, including ours, showed that a SC mutant can invade a SI population under a wide range of parameters (Charlesworth and Charlesworth, 1979 ; Gervais et al., 2011 ; Porcher and Lande, 2005 ; Uyenoyama et al., 2001). Indeed, maintenance of SI is expected to occur in populations with high inbreeding depression and the conditions strongly depend on the self-pollination rate. Our results showed that the critical inbreeding depression is higher for lower self-pollination rates, which is consistent with expectations of infinite population models with constant inbreeding depression (Charlesworth and Charlesworth, 1979, appendix), although the critical inbreeding depression seems higher in our model than predicted by previous models for a given number of SI alleles. However, with our approach, we may have overestimated the critical inbreeding depression. Porcher and Lande (2005) considered an infinite population in which homozygosity only occurs by selfing. They showed that for high selfing rates, a SC mutant can invade a SI population regardless of inbreeding depression because deleterious mutations are more easily purged. On the contrary, as mentioned above, our simulations showed that maintenance of SI is easier for high selfing rates.

Our results showed that conditions for maintenance of SI are mainly driven by inbreeding depression and self-pollination rate (Figure 2.2), inbreeding depression being indirectly determined by the combination of the other parameters. Hence, measuring inbreeding depression and selfing rate in natural populations could thus be sufficient to determine whether any given population is expected to maintain SI.

### 2.4.2 Population size and number of SI alleles

Previous models on the maintenance of SI considered the number of SI alleles as a fixed parameter and showed that the higher the number of SI alleles in the population, the more easily SI is maintained (Charlesworth and Charlesworth, 1979 ; Porcher and Lande, 2005). However, the number of SI alleles is known to depend on the population size (Yokoyama and Hetherington, 1982). Therefore, we considered the number of SI alleles as a variable rather than a parameter, and as expected we observed

a lower number of SI alleles maintained in smaller populations. Yet at the same time, the purge of deleterious mutations is expected to be more efficient in smaller populations (Bataillon and Kirkpatrick, 2000 ; Glémin, 2003), reducing inbreeding depression. Hence, these two indirect effects of population size affect the maintenance of SI in the same direction, *i.e.* lower inbreeding depression and lower number of SI alleles, leading to SC mutants invading more easily a small population with lower inbreeding depression and a lower number of SI alleles than a large population.

We also observed that, in cases where SI was maintained, the number of SI alleles decreased as the self-pollination rate increased, yet the effective selfing rate remained almost constant and very low ( $< 0.015$ ) because of the low frequency of SC mutants ( $< 0.02$ ).

### 2.4.3 Recombination rate

Glémin et al. (2001) suggested that inbreeding depression should be lower for higher recombination rate between between the S-locus and a deleterious mutation. On the contrary, our results showed a reverse pattern, with lower inbreeding depression for lower recombination rate. Two explanations can be put forward to account for this discrepancy. First, Glémin et al. (2001)'s model considers only one locus under purifying selection partially linked to the S-locus whereas our model considers a potential infinity of loci under purifying selection. Second, another major difference between our model and Glémin et al. (2001) is that SC mutants are introduced, potentially causing the breakdown of SI. Although the most important difference was observed in cases where SI was maintained, it would be interesting to determine whether a SC mutant can influence the evolution of inbreeding depression even if it has a very low frequency.

### 2.4.4 Properties of mutations

We considered that all deleterious mutations along the genome have the same properties in terms of selection and dominance coefficients, yet inbreeding depression in natural populations is due to both highly deleterious recessive mutations and partially recessive mildly deleterious mutations (Charlesworth and Willis, 2009). However, purging is most effective on highly deleterious recessive mutations whereas mildly deleterious mutations are not easily purged by inbreeding (Lande et al., 1994). Consequently, the proportion of inbreeding depression due to mildly deleterious mutations is expected to be almost

constant over time. Porcher and Lande (2005) showed that the introduction into their model of constant background inbreeding depression due to partially recessive mildly deleterious mutations in addition to inbreeding depression due to nearly recessive lethal mutations had two effects : the critical inbreeding depression above which SI is maintained is lower, and SI can be maintained even for highest selfing rates.

### **2.4.5 Polymorphism**

Our results showed that stable polymorphism with both SC and SI alleles is expected only for a very small range of parameters, which is similar to Porcher and Lande (2005). This is consistent with observations that SC alleles are rarely found in natural SI populations (Stone, 2002). However, some cases have been reported and self-incompatibility often appears to be quantitative rather than qualitative, with some partially SC alleles (Mena-Ali and Stephenson, 2007 ; Paape et al., 2011). Moreover, the incidence of polymorphism may be underestimated because some SC alleles can be interpreted as errors, among other things because of the variability in the expression of SI.

The conditions required to observe a polymorphism with both SI and SC alleles has important implications for the evolution of SI. The S-locus is composed of two linked genes, one expressed in pistil, and one expressed in pollen, and the SI reaction is functional if the two genes express cognate specificities. Consequently, the evolution of new S-allele requires two mutations, one on each gene, the first mutation leading to a SC intermediate, and the second one restoring SI. Diversification of SI can thus only occur if SI and SC alleles are both maintained in a population, at least during a sufficient time so that a mutation restoring SI and thus creating a new SI allele appears (Gervais et al., 2011). Our results allowing for the joint evolution of SI and inbreeding depression in a finite population with linkage suggest that conditions allowing for diversification of SI thus seem stringent.

### **2.4.6 Conclusion**

All parameters together, the conditions for maintenance of SI we observed in terms of inbreeding depression and self-pollination rate are very similar to those expected by models with constant inbreeding depression. Evolution of inbreeding depression does not seem to be the key to understand the maintenance of SI.

## 2.5 Appendix

### Infinite population model with constant inbreeding depression in seeds

Let us consider an infinite population with  $n$  ( $n \geq 3$ ) SI alleles denoted  $S_i$  ( $1 \leq i \leq n$ ) and a SC allele denoted  $S_C$ . The frequencies of SI alleles at equilibrium are all equal to  $p$ . The frequency of  $S_C$  is thus  $q = 1 - np$ . Three kinds of genotypes can be found in this population : SI individuals  $S_i S_j$  ( $1 \leq i, j \leq n; i \neq j$ ), partially SC heterozygous individuals  $S_i S_C$ , and fully SC homozygous individuals  $S_C S_C$ . Let  $x_{ij}$ ,  $x_{iC}$  and  $x_{CC}$  be the frequencies of these three genotypes, respectively. Each individual receives a proportion of self-pollen  $\alpha$  and a proportion  $1 - \alpha$  comes from other plants. Selfing rate of fully SC individuals  $S_C S_C$  is thus  $\alpha$ . Partially SC plants  $S_i S_C$  can be fertilized by all haploid genotypes but one, and half of the pollen they produce is SI. Selfing rate of plants  $S_i S_C$  is thus :

$$a = \frac{\alpha/2}{\alpha/2 + (1 - \alpha)(1 - p)}. \quad (2.2)$$

Inbreeding depression takes place in inbred seeds that suffer a reduction of fitness of  $1 - \delta$  relative to outcrossed seeds. The average fitness of the population is thus :

$$\bar{W} = 1 - \delta(\alpha x_{CC} + a x_{iC}). \quad (2.3)$$

The genotypic frequencies after selection are :

$$x'_{CC} = \frac{1}{\bar{W}} \left( x_{CC} (\alpha (1 - \delta) + (1 - \alpha) q) + \frac{x_{iC}}{2} \left( a (1 - \delta) + \frac{(1 - a) q}{1 - p} \right) \right), \quad (2.4)$$

$$x'_{iC} = \frac{1}{\bar{W}} \left( x_{CC} (1 - \alpha) (1 - q) + \frac{x_{iC}}{2} (a (1 - \delta) + (1 - a)) + \frac{x_{ij} q}{1 - 2p} \right), \quad (2.5)$$

$$x'_{ij} = \frac{1}{\bar{W}} \left( \frac{x_{iC} (1 - a) (n - 1) p}{2 (1 - p)} + \frac{x_{ij} (n - 2) p}{1 - 2p} \right). \quad (2.6)$$

We now consider the introduction of a rare SC allele in a SI population. Within such

a population, there is no selfing. the mean fitness of the population is thus  $\bar{W} = 1$ . When  $S_C$  is extremely rare ( $q \ll 1$ ,  $x_{CC} \sim 0, x_{iC} \sim 2q$  and  $x_{ij} \sim 1$ ), the recursion equations can be approximated by :

$$x'_{CC} \sim \frac{x_{iC}\alpha(1-\delta)}{2(\alpha+2(1-\alpha)(1-1/n))}, \quad (2.7)$$

$$x'_{iC} \sim \frac{x_{iC}}{2} \left( \frac{\alpha(1-\delta)}{\alpha+2(1-\alpha)(1-1/n)} + \frac{(1-\alpha)(1-1/n)}{\alpha/2+(1-\alpha)(1-1/n)} \right) + \frac{q}{1-2/n}. \quad (2.8)$$

The proportion of  $S_C$  after selection is approximated by :

$$q' \sim q \left( \frac{\alpha(1-\delta) + \alpha(1-\delta)/2 + (1-\alpha)(1-1/n)}{\alpha+2(1-\alpha)(1-1/n)} + \frac{1}{2-4/n} \right). \quad (2.9)$$

The change in frequency of  $S_C$  is thus approximated by :

$$\Delta_q \sim q \left( \frac{\alpha(1-\delta) + \alpha(1-\delta)/2 + (1-\alpha)(1-1/n)}{\alpha+2(1-\alpha)(1-1/n)} + \frac{1}{2-4/n} \right). \quad (2.10)$$

The SC allele is unable to increase its frequency for :

$$\delta > \frac{2(n-1)(2+(n-2)\alpha)}{3n\alpha(n-2)} \quad (2.11)$$

## Chapitre 3

# Origin and diversification dynamics of self-incompatibility haplotypes

**Contribution des auteurs.** J'ai développé et analysé le modèle analytique, ainsi que le programme de simulations, et j'ai rédigé l'article. Adrienne Ressayre a participé au développement du modèle analytique et à la rédaction de l'article. Vincent Castric et Sylvain Billiard ont supervisé le travail.

## 3.1 Introduction

Genes involved in recognition systems, such as the major histocompatibility complex in vertebrates (Solberg et al., 2008), mating types either in fungi (Billiard et al., 2011) or protists (Phadke and Zufall, 2009) and self-incompatibility (SI) in plants (Lawrence, 2000) typically show extraordinary high levels of genetic diversity. SI is widespread in Angiosperms (found in more than 100 families, Iqbal et al. 2008), and is highly multiallelic (up to 200 SI haplotypes, Lawrence 2000). This system enables hermaphrodite plants to avoid selfing and mating with close relatives and is based on recognition and rejection of pollen by pistils if they express cognate SI specificities. In many species, SI specificity is controlled by a single genetic locus (the S-locus) composed of two linked genes, one expressed in pollen and the other in styles (Takayama and Isogai, 2005). The maintenance of high diversity in gametophytic self-incompatibility (GSI) is easily explained by negative frequency-dependent selection, whereby individuals with a rare SI haplotype can fertilize more partners than individuals with a common SI haplotype. As a consequence, rare SI haplotypes are unlikely to be lost by drift, especially new SI haplotypes arising by mutation (Wright, 1939). A major unsolved puzzle, though, is how new SI haplotypes appear, while it is a long-standing question (Fisher, 1961 ; Lewis, 1949). Some information exists about sequence changes affecting the specificity of the pollen or pistil protein, e.g. in the *Solanum chacoense* pistil gene (Matton et al., 1999), and the *Brassica oleracea* pollen gene (Chookajorn et al., 2004). These investigations showed that the substitution of just a few amino-acids in the pollen or pistil sequences may be sufficient to alter specificity. Yet there is no plausible evolutionary scenario for the emergence of new SI haplotypes, because, to create a new functional SI haplotype, the two genes must both change appropriately. The large number of SI haplotypes typically found in SI species nevertheless shows that such diversification occurred repeatedly.

Three evolutionary models have so far been proposed to explain how new SI haplotypes can arise. Matton et al. (1999) proposed a verbal model based on the observation in experiments with *S. chacoense* of a dual-specificity artificial chimeric S-RNase protein (produced by the style gene) able to reject two distinct pollen specificities. Their model considered a mutation in the style gene conferring dual specificity, recognising a new pollen type, not yet present, as well as its initial pollen type. The pollen gene of this haplotype might then mutate to acquire the new specificity, and the ancestral pistil recognition type might be lost, resulting in a new functional SI haplotype. This scenario



was criticized because the new pollen allele would transiently be rejected by two different haplotypes, placing it at a selective disadvantage when challenged against its ancestral non-mutated copy (Uyenoyama and Newbigin 2000, Charlesworth 2000), and is thus expected to be lost from the population. Furthermore, this scenario involves the unlikely occurrence of three consecutive mutations on the same haplotype (Charlesworth, 2000).

A second verbal model is based on the observation that the *B. oleracea* pollen gene can tolerate several amino acid substitutions without affecting its binding affinity or specificity (Chookajorn et al., 2004). These authors proposed that nearly-neutral standing variation exists within each functional allelic class. If, by chance, recognition became stronger among a subset of haplotypes within a class than with the other haplotypes of the same class, selection might reinforce this new interaction and reduce recognition of the remaining haplotypes in the class, eventually disrupting recognition between the two classes. This model did not specify the selective force reinforcing interactions within nascent functional classes, and did not consider inbreeding depression, the main evolutionary force maintaining and driving the evolution of SI. Self-compatible haplotypes can invade SI populations (Charlesworth and Charlesworth, 1979), but the conditions under which this model can generate new functional SI haplotypes are unclear.

The only population genetics model of new SI haplotypes involves self-compatible (SC) intermediates (Uyenoyama et al., 2001). SC pollen mutants expressing a new pollen specificity can invade populations and be stably maintained at an intermediate frequencies along with functional SI haplotypes when inbreeding depression is high, and proportion of self-pollen is low to intermediate (Charlesworth and Charlesworth 1979, Uyenoyama et al. 2001). Long-term maintenance of pollen mutants was proposed to allow diversification, with new SI haplotypes arising through a single compensatory mutation in the pistil gene of the haplotype carrying the pollen mutant. However, the haplotype in which the pollen mutant arose is almost always excluded at equilibrium. Thus, even though new SI haplotypes eventually appear, this almost always involved loss of their ancestral haplotype, resulting in replacement of SI haplotypes rather than diversification, suggesting that an increased number of SI haplotypes is almost impossible in a panmictic population. The authors therefore suggested that diversification might occur in subdivided populations, but did not model subdivision.

We extend the analytical model of Uyenoyama et al. (2001) by studying the conditions for the evolution of new SI haplotypes, and also by simulations. We first study the conditions under which diversification at the S-locus is expected to occur in an infi-

nite population, in terms of the proportion of self-pollen deposited, and the inbreeding depression, and then investigated the effect of genetic drift using simulations including recurrent mutations for new pollen and pistil specificities.

## 3.2 Models

We assumed gametophytic self-incompatibility (GSI), i.e. pollen specificities determined by the pollen's haploid genome, and style specificities determined co-dominantly by the diploid genome of the style. Pollinations are compatible if the pollen specificity of the donor differs from both specificities expressed by the style of the recipient (de Nettancourt, 2001). Following Uyenoyama et al. (2001), we assumed two completely linked genes,  $A$  determining the style type, and  $B$  the pollen specificity. Alleles  $A_i$  and  $B_j$  segregate at the  $A$  and  $B$  genes, and express specificities  $i$  and  $j$ . Functionally incompatible haplotypes carry identical specificities at the two genes, e.g.  $A_iB_i$ , and we denote haplotypes by  $S_i$  (Table 3.1 gives our notation). Under these assumptions, a new SI haplotype  $A_uB_u$  requires two mutations, one in each gene.

TABLE 3.1 – Genotypes and associated phenotypes produced in a population composed of  $n - 1$  SI haplotypes  $S_i$ , the ancestral haplotype  $S_n$  and the different studied mutants (the pollen-part mutant  $S_b$  and the compensated mutant  $S_{n+1}$ ).

Genotype	Frequency	Rejected haplotypes	Proportion of compatible pollen	Selfing rate
$S_iS_j$	$x_{ij}$	$S_i, S_j$	$N_{ij}$	0
$S_bS_b$	$x_{bb}$	$S_n$	$N_{bb}$	$s_{bb}$
$S_bS_n$	$x_{bn}$	$S_n$	$N_{bn}$	$s_{bn}$
$S_bS_i$	$x_{bi}$	$S_n, S_i$	$N_{bi}$	$s_{bi}$
$S_nS_i$	$x_{ni}$	$S_n, S_i$	$N_{ni}$	0
$S_bS_{n+1}$	$x_{bn+1}$	$S_b, S_n, S_{n+1}$	$N_{bn+1}$	0
$S_nS_{n+1}$	$x_{nn+1}$	$S_b, S_n, S_{n+1}$	$N_{nn+1}$	0
$S_iS_{n+1}$	$x_{in+1}$	$S_b, S_{n+1}, S_i$	$N_{in+1}$	0

### 3.2.1 Infinite population model

In the infinite population model, the starting population had equal frequencies of  $n$   $S_i$  haplotypes ( $1 \leq i \leq n$ ). We modelled inbreeding depression by assuming that each plant received a proportion  $\alpha$  of self-pollen (rate of self-pollen deposition) and that selfed offspring have a probability  $1 - \delta$  of reaching maturity, relative to outcrossed offspring. The average fitness of the population was thus

$$\bar{W} = 1 - \delta \sum_{uv} s_{uv} x_{uv}, \quad (3.1)$$

with  $x_{uv}$  the frequency of genotype  $S_u S_v$  (see Table 3.1) and  $s_{uv}$  its selfing rate. We sequentially introduced different mutants and followed their fate in the population. Because Uyenoyama et al. (2001) showed that diversification can proceed only through pollen-part mutants, we give only the results when the first mutation affects the pollen type.

**Pollen-part mutation (first mutation).** Using Mathematica 7.0 (Wolfram Research, 2008), we first investigated the fate of a mutation in the pollen component of haplotype  $S_n$ , generating haplotype  $S_b = A_n B_{n+1}$ . Following the introduction of one such mutation, the population has  $n + 1$  distinct haplotypes:  $n - 1$  unaffected haplotypes  $S_i$  ( $1 \leq i < n$ ), the ancestral haplotype  $S_n$ , and the pollen-part mutant  $S_b$ , with respective frequencies in the pollen pool  $p$ ,  $p_n$  and  $p_b$ . The population has  $n(n - 1)/2 + n - 1 + 1$  diploid genotypes, the  $n(n - 1)/2$   $S_i S_j$  genotypes (fully SI when  $1 \leq i, j \leq n$ ), the  $n$   $S_i S_b$  genotypes (partially self-compatible when  $1 \leq i \leq n$ ), and the  $S_b S_b$  genotype (which can self-fertilize, but rejects  $S_n$  pollen). Assuming equal pollen production among diploid genotypes and given the haplotypes rejected by each genotype (Table 3.1), the proportions of compatible pollen received by each genotype are

$$\begin{aligned} N_{ij} &= 1 - 2p \\ N_{bb} = N_{bn} &= 1 - p_n \\ N_{bi} = N_{ni} &= 1 - p_n - p. \end{aligned} \quad (3.2)$$

where the double subscripts indicate the diploid genotypes ( $S_i S_j$ ,  $S_b S_b$ ,  $S_b S_n \dots$ ).

Hence, the selfing rates of SC genotypes are

$$\begin{aligned}
 s_{bb} &= \frac{\alpha}{\alpha + (1 - \alpha)N_{bb}} \\
 s_{bn} &= \frac{\alpha}{\alpha + 2(1 - \alpha)N_{bn}} \\
 s_{bi} &= \frac{\alpha}{\alpha + 2(1 - \alpha)N_{bi}}.
 \end{aligned} \tag{3.3}$$

The factor of 2 in the denominators of  $s_{bn}$  and  $s_{bi}$  arises because only half of the pollen of self-compatible (SC) heterozygotes is self-compatible.

Using equations 3.1, 3.2 and 3.3, we computed the recursion equations describing the genotypic frequency change in a generation, which allowed us to investigate the evolutionary fate of all haplotypes in the population, and in particular the pollen-part mutant  $S_b$ . The resulting recursion equations are given in Appendix A1.

**Compensatory (second) mutation.** We next introduced a mutation in the pistil gene of the pollen-part mutant  $S_b$ , generating a fully functional new SI haplotype, denoted by  $S_{n+1} = A_{n+1}B_{n+1}$ . Let  $p_{n+1}$  be the frequency of  $S_{n+1}$  in the pollen pool. In the population,  $n(n+1)/2 + n + 1 + 1$  diploid genotypes can now be formed, the  $n(n+1)/2$   $S_iS_j$  genotypes (fully SI if  $1 \leq i, j \leq n+1$ ), the  $n$  partially self-compatible  $S_iS_b$  genotypes ( $1 \leq i \leq n$ ), the genotype  $S_bS_{n+1}$  (which is fully self-incompatible and rejects the three haplotypes  $S_n$ ,  $S_b$  and  $S_{n+1}$ ), and the  $S_bS_b$  genotype, which can self but rejects  $S_n$  pollen. The proportions of compatible pollen for each genotype are

$$\begin{aligned}
 N_{ij} &= 1 - 2p \\
 N_{bb} = N_{bn} &= 1 - p_n \\
 N_{bi} = N_{ni} &= 1 - p_n - p \\
 N_{bn+1} = N_{nn+1} &= 1 - p_b - p_n - p_{n+1} \\
 N_{in+1} &= 1 - p_b - p_{n+1} - p.
 \end{aligned} \tag{3.4}$$

The selfing rates  $s_{bb}$ ,  $s_{bn}$  and  $s_{bi}$  defined in equation 3.3 remain unchanged, and the recursion equations needed to investigate the evolutionary fate of the compensated haplotype  $S_{n+1}$  and its SC ancestral  $S_b$  are given in Appendix A2.

**Evolution of the system.** For diversification to occur, four conditions are clearly necessary (Uyenoyama et al., 2001) :

- (i) The pollen-part mutant  $S_b$  must increase in frequency when rare.
- (ii)  $S_b$  must not fix (*i.e.* a functional SI system must be retained).
- (iii)  $S_n$  (the ancestral allele from the pollen-part mutant arises) must not be excluded by the introduction of  $S_b$ .
- (iv) The mutation in the pistil gene that creates the new fully functional SI haplotype,  $S_{n+1}$ , must increase in frequency when both  $S_b$  and  $S_n$  are present in the population.

We determined parameter values under which these conditions were jointly satisfied. Conditions (i) and (ii) were investigated using local stability analysis. For condition (i), we studied invasion of a population with  $n$   $S_i$  haplotypes by  $S_b$ , and for condition (ii) we studied the stability of the equilibrium with  $S_b$  fixed. We could not use this approach to investigate conditions (iii) and (iv) because we found no tractable analytical solution for the equilibrium when  $n - 1$   $S_i$  and  $S_b$  were all present in the population. We therefore used recursion equations A1 and A2 to numerically approximate the parameter values satisfying these conditions. For each of 5 to 10 values of the self-pollen deposition parameter ( $\alpha$ ), we varied inbreeding depression ( $\delta$ ) values in steps of  $10^{-3}$ , and followed the population for 10,000 generations. Haplotypes with frequencies less than  $10^{-5}$  were considered lost from the population. Specifically, for condition (iii) we introduced  $S_b$  at low frequency ( $x_{bi} = 2 \cdot 10^{-5}$ , *i.e.*  $p_b = 10^{-5}$ ) into an initially isoplethic population (with all functional SI haplotypes equally frequent), and determined the value of  $\delta$  for which  $S_n$  was not eliminated when  $S_b$  spread. For condition (iv), we started with a population at equilibrium with  $n$   $S_i$  haplotypes plus  $S_b$ , as obtained under condition (iii), and introduced the  $S_{n+1}$  haplotype at low frequency ( $x_{in+1} = 2 \cdot 10^{-5}$ , *i. e.*  $p_{n+1} = 10^{-5}$ ), to determine the critical  $\delta$  value for  $S_n + 1$  to increase in frequency.

### 3.2.2 Finite populations simulation

We assumed a population of  $N$  diploid individuals in which all adults died after zygote production. The initial population contained  $n$  different SI haplotypes  $S_i = S_{ii}$  in equal frequencies. We assumed a maximum of  $k$  different specificities ( $1 \leq i \leq k$ ). The life cycle included three steps : mutation, fertilization during which selection on S-alleles during pollination plays, and viability selection through inbreeding depression.

**Mutation.** The number of mutations occurring each generation ( $U$ ) was randomly drawn from a Poisson distribution with parameter  $4N\mu$ , where  $\mu$  is the mutation rate per gene per generation. The factor 4 arises because mutations arise in both the pollen and pistil genes of the  $2N$  chromosomes in the population.  $U$  chromosomes were randomly drawn with replacement in the population. For each chromosome, the pollen or pistil gene was mutated with equal probability, and the specificity was randomly changed to one of the remaining  $k - 1$  specificities. Hence, unlike the infinite population model, the order of mutations is not constrained.

**Pollination.** Plants were assumed to produce infinite numbers of pollen grains and ovules. Hence we assume no pollen limitation. Diploid individuals were randomly drawn with replacement as maternal plants. Let us consider a plant with genotype  $S_{ij}S_{gh}$  which expresses specificities  $i$  and  $g$  in pistil and  $j$  and  $h$  in pollen, with  $\{g, h, i, j\} \in \llbracket 1, k \rrbracket^4$ . The probability for this plant to self-fertilize was

$$s_{ij,gh} = \frac{\alpha(\gamma_{ij}\gamma_{gj} + \gamma_{gh}\gamma_{ih})}{\alpha(\gamma_{ij}\gamma_{gj} + \gamma_{gh}\gamma_{ih}) + 2(1 - \alpha)\sum_{m,n} p_{mn}\gamma_{in}\gamma_{gn}}, \quad (3.5)$$

where  $\alpha$  is the proportion of self pollen received,  $p_{mn}$  the frequency of the  $S_{mn}$  haplotype in the population, and  $\gamma_{ij}$  is an indicator variable such that  $\gamma_{ij} = 0$  if  $i = j$  (i.e.  $B_j$  pollen is rejected by  $A_i$  pistils) and  $\gamma_{ij} = 1$  if  $i \neq j$  (compatible mating).

If the plant self-fertilizes, it could be fertilized by  $S_{ij}$  with probability  $f_{ij} = \gamma_{ij}\gamma_{gj}/(\gamma_{ij}\gamma_{gj} + \gamma_{gh}\gamma_{ih})$  or by  $S_{gh}$  with probability  $f_{gh} = 1 - f_{ij}$ . Otherwise, the probability that it was fertilized by a pollen with haplotype  $S_{uv}$  was

$$f_{uv} = \frac{p_{uv}\gamma_{iv}\gamma_{gv}}{\sum_{m,n} p_{mn}\gamma_{in}\gamma_{gn}}. \quad (3.6)$$

**Inbreeding depression.** Finally, to form the zygote genotype, a maternal chromosome was drawn randomly with probability  $1/2$ . Selfed seeds survived with probability  $1 - \delta$ .  $N$  individuals were generated.

**Diversification, loss or maintenance of SI.** We first simulated an initial population of  $n = 5$  different SI haplotypes in equal frequencies and  $k = 20$ . We simulated four population sizes ( $N = 50, 500, 5000$  and  $50000$ ) and five mutation rates ( $\mu = 5.10^{-7}, 5.10^{-6}, 5.10^{-5}, 5.10^{-4}$  and  $5.10^{-3}$ ). We defined four possible different final states of the

population :

- (1) Loss of SI : only SC haplotypes present in the population.
- (2) Maintenance of SI : no SC haplotypes present and the number of SI haplotypes remained stable or decreased ( $n \leq 5$ ).
- (3) Polymorphism : both SI and SC haplotypes present at equilibrium.
- (4) Diversification :  $n > 5$  SI haplotypes and no SC haplotypes present.

One hundred runs were performed for each set of parameters. The simulations were stopped when either diversification occurred (more than five SI haplotypes and no SC haplotypes in high frequencies), or all SI haplotypes were lost. Otherwise (no new SI haplotypes and/or presence of SC haplotypes at high frequencies), the system was run for  $10^5$  generations. Since we were interested in the final state of the population, and since SC and SI haplotypes could arise by mutation at any time, we then arbitrarily characterized the state of the system based on haplotypes present at frequencies  $> 0.01$  (0.05 for  $N = 50$ ).

**Dynamics of diversification.** In the previous section we were interested only in the final evolutionary outcome. To investigate the numbers of SI haplotypes over time, and the times between successive diversification events, we performed further simulations for  $10^6$  generations with  $N = 5000$ ,  $\mu = 5 \cdot 10^{-5}$ ,  $\delta = 0.9$  and  $\alpha$  values of 0.2 and 0.4. Runs were started at isoplethy with three SI haplotypes, and we studied three  $k$  values, 20, 100 and 200. One hundred runs were performed for each set of parameters.

To investigate the detailed dynamics of SI haplotype gains and losses, we also tracked the genotypes in the population for the 150,000 generations for 20 replicates, sampling genotypic frequencies every 1,000 generations, using the following approximations. Since new SI haplotypes necessarily arise through SC intermediates, their rate of gain should be approximated by the population's total frequency of SC haplotypes. Similarly, since in our model the loss of haplotypes is caused by genetic drift of low frequency haplotypes, and all functional SI haplotypes in GSI systems have equal expected equilibrium frequencies, the rate of loss should be approximated by the average frequency of SI haplotypes.

## 3.3 Results

### 3.3.1 Infinite populations

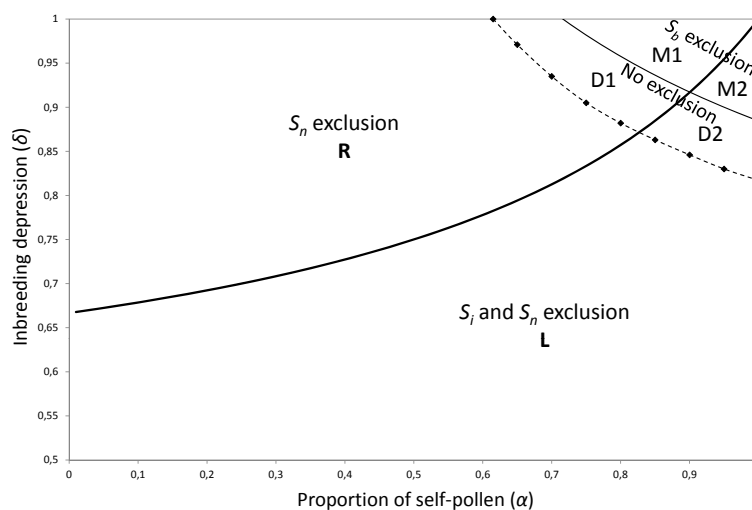
**Evolutionary outcomes.** Extending the stability analysis of the Uyenoyama et al. (2001) model (Appendix A3) allowed us to divide the parameter space into six regions (L, M1, M2, R, D1, D2, see Figure 3.1a), instead of the four originally described (Table 3.2). Uyenoyama et al. (2001)’s four regions were defined based only on the frequency change of the pollen-part mutant  $S_b$  when rare or frequent. Our inclusion of the fate of the ancestral SI haplotype  $S_n$  splits two of the regions where  $S_b$  increases in frequency when rare, depending whether rare  $S_n$  (compensating) mutations increase or decrease. The six resulting regions correspond to four qualitatively different evolutionary patterns for  $S_b$ ,  $S_n$  and  $S_{n+1}$  haplotypes. Three evolutionary outcomes were identified by Uyenoyama et al. (2001) : in region L, the SI system was lost, whereas it was maintained unaltered in regions M1 and M2, or haplotypes were replaced (region R). Our analysis identifies additional regions D1 and D2 in which new SI haplotypes evolve. We next describe these regions in detail.

TABLE 3.2 – Evolution of the frequency of  $S_b$  near fixation and of  $S_b$ ,  $S_n$  and  $S_{n+1}$  near exclusion, in the six described regions. A “+” indicates that the frequency increases. A “-” indicates that the frequency decreases. Regions M1, M2, R+D1 and L+D2 correspond respectively to the regions E, D, P and S described in Uyenoyama et al. (2001).

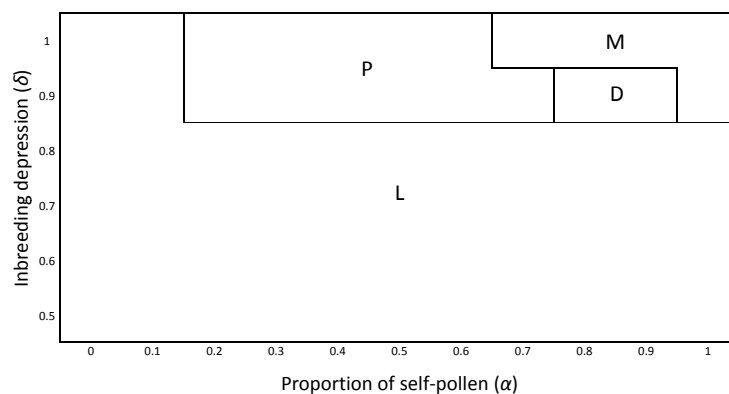
	$S_b$ frequent	$S_b$ rare	$S_n$ rare	$S_{n+1}$ rare
L	+	+	-	-
R	-	+	-	+
M1	-	-	+	+
M2	+	-	+	+
D1	-	+	+	+
D2	+	+	+	+

Low inbreeding depression yields outcomes in region L (unconditional increase in the frequency of  $S_b$ , to fixation excluding all functional haplotypes, with loss of SI). With high inbreeding depression and a high proportion of self-pollen deposition, regions M1 and M2 result ( $S_b$  does not increase in frequency when introduced at low initial frequency and an unaltered SI is maintained; the two regions differ in the behaviour





(a) In infinite population



(b) In a finite population

**FIGURE 3.1** – Limits of the regions where the different evolutionary outcomes are observed, for an initial number of S-haplotypes  $n = 5$ . (a) In infinite population : plain lines are the results of the stability analysis (thick line : fate of the pollen part mutant  $S_b$  when it is frequent ; thin line : evolutionary outcome when  $S_b$  is rare ; see text for further details). The dashed line limit was obtained through numerical iterations and delimits the region where haplotype  $S_n$  was not excluded after the introduction of the pollen-part mutant  $S_b$ . Diamonds refer to the parameters that we investigated to approximate this limit. We define four regions according to which haplotypes are excluded when  $S_b$  is introduced at low frequency in the population. These four regions are : the loss region (L), when the pollen-part mutant  $S_b$  goes to fixation ; the replacement region (R), when only  $S_n$  is excluded ; the maintenance regions (M1 + M2), when  $S_b$  alone is excluded ; the diversification regions (D1 + D2), when no haplotypes are excluded, a new S-haplotype can emerge. (b) In a finite population ( $N = 5000$ ,  $\mu = 5.10^{-7}$ ,  $k = 20$ ) : we defined four regions according to which evolutionary outcome was the most frequent in our simulations for a given set of parameters : the loss region (L, where loss of the SI system was mainly observed) ; the polymorphism region (P, where the populations were composed of a mix of SC and SI haplotypes) ; The maintenance region (M, where only SI haplotypes were observed and the number of specificities did not increase) ; the diversification region (D, where no SC haplotypes were observed and the number of SI haplotypes increased).

of  $S_b$  near fixation, effectively fixing in M2, but decreasing in frequency in M1). When inbreeding depression is high and self-pollen deposition low to intermediate, region R results. In this region of parameter space,  $S_b$  establishes a polymorphism along with the pre-existing  $S_i$  haplotypes, but this is always accompanied by the exclusion of the ancestral  $S_n$  haplotype. Subsequent introduction of  $S_{n+1}$  leads to the exclusion of  $S_b$ , resulting in a fully SI population with the  $n - 1$  initial  $S_i$  haplotypes and the new haplotype  $S_{n+1}$  replacing  $S_n$ , with no increase in the number of SI haplotypes.

Finally, for intermediate to high values of both  $\delta$  and  $\alpha$ , diversification is possible. Such situations yield regions D1 and D2, in which  $S_b$  also establishes a polymorphism at equilibrium (like region R), but without loss of  $S_n$  ( $S_b$  fixes in region D2, but in D1 it decreases in frequency when initially present at high frequency). The frequency of  $S_b$  has one stable equilibrium in region D1, and two equilibria, one stable and one unstable, in D2 (Table 3.2); the stable equilibrium in D2 has a lower  $S_b$  frequency than the unstable one. Because  $S_n$  is also maintained in regions D1 and D2, the compensatory pistil mutation  $S_{n+1}$  can produce an increase the number of SI haplotypes. Indeed, the  $S_{n+1}$  mutation always led to the exclusion of  $S_b$  (see condition (iv) in Appendix A3).

The relative sizes of the different regions depended on the initial number of SI haplotypes ( $n$ ), with an increase of the M regions and a decrease of the R and D regions as  $n$  increases (Figure 3.2). This relates to the possibility of polymorphism for pollen-part mutation until  $n$  becomes large (Charlesworth and Charlesworth, 1979).

### 3.3.2 Finite population simulations

Our simulations starting with  $n = 5$  support the analytical predictions, but genetic drift and recurrent pollen-part and compensatory pistil mutation can increase the parameter space where diversification can occur. Fig. 1b shows four different evolutionary outcomes : region L, where SC haplotypes became fixed, M, where SI was maintained, a polymorphism region (P), where both SC and SI haplotypes are present, and a diversification region (D), where the final population contained more than the initial number of SI haplotypes, and no SC haplotypes. Note that Figure 3.1b is a simplified representation of the results for  $N = 5000, \mu = 5.10^{-7}, k = 20$  (detailed results are shown in Figures 3.3, S1, S2 and S3).

Whereas the relative positions of these different zones are largely independent of the parameter values, their sizes depend strongly on the population size and mutation rate.

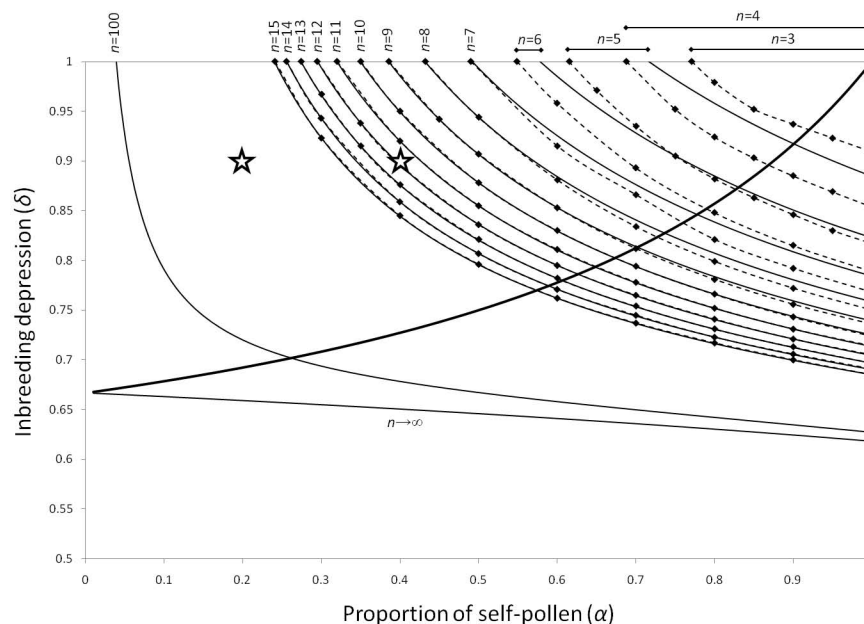


FIGURE 3.2 – Effect of the initial number of specificities ( $n$ ) on the different evolutionary outcomes in an infinite population. Plain and dashed lines define the limits of each region (L, R, M1, M2, D1 and D2, see legend of Figure 3.1). The diversification region is indicated on top of the panel for the different  $n$ . Note that some diversification regions are overlapping. The area of diversification region decreases when  $n$  increased, except for  $n = 3$  and  $n = 4$ . The diversification region was also displaced toward smaller values of  $\delta$  and  $\alpha$  when  $n$  increased. Stars indicate the values of  $\delta$  and  $\alpha$  that were chosen to investigate the diversification dynamics across  $10^6$  generations in Figure 3.4.

Except when the values of  $\mu$  and  $N$  were both low, region D was considerably larger than in the analytical model (Figure 3.3), extending over much of the R region determined for an infinite population. Indeed, for high  $N\mu$ , diversification occurred even with high inbreeding depression and self-pollen deposition (whereas, with low  $N\mu$ , the system was maintained, as in an infinite population, see Figure S1). The parameter values for which SI was lost were similar to those for an infinite population (Figure S2), except for high inbreeding depression and low self-pollen deposition where SI is lost in finite populations whereas replacement is expected in infinite population. As expected from the fact that a compensated mutant  $S_{n+1}$ , always excludes its SC ancestor  $S_b$ , coexistence of SI and SC haplotypes was very rare, observed mainly with low mutation rates and intermediate  $N$  (Figure S3).

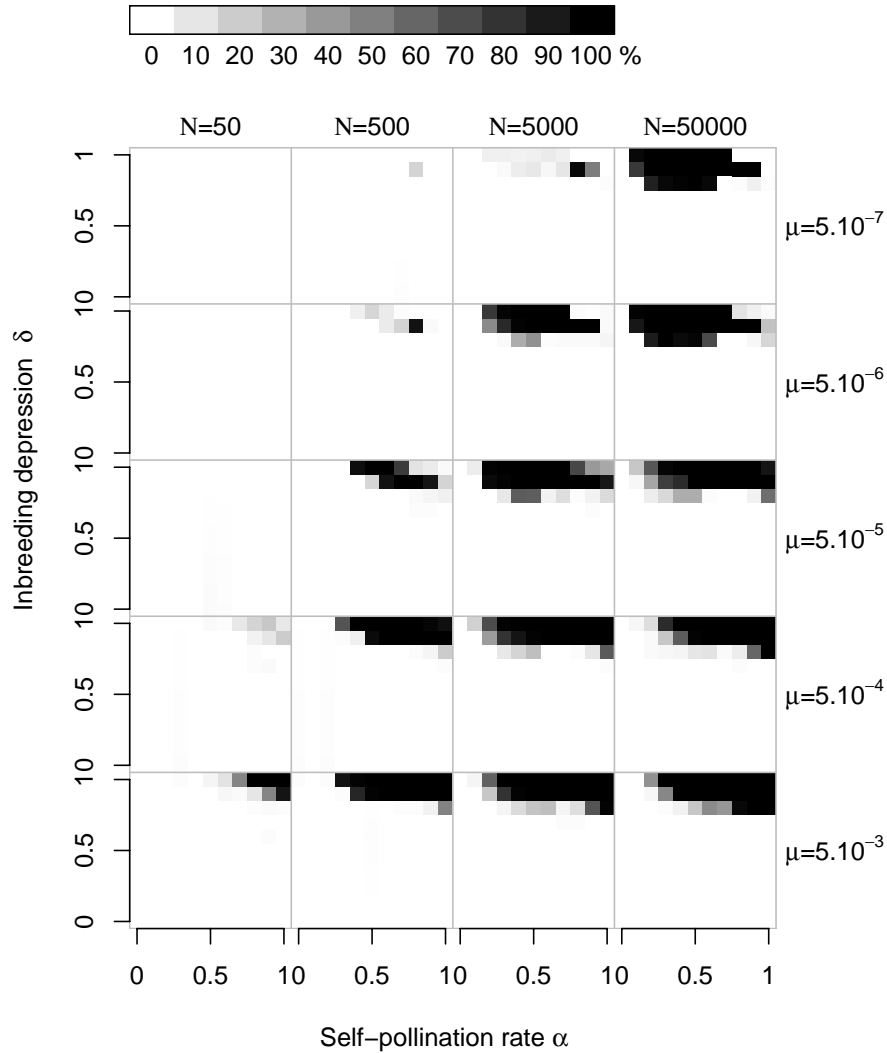
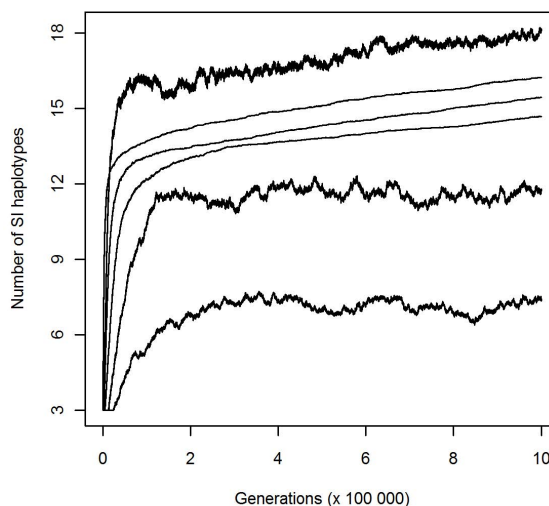


FIGURE 3.3 – The diversification region in a finite population for different population sizes and mutation rates, with an initial number of specificities  $n = 5$ . Grid cells are figured in shades of grey that are proportional to the number of simulations where diversification was observed ( $n > 5$  at the end of the simulation), with black squares corresponding to sets of parameters under which the evolutionary outcome was only diversification and white squares to cases under which no diversification was observed. A hundred replicate runs were performed for each combination of parameter values (population size  $N$ , mutation rate  $\mu$ ).  $k = 20$  for all simulations.

**Dynamics of diversification.** With  $n = 3$  in the initial population, there was generally loss of some SI haplotypes in the early generations, since, with such low numbers of SI haplotypes, an  $S_b$  pollen-part mutant is expected to exclude its ancestral SI haplotype  $S_n$  (see Figure 3.2). However, the populations rapidly returned to  $n > 3$ . For the parameter values investigated ( $N = 5000$ ,  $\mu = 5.10^{-5}$ ,  $\delta = 0.9$ ,  $\alpha = 0.2$  and  $0.4$ ), the numbers of SI haplotypes in the populations increased from 3 to between 7 and 18 within  $10^6$  generations (Figure 3.4). For  $\alpha = 0.2$ , the population had apparently reached a stationary number of haplotypes. Diversification always proceeded through pollen-part mutants followed by compensatory mutations in the pistil gene, confirming for finite populations this prediction from the analytical model (Uyenoyama et al., 2001).

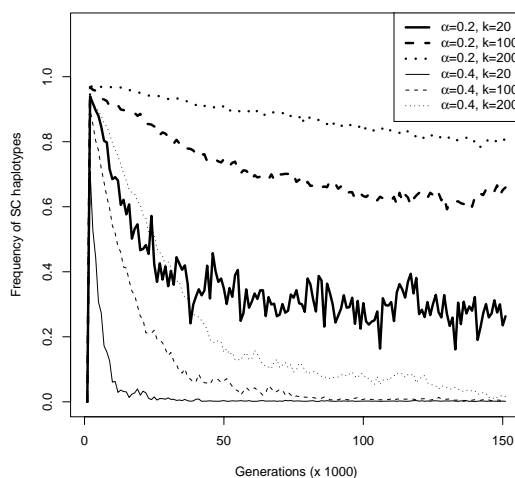


**FIGURE 3.4** – Diversification dynamics in finite populations across  $10^6$  generations. The number of SI haplotypes in the population was averaged over a hundred replicates. Simulations were performed with  $N = 5,000$ ,  $\mu = 5.10^{-5}$ ,  $\delta = 0.9$ , the initial number of SI haplotypes  $n = 3$  and for three different values of the maximum number of specificities ( $k$ ) and two values of the proportion of self-pollen ( $\alpha$ ). Thick line :  $\alpha = 0.2$ , thin line :  $\alpha = 0.4$ . For a given  $\alpha$ , curves correspond to  $k = 20, 100$  and  $200$  from top to bottom.

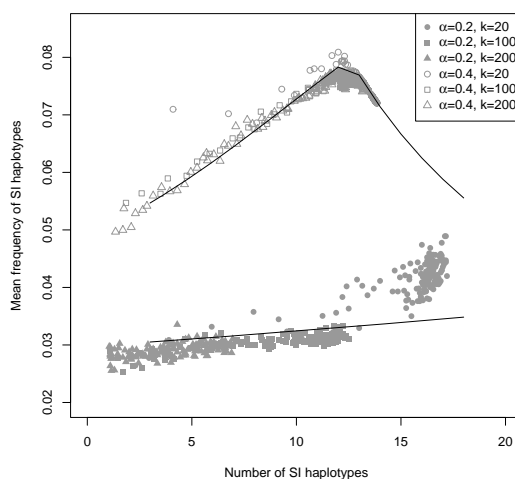
The mean waiting time between successive diversification events (see Methods for the approximations used to study this) increased with  $n$ , with two distinct phases. Haplotype numbers first increased rapidly and then suddenly either stabilized or increased much more slowly. In the initial fast diversification, the overall frequency of SC haplo-

types was very high (Figure 3.5a), implying a high birth rate of SI haplotypes, but the overall frequency of SC haplotypes then decreased, thus slowing down diversification as  $n$  increased (Figure 3.5a). Interestingly, during the same time, the rate of loss of functional SI also decreased (as estimated from the increased mean frequency of SI haplotypes, see Methods; Figure 3.5b). This suggests that the decrease of the birth rate more than outweighs the decreased death rate, and therefore that the slowdown of diversification is due to the decreased birth rate. This is consistent with the fact that the ability of pollen-part mutations to invade decreases when  $n$  is high.

SI haplotype diversification rate increased with decreasing  $k$ , especially during the first phase (Figure 3.4), which explains the lower number of SI haplotypes at a given time. We interpret this in terms of the increased probability  $1/(k - 1)$  that compensatory mutations can occur as  $k$  decreases. However, the value of  $k$  should not affect diversification probability for given values of  $\alpha$  and  $\delta$ . In other words, we do not expect that the parameter space for diversification (Figure 3.3) would shrink as  $k$  increases. The detailed dynamics depends on the level of self-pollen deposition. With high self-pollen deposition ( $\alpha = 0.4$ ), the overall frequency of SC haplotypes was near zero (Figure 3.5a), implying a very low rate of gain of new SI haplotypes, and the frequency of SI haplotypes was high (Figure 3.5b), suggesting a low probability of haplotype loss by genetic drift. Diversification therefore occurred with slow haplotype turnover, with both birth and death rates were low. For lower proportions of self-pollen ( $\alpha = 0.2$ ), however, SC haplotypes were at higher frequency (Figure 3.5a) but SI haplotypes were less frequent (Figure 3.5b), i.e. diversification occurred with fast haplotype turnover, and high birth and death rates. This is consistent with the observation of more stochastic variation in the number of haplotypes in the population for the lower  $\alpha$  value (Figure 3.4), and with additional simulations with larger population size, up to  $N = 10000$  (data not shown), which did not detectably affect the final number of haplotypes when  $\alpha = 0.4$  (i.e. there was no effect of genetic drift on the dynamics of diversification) but did increase the final number of haplotypes for  $\alpha = 0.2$ .



(a) Mean cumulated frequency of all SC haplotypes as a function of time.



(b) Mean frequency of SI haplotypes as a function of their total number.

**FIGURE 3.5** – Genotypic composition of the population in the first 150,000 generations to investigate the dynamics of birth and death of SI haplotypes (genotypic frequencies sampled every 1000 generations in 20 replicates for  $N = 5000$ ,  $\mu = 5.10^{-5}$  and  $\delta = 0.9$ ). (a) Mean cumulated frequency of all SC haplotypes as a function of time. After an initial rapid increase in the total frequency of SC haplotypes, their frequency started to decrease as the number of SI haplotypes increased (see Figure 3.4). (b) Mean frequency of SI haplotypes as a function of their total number. Symbols are mean frequencies in simulations and lines are the expected frequencies at deterministic equilibrium in an infinite population (computed using equations A1 with a frequency of  $S_n p_n = 0$  and setting  $n = x + 1$  when  $x$  SI haplotypes were segregating within the population). For both values of  $\alpha$ , the mean frequency of SI haplotypes increased with their total number. This is due to the fact that when the number of SI haplotypes increases, the total frequency of SC haplotypes decreases, thus leaving a larger range of frequency for SI haplotypes. When  $\alpha = 0.4$ , the mean frequency of SI haplotypes decreased when their number was larger than 12 because the frequency of SC haplotypes approached zero. There was no similar threshold value of the number of SI haplotypes in the case  $\alpha = 0.2$  because the frequency of SC haplotypes remained high.

## 3.4 Discussion

### 3.4.1 Importance of genetic drift and recurrent mutations.

Allelic diversification is clearly possible in an infinite, undivided population, and the region of parameter space where it can occur can be large, especially when the number of SI haplotypes in the population is  $< 6$ . Our simulations show that, in finite populations, diversification is generally possible under a larger set of parameter values than in an infinite population (with the exception of very small populations and/or very low mutation rates). The difference was largely dependent on the proportion of self-pollen deposition, low values of which can greatly extend the lower limit of the diversification region, so that it could cover most of the region where replacement of SI haplotypes, without diversification, is expected in an infinite population. In contrast, the critical inbreeding depression allowing diversification in finite populations never extended beyond that for an infinite population, for the parameter space we explored (even for the largest population sizes and mutation rates). Hence, while stochastic processes are generally associated with reduced genetic diversity within small, finite populations, our results show that diversification of SI haplotypes behaves in the opposite way, and that the forces promoting diversification can outweigh stochastic loss of SI haplotypes, which becomes strongly limited due to frequency-dependent selection under conditions when SI is favored.

### 3.4.2 The role of inbreeding depression.

Our results support previous conclusions that diversification requires high inbreeding depression (Charlesworth and Charlesworth, 1979 ; Uyenoyama, 1988a,b). However, these conclusions were based on a fixed inbreeding depression parameter, ignoring its evolution along as the outcrossing rate changes. Porcher and Lande (2005) showed that, for high inbreeding depression due to deleterious mutations, and high selfing rates, mutations may be efficiently purged, eventually leading to the loss of SI. Our study again treated inbreeding depression as a fixed parameter, even though clearly invasion of a population by an  $S_b$  allele may allow purging of deleterious mutations, making the conditions for diversification increasingly stringent. However, Porcher and Lande (2005) also showed, in single-locus models for S-alleles, that SC haplotypes can be maintained in coexistence with SI haplotypes under a wide range of parameter values, especially when



selfing rate is low or when there is a sheltered load. Moreover, our results showed that inbreeding depression must be high for diversification to occur, while the self-pollination rate may be moderate to high. This should now be added to further studies of two-gene models for S-allele evolution.

### 3.4.3 Dynamics of diversification.

The observed increase in the time between successive diversification events may be due to the fact that, for a given combination of self-pollen deposition and inbreeding depression, the expected frequency of pollen-part intermediates decreases as  $n$  increases, which decreases the probability of a compensatory (pistil) mutation occurring in the intermediate haplotype before it is lost by drift. The prediction that diversification is not constant through time may explain some otherwise puzzling features of SI systems, including cases of rapid S-allele diversification, and the unexpectedly long terminal branches typically observed at SI genes.

### 3.4.4 S-allele numbers and the shape of phylogenies.

The observation that all SI haplotypes in *Physalis* and *Witheringia* belong to only three ancestral lineages (Paape et al. 2008; see also Miller et al. 2008 for the genus *Lycium*), was interpreted by these authors in terms of rapid S-allele diversification after demographic bottlenecks (when the number of SI haplotypes will tend to be small). Further, Castric and Vekemans (2007) estimated much stronger positive selection among *Brassica* SI haplotypes than among the more anciently diverged SI haplotypes of *Arabidopsis* species. This might suggest rapid and recent diversification of *Brassica* SI haplotypes from just two ancestral lineages.

The parameter space where diversification is expected to occur in an infinite population shrinks rapidly with increased  $n$ , and the D region is also displaced towards lower self-pollen deposition rates and inbreeding depression. Initially rapid diversification becomes slow as SI haplotype number increases, eventually stopping almost completely, yielding a threshold value of  $n$  that allows diversification. For  $\alpha = 0.2$ , haplotype turnover was rapid during diversification, and the overall frequency of SC haplotypes remained high when diversification ceased.

Our simulations yield many fewer SI haplotypes than are typically observed in natural populations, where the number of SI haplotypes commonly reaches 50-100 in species-wide

range samples (Lawrence, 2000). This is probably not due to the low effective population size assumed in our simulations, because, at least for high proportions of self-pollen ( $\alpha = 0.4$ ), the SI haplotype number maintained was apparently independent of population size (results not shown). Moreover, under the classical single-gene mutation-drift-selection theoretical framework where a single mutation gives rise to a new functional SI haplotypes, only a very low mutation rate is needed to yield high equilibrium numbers of SI haplotypes in a population of 5,000 plants. In such models, to yield as few as 17 SI haplotypes at equilibrium (the maximum number of SI haplotypes we observed in our simulations), an unrealistically low mutation rate of  $10^{-20}$  must be assumed (Vekemans and Slatkin, 1994). Thus, although our model can generate diversification, it does not fully explain S-allele polymorphism, suggesting that other factors, such as population structure, are probably needed to explain the diversity observed, as proposed by Uyenoyama et al. (2001).

Allelic genealogies of genes under balancing selection extend deep in the evolutionary past, and thus can reveal ancient evolutionary processes such as the number of breeding individuals in the species as a whole across its evolutionary history or the strength of balancing selection, but these evolutionary inferences assume a constant diversification rate (*e.g.* Vekemans and Slatkin 1994, Takahata 1990). Our results suggest that this assumption may be violated, since the diversification rate decreases with the number of functional SI haplotypes in the population. The diversification rate will not be constant if the equilibrium number of SI haplotypes has only recently been reached, and the turnover rate is slow. In contrast, if turnover is fast enough and/or haplotype number stopped increasing a sufficiently long time ago, the rate can be considered constant.

Our results are consistent with the finding that terminal branches in the phylogenies of SI haplotypes are longer than expected by models that assume that balancing selection simply extends the lengths of branches of the coalescent tree, relative to a neutral model (Takahata 1990, Uyenoyama 1997). Two main hypotheses have been proposed to explain this. First, Uyenoyama (1997) proposed the accumulation of deleterious mutations linked to individual SI haplotypes (called sheltered load); a new SI haplotype would then share linked mutations with its ancestral copy and would thus be excluded from the population. There is some evidence for existence of sheltered load (Stone 2004, Llaurens et al. 2009). Second, Schierup et al. (2001) proposed that recombination could explain long terminal branches, by creating haplotypes whose sequence are composed of fragments with different evolutionary histories, thus homogenizing sequence

divergence among them. Recombination may occur among SI haplotypes in *Arabidopsis* (Castric et al., 2010), *Petunia inflata* (Wang et al., 2001), *Prunus dulcis* (Ortega et al., 2006) and other Solanaceae and Rosaceae (Vieira et al., 2003). Our results here suggest that long terminal branches could simply be an intrinsic property of the diversification process, through the slowing down of the S-allele diversification process as SI haplotype numbers increase. Furthermore, it is striking that such a slowing down of diversification is apparently encountered in very different mechanisms such as the appearance of mating-types in fungi (May et al., 1999) and speciation (*e.g.* Morlon et al. 2010 ; Quental and Marshall 2010). We might thus speculate the intriguing possibility that such a phenomenon is a general pattern.

### 3.4.5 The mutation process.

Our hypothesis for the molecular mechanism responsible for recognition between pollen and pistil is oversimplified, particularly the limited number of possible allelic states ( $k$ -allele model) and the strict recognition reactions between pollen and pistils with cognate specificities. The pollen and pistil genes differ in size and function (Takayama and Isogai, 2005), and in the numbers of amino-acid sites evolving under positive selection. Estimates of this number are between 13 and 31 in the pistil gene S-RNase of Solanaceae (Igic et al., 2007 ; Savage and Miller, 2006 ; Takebayashi et al., 2003), versus 23 in the pollen gene SFB in *Prunus* (Nunes et al., 2006). In *Brassica*, estimates are between 26 and 44 in the pistil gene SRK and about 47 in the pollen-S gene SCR (Castric and Vekemans, 2007 ; Sainudiin et al., 2005 ; Takebayashi et al., 2003). Mutagenesis experiments suggest that several mutations are generally required to shift specificities (at least 4 amino-acids in *Solanum chacoense* S-RNases, Matton et al. 1999 ; 4 amino-acids in *Brassica oleracea* SCR, Chookajorn et al. 2004). Compensatory mutations may thus require several mutational steps, possibly making coevolution and diversification more difficult than in our model.

Now that *Arabidopsis thaliana* has been established as a model plant for mechanistic studies of SI (Nasrallah et al., 2002 ; Tsuchimatsu et al., 2010), it will be feasible to study in detail the steps required for allele diversification, at least in this plant.

## 3.5 Appendix

Here we detail the recursive equations for each genotype present in a population composed of  $n$  SI haplotypes and the different studied mutants. Definition of the different genotypes and their associated variables are given in Table 3.1.

**A1.Pollen-part mutation.** Genotypic frequencies in the next generation in a population composed of  $n - 1$  SI haplotypes  $S_i$ , the ancestral SI haplotype  $S_n$  and the pollen-part mutant haplotype  $S_b$  are

$$\begin{aligned}
 x'_{ij} &= \frac{1}{\bar{W}} \left( x_{ij} \frac{(n-3)p}{N_{ij}} + x_{bi}(1-s_{bi}) \frac{(n-2)p}{2N_{bi}} + x_{ni} \frac{(n-2)p}{2N_{ni}} \right) \\
 x'_{bb} &= \frac{1}{\bar{W}} \left( x_{bb} \left( s_{bb}(1-\delta) + (1-s_{bb}) \frac{p_b}{N_{bb}} \right) + \frac{x_{bn}}{2} \left( s_{bn}(1-\delta) + (1-s_{bn}) \frac{p_b}{N_{bn}} \right) \right. \\
 &\quad \left. + \frac{x_{bi}}{2} \left( s_{bi}(1-\delta) + (1-s_{bi}) \frac{p_b}{N_{bi}} \right) \right) \\
 x'_{bn} &= \frac{1}{\bar{W}} \left( \frac{x_{bn}}{2} \left( s_{bn}(1-\delta) + (1-s_{bn}) \frac{p_b}{N_{bn}} \right) + x_{ni} \frac{p_b}{2N_{ni}} \right) \\
 x'_{bi} &= \frac{1}{\bar{W}} \left( x_{ij} \frac{p_b}{N_{ij}} + x_{bb}(1-s_{bb}) \frac{(n-1)p}{N_{bb}} + x_{bn}(1-s_{bn}) \frac{(n-1)p}{2N_{bn}} \right. \\
 &\quad \left. + \frac{x_{bi}}{2} \left( s_{bi}(1-\delta) + (1-s_{bi}) \frac{(p_b + (n-2)p)}{N_{bi}} \right) + x_{ni} \frac{p_b}{2N_{ni}} \right) \\
 x'_{ni} &= \frac{1}{\bar{W}} \left( x_{ij} \frac{p_n}{N_{ij}} + x_{bn}(1-s_{bn}) \frac{(n-1)p}{2N_{bn}} + x_{ni} \frac{(n-2)p}{2N_{ni}} \right),
 \end{aligned}$$

where  $\bar{W}$ ,  $N_g$  and  $s_g$  are respectively defined in equations 3.1, 3.2 and 3.3, and where  $p$ ,  $p_n$  and  $p_b$ , the frequencies of  $S_i$ ,  $S_n$  and  $S_b$ , are

$$\begin{aligned}
 p &= \frac{1}{(n-1)} \left( x_{ij} + \frac{x_{bi} + x_{ni}}{2} \right) \\
 p_n &= \frac{x_{bn} + x_{ni}}{2} \\
 p_b &= x_{bb} + \frac{x_{bn} + x_{bi}}{2}.
 \end{aligned}$$

**A2. Compensatory mutation.** In the case of a population composed of  $n - 1$  SI haplotypes  $S_i$ , the ancestral SI haplotype  $S_n$ , the pollen-part mutant haplotype  $S_b$  and

the new SI haplotype  $S_{n+1}$ , the genotypic frequencies in the next generation are

$$\begin{aligned}
x'_{ij} &= \frac{1}{\bar{W}} \left( x_{ij} \frac{(n-3)p}{N_{ij}} + x_{bi} (1-s_{bi}) \frac{(n-2)p}{2N_{bi}} + x_{ni} \frac{(n-2)p}{2N_{ni}} + x_{in+1} \frac{(n-2)p}{2N_{in+1}} \right) \\
x'_{bb} &= \frac{1}{\bar{W}} \left( x_{bb} \left( s_{bb} (1-\delta) + (1-s_{bb}) \frac{p_b}{N_{bb}} \right) + \frac{x_{bn}}{2} \left( s_{bn} (1-\delta) + (1-s_{bn}) \frac{p_b}{N_{bn}} \right) \right. \\
&\quad \left. + \frac{x_{bi}}{2} \left( s_{bi} (1-\delta) + (1-s_{bi}) \frac{p_b}{N_{bi}} \right) \right) \\
x'_{bn} &= \frac{1}{\bar{W}} \left( \frac{x_{bn}}{2} \left( s_{bn} (1-\delta) + (1-s_{bn}) \frac{p_b}{N_{bn}} \right) + x_{ni} \frac{p_b}{2N_{ni}} \right) \\
x'_{bi} &= \frac{1}{\bar{W}} \left( x_{ij} \frac{p_b}{N_{ij}} + x_{bb} (1-s_{bb}) \frac{(n-1)p}{N_{bb}} + x_{bn} (1-s_{bn}) \frac{(n-1)p}{2N_{bn}} \right. \\
&\quad \left. + \frac{x_{bi}}{2} \left( s_{bi} (1-\delta) + (1-s_{bi}) \frac{(p_b + (n-2)p)}{N_{bi}} \right) + x_{ni} \frac{p_b}{2N_{ni}} + x_{bn+1} \frac{(n-1)p}{2N_{bn+1}} \right) \\
x'_{ni} &= \frac{1}{\bar{W}} \left( x_{ij} \frac{p_n}{N_{ij}} + x_{bn} (1-s_{bn}) \frac{(n-1)p}{2N_{bn}} + x_{ni} \frac{(n-2)p}{2N_{ni}} + x_{nn+1} \frac{(n-1)p}{2N_{nn+1}} + x_{in+1} \frac{p_n}{2N_{in+1}} \right) \\
x'_{bn+1} &= \frac{1}{\bar{W}} \left( x_{bb} (1-s_{bb}) \frac{p_{n+1}}{N_{bb}} + x_{bn} (1-s_{bn}) \frac{p_{n+1}}{2N_{bn}} + x_{bi} (1-s_{bi}) \frac{p_{n+1}}{2N_{bi}} \right) \\
x'_{nn+1} &= \frac{1}{\bar{W}} \left( x_{bn} (1-s_{bn}) \frac{p_{n+1}}{2N_{bn}} + x_{ni} \frac{p_{n+1}}{2N_{ni}} + x_{in+1} \frac{p_n}{2N_{in+1}} \right) \\
x'_{in+1} &= \frac{1}{\bar{W}} \left( x_{ij} \frac{p_{n+1}}{N_{ij}} + x_{bi} (1-s_{bi}) \frac{p_{n+1}}{2N_{bi}} + x_{ni} \frac{p_{n+1}}{2N_{ni}} + x_{bn+1} \frac{(n-1)p}{2N_{bn+1}} + x_{nn+1} \frac{(n-1)p}{2N_{nn+1}} \right. \\
&\quad \left. + x_{in+1} \frac{(n-2)p}{2N_{in+1}} \right)
\end{aligned}$$

where  $\bar{W}$ ,  $N_g$  and  $s_g$  are respectively defined in equations 3.1, 3.4 and 3.3, and where  $p$ ,  $p_n$ ,  $p_b$  and  $p_{n+1}$ , the frequencies of  $S_i$ ,  $S_n$ ,  $S_b$  and  $S_{n+1}$ , are

$$\begin{aligned}
p &= \frac{1}{(n-1)} \left( x_{ij} + \frac{x_{bi} + x_{ni} + x_{in+1}}{2} \right) \\
p_n &= \frac{x_{bn} + x_{ni} + x_{nn+1}}{2} \\
p_b &= x_{bb} + \frac{x_{bn} + x_{bi} + x_{bn+1}}{2} \\
p_{n+1} &= \frac{x_{bn+1} + x_{nn+1} + x_{in+1}}{2}.
\end{aligned}$$

**A3. Stability analysis.** Condition (i) : Fate of haplotype  $S_b$  when rare. Local stability analysis of the equilibrium when  $S_b$  was absent ( $x_0 = (n - 2)/n; x_4 = 2/n; x_1 = x_2 = x_3 = 0$ ) provides the parameter space where  $S_b$  can increase in frequency when rare. The parameter space we found where  $S_b$  increases when rare was identical to that described by Uyenoyama et al. (2001; Eq. A1) and is delimited by the thin solid line in Figure 3.1a for  $n = 5$ , and in Figure 3.2 for different values of  $n$ . Above this line,  $S_b$  did not increase in frequency when rare and was excluded from the population, *i.e.* when the self-pollen rate  $\alpha$  and the inbreeding depression  $\delta$  were high. Below this line,  $S_b$  increased in frequency and may either invade the population (resulting in the loss of SI) or be maintained at intermediate frequencies along with functional haplotypes (see condition (ii) below). Note that when the population contained only three SI haplotypes ( $n = 3$ ), the parameter space was empty, *i.e.*  $S_b$  was always expected to increase in frequency, and that for  $n = 4$ , the parameter space was limited to the point  $\alpha = 1$  and  $\delta = 1$ . Figure 3.2 shows that the parameter space of condition (i) became larger as the number of SI haplotypes segregating in the population ( $n$ ) increased.

Condition (ii) : Fate of the haplotype  $S_b$  when frequent. Local stability analysis of the equilibrium where  $S_b$  was fixed ( $x_1 = 1; x_0 = x_2 = x_3 = x_4 = 0$ ) gives the parameter space where  $S_b$  increased in frequency when already frequent, *i.e.* where it reaches fixation. We found that, for any  $n$ ,  $S_b$  increased when already frequent if

$$\delta < \frac{\alpha - 2}{2\alpha - 3}. \quad (3.7)$$

This threshold corresponds to the critical inbreeding depression above which an SI haplotype can enter a SC population from a low initial frequency (Charlesworth and Charlesworth 1979; Eq. 2). This threshold is shown on Figures 3.1a and 3.2 as the thick solid line and is independent from the number of SI haplotypes in the population.

Condition (iii) : Fate of haplotype  $S_n$  when  $S_b$  was present. The parameter space where  $S_n$  was excluded when  $S_b$  was present was determined numerically and is delimited by the dashed lines on Figure 3.1a for  $n = 5$  and on Figure 3.2 for different values of  $n$ . The parameter space for which  $S_b$  and  $S_n$  can both be maintained in the population at the same time is located between the thin solid line, condition (i), and the dashed line, condition (iii). Figure 3.1a shows that  $S_n$  was not excluded by  $S_b$  for high values of  $\alpha$  and  $\delta$ . Figure 3.2 shows that the parameter space for which  $S_n$  is not excluded became smaller and was displaced toward smaller values of  $\alpha$  and  $\delta$  when  $n$  increased. Practically,

we were not able to distinguish between conditions (i) and (iii) when  $n > 10$ . Notice that for  $n = 3$  and  $n = 4$ , the parameter space for condition (iii) was large, whereas it either did not exist or was reduced to a single point for condition (i).

Condition (iv) : Fate of haplotype  $S_{n+1}$  when both  $S_n$  and  $S_b$  were present. Our results showed that the introduction of the new SI haplotype  $S_{n+1}$  in a population composed of  $S_b$  and  $S_i$  haplotypes (including or not the ancestor  $S_n$ ) always led to the exclusion of the pollen-part mutant  $S_b$ , resulting in a fully self-incompatible population including a new SI haplotype. In cases where  $S_n$  was maintained in polymorphism with the pollen-part mutant  $S_b$ , this process can thus potentially lead to an increase in the number of SI haplotypes in the population.

### 3.6 Supplemental

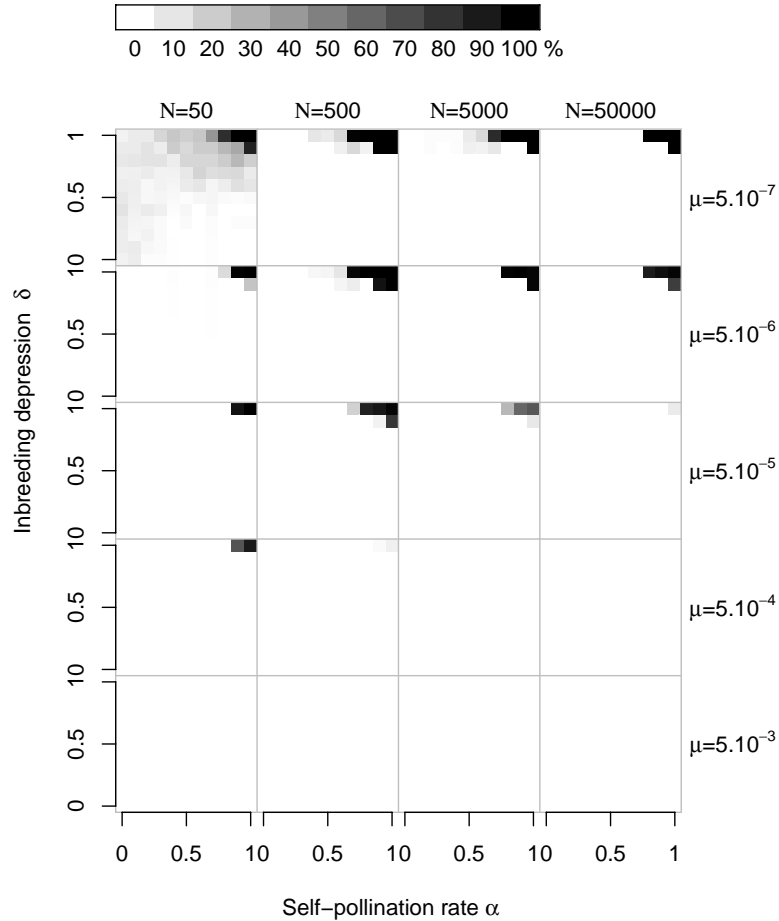
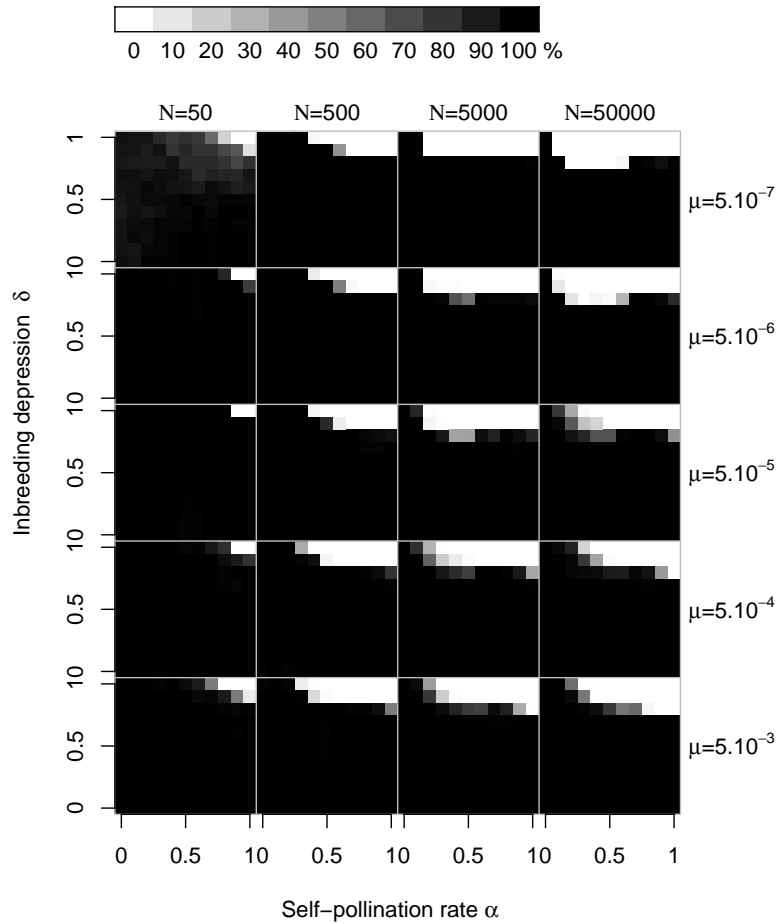
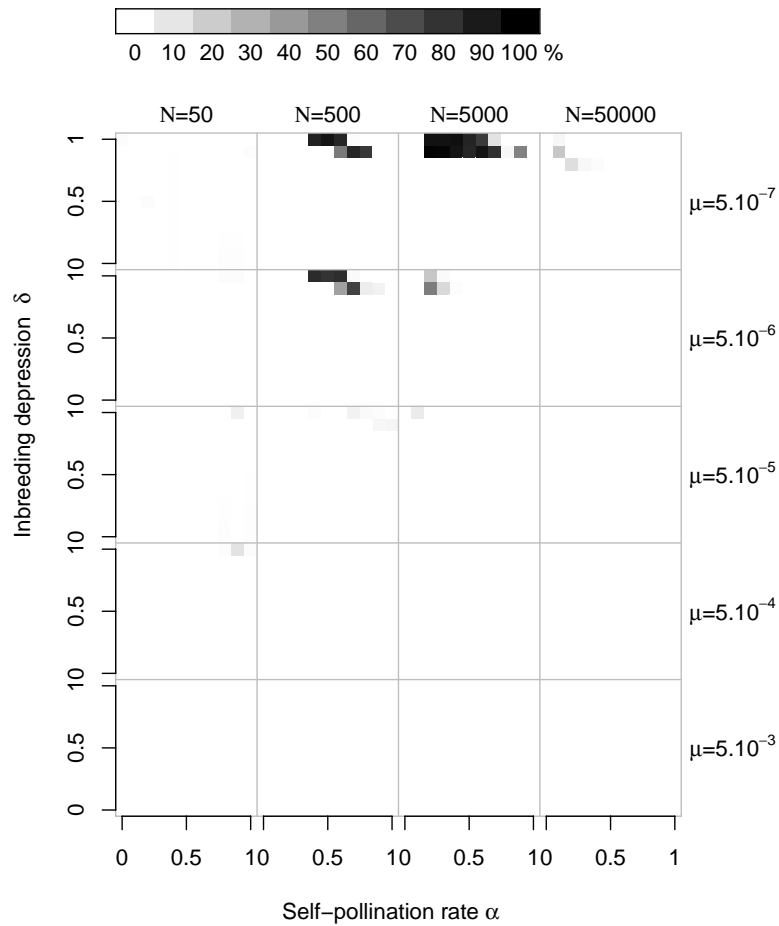


FIGURE S1 – The maintenance region in a finite population for different population sizes and mutation rates, with an initial number of specificities  $n = 5$ . Grid cells are figured in shades of grey that are proportional to the number of simulations where maintenance was observed ( $3 \leq n \leq 5$  at the end of the simulation), with black squares corresponding to sets of parameters under which the evolutionary outcome was only maintenance and white squares to cases under which no maintenance was observed. A hundred replicate runs were performed for each combination of parameter values (population size  $N$ , mutation rate  $\mu$ ).  $k = 20$  for all simulations.





**FIGURE S2** – The loss region in a finite population for different population sizes and mutation rates, with an initial number of specificities  $n = 5$ . Grid cells are figured in shades of grey that are proportional to the number of simulations where diversification was observed ( $n = 0$  at the end of the simulation), with black squares corresponding to sets of parameters under which the evolutionary outcome was only loss and white squares to cases under which no loss was observed. A hundred replicate runs were performed for each combination of parameter values (population size  $N$ , mutation rate  $\mu$ ).  $k = 20$  for all simulations.



**FIGURE S3** – The polymorphism region in a finite population for different population sizes and mutation rates, with an initial number of specificities  $n = 5$ . Grid cells are figured in shades of grey that are proportional to the number of simulations where a polymorphic population was observed with a mix of SC haplotypes, pollen-part mutant and function SI haplotypes. Black squares corresponding to sets of parameters under which the evolutionary outcome was only polymorphism and white squares to cases under which no polymorphism was observed. A hundred replicate runs were performed for each combination of parameter values (population size  $N$ , mutation rate  $\mu$ ).  $k = 20$  for all simulations.

# Chapitre 4

## Discussion

La perte de l'auto-incompatibilité au profit de l'auto-compatibilité est l'une des transitions les plus fréquemment observées dans l'évolution des Angiospermes. Cela peut sembler concordant avec les résultats des différents modèles d'évolution de l'auto-incompatibilité, selon lesquels les conditions permettant à une population auto-incompatible de résister à l'invasion d'un haplotype auto-compatible sont très restreintes (Charlesworth and Charlesworth, 1979 ; Porcher and Lande, 2005 ; Uyenoyama et al., 2001). Mais l'auto-incompatibilité est encore présente chez plus de la moitié des espèces d'Angiospermes, ce que ne parviennent pas à expliquer les modèles actuels. Notre modèle diffère des précédents par divers aspects. Il considère notamment une taille de population finie, des mutations récurrentes vers l'auto-compatibilité, un nombre d'allèles S variable, et une dépression de consanguinité dynamique modélisée par des mutations délétères partiellement liées entre elles et au locus S. Malgré ces différences qui rendent notre modèle plus réaliste, nos résultats sont très similaires aux attendus des modèles précédents, et particulièrement pour les modèles considérant une dépression de consanguinité constante (Charlesworth and Charlesworth, 1979 ; Uyenoyama et al., 2001). En effet, les nombreux paramètres que nous avons considérés (taille de population, taux de mutations délétères, coefficients de dominance et de sélection des mutations délétères...) influencent la dépression de consanguinité, qui est elle-même corrélée avec l'évolution du système d'auto-incompatibilité. Le maintien de l'auto-incompatibilité est toujours associé à une forte dépression de consanguinité, supérieure à  $2/3$ , et cette dépression de consanguinité est d'autant plus forte que le taux d'auto-pollinisation diminue. La question du maintien de l'auto-incompatibilité reste donc ouverte, et la modélisation explicite de la dépression de consanguinité ne semble pas être la clef permettant d'y répondre.

Si nos résultats sont très proches de ceux observés pour des modèles considérant une dépression de consanguinité constante (Charlesworth and Charlesworth, 1979 ; Uyenoyama et al., 2001), ils divergent légèrement de ceux de Porcher and Lande (2005), qui ont considéré une dépression de consanguinité dynamique dans une population de taille infinie. En effet, dans leur cas, la population auto-incompatible est toujours envahie par le mutant auto-compatible lorsque le taux d'auto-pollinisation est très élevé (sauf dans le cas où l'on ajoute au modèle un fardeau lié et/ou une dépression de consanguinité constante). Notre modèle et le leur diffèrent par divers aspects. En raison de la taille finie de la population, les mutations délétères peuvent se retrouver à l'état homozygote chez des individus issus de croisement entre apparentés, alors que cela n'est possible que chez les individus issu d'autofécondation dans le modèle de Porcher and Lande (2005). De plus, nous considérons des mutations partiellement liées entre elles et avec le locus S, alors que tous les loci sont indépendants dans le cas de Porcher and Lande (2005). Par conséquent, il serait intéressant de déterminer quelles hypothèses de modélisation sont responsables de la différence observée.

Glémin et al. (2001) ont étudié en population finie l'effet de l'auto-incompatibilité sur l'évolution de la dépression de consanguinité. Ils ont montré que la dépression de consanguinité était plus faible lorsque le taux de recombinaison était plus élevé, ce qui est contraire à nos résultats. Mais leur modèle est différent du notre sous plusieurs aspects. D'une part, ils ont utilisé un modèle à deux locus partiellement liés : le locus S et un locus soumis à sélection purifiante. Il serait intéressant de savoir si cette différence est due à la recombinaison entre le locus S et les mutations délétères ou bien à la recombinaison entre mutations délétères. Cet effet pourrait être étudié en utilisant un modèle à trois locus considérant des taux de recombinaison différents entre le locus S et deux locus sous sélection purifiante. Une alternative serait de modifier notre modèle de simulation de manière à ce que le locus S soit sur un chromosome différent. Ainsi, seules les mutations délétères seraient partiellement liées. D'autre part, Glémin et al. (2001) ne se sont pas intéressés au maintien de l'auto-incompatibilité et ont donc considéré une population auto-incompatible sans y introduire de mutants auto-compatibles. La différence entre nos résultats et les leurs en terme d'effet de la recombinaison concerne les cas identifiés comme "maintien de l'auto-incompatibilité", c'est-à-dire les cas pour lesquels la fréquence de l'haplotype auto-compatible est inférieure à 2%. Néanmoins, il serait intéressant de savoir si la présence d'un auto-compatible, même en faible fréquence, peut influencer l'évolution de la dépression de consanguinité. Il suffirait pour cela de comparer nos résultats avec

des simulations de populations auto-incompatibles sans y introduire de mutants auto-compatibles.

L'extrême diversité observée au locus S, l'un des locus les plus diversifiés du monde vivant, soulève la question des processus et des conditions permettant l'évolution de nouveaux haplotypes S. La principale difficulté dans la compréhension de ce phénomène réside dans la structure même du locus S, constitué de deux gènes liés. Un haplotype S n'est fonctionnel, c'est-à-dire auto-incompatible, que si les deux gènes pollen et pistil expriment la même spécificité. L'évolution de nouveaux haplotypes S nécessite donc la coévolution de ces deux gènes. Seuls Uyenoyama et al. (2001) avaient jusqu'à présent proposé un modèle analytique de diversification au locus S, permettant d'expliquer l'évolution d'un nouvel haplotype S via une première mutation dans le pollen, suivie d'une mutation compensatrice dans le pistil. Mais l'haplotype intermédiaire auto-compatible entraînait quasi-systématiquement l'exclusion de l'haplotype ancestral. Ce modèle ne permettait donc pas d'expliquer la diversification à proprement parler au locus S, c'est-à-dire une augmentation du nombre d'haplotypes S. Après avoir repris et analysé ce modèle, nous l'avons transposé dans une population de taille finie, avec des mutations récurrentes entraînant l'apparition de nouvelles spécificités à l'un ou l'autre des deux gènes pollen ou pistil. Nos résultats montrent que les conditions permettant la diversification du système d'auto-incompatibilité sont beaucoup plus étendues qu'attendu en population infinie, et recouvrent notamment des gammes de paramètres pour lesquelles on s'attendait à observer l'exclusion de l'haplotype auto-compatible intermédiaire, et donc le maintien du système d'auto-incompatibilité sans évolution du nombre d'haplotypes S. Ceci peut s'expliquer par le fait que, par dérive, les haplotypes auto-compatibles, bien que contre-sélectionnés, peuvent être maintenus dans la population suffisamment longtemps pour qu'une mutation compensatrice apparaisse. La dérive, souvent considérée comme un facteur réduisant la diversité, peut dans certains cas être au contraire un mécanisme favorisant fortement la diversification.

Nos deux modèles M (pour Maintien) et D (pour Diversification) ne considèrent pas les mêmes hypothèses et ne présentent donc pas les mêmes limites (voir Chapitre 1). Les principales différences entre ces deux modèles sont la structure du locus S (un seul gène dans le cas du modèle M, alors que le modèle D considère des mutations indépendantes sur les deux gènes pollen et pistil) et la modélisation de la dépression de consanguinité (dynamique dans le cas du modèle M et constante dans le cas du modèle D). Nous avons vu que les conditions permettant le maintien de l'auto-incompatibilité d'après

notre modèle M sont très similaires à celles déterminées par les modèles précédents (Charlesworth and Charlesworth, 1979 ; Porcher and Lande, 2005 ; Uyenoyama et al., 2001). Notre modèle D nous a permis de déterminer des conditions favorables à la diversification du locus S. Il s'avère que, au moins pour des tailles de population et des taux de mutation au locus S élevés, ces conditions correspondent globalement aux conditions favorisant le maintien de l'auto-incompatibilité. Cette région de paramètres, correspondant à une dépression de consanguinité élevée et à des taux d'autofécondation élevés ou intermédiaires, est donc très conservée d'un modèle à l'autre. On peut donc supposer que le fait d'introduire dans le modèle M des hypothèses du modèle D, ou inversement, n'aura pas un impact important sur les conclusions en terme de conditions pour le maintien ou la diversification du système d'auto-incompatibilité.

Les résultats de notre étude de la diversification au locus S montrent que le taux de diversification n'est pas constant au cours du temps, mais tend à devenir plus faible quand le nombre d'haplotypes S dans la population augmente. Notre modèle montre donc que la diversification au locus S est possible, mais ne permet pas d'expliquer la totalité de la diversité observée dans les populations naturelles. Nous avons considéré une population non-structurée. Or, la migration efficace des allèles S est très élevée en raison de la sélection fréquence-dépendante négative. Si un allèle S arrive par migration dans une population dans laquelle il est absent, il sera compatible avec la totalité des génotypes et augmentera donc en fréquence. Le fait de considérer une population structurée pourrait donc permettre d'augmenter la diversité au sein des différents dèmes, si les processus de diversification au sein de chaque dème mènent à l'évolution d'haplotypes S différents, puis que ces différents haplotypes migrent d'une population à l'autre. La diversification des haplotypes S serait donc favorisée par une structuration inter-dèmes relativement forte. Cependant, la migration des haplotypes intermédiaires auto-compatibles est également favorisée par la sélection fréquence-dépendante négative et pourrait au contraire rendre plus difficile les processus de diversification. La structure des populations a également un effet sur la dépression de consanguinité, et pourrait donc indirectement influencer l'évolution de l'auto-incompatibilité vers son maintien ou sa perte. En effet, si l'on compare avec la dépression de consanguinité attendue en population non structurée, la dépression de consanguinité intra-dème est plus faible, tandis que la dépression de consanguinité inter-dème est plus élevée (Glémin et al., 2003 ; Roze and Rousset, 2004), ce qui pourrait faciliter la perte de l'auto-incompatibilité. Il est donc difficile de savoir *a priori* l'effet qu'aurait la prise en compte d'une population structurée sur les conditions

de diversification au locus S. La prise en compte de la structuration des populations représente donc une perspective claire d'amélioration de notre modèle de diversification.

La plupart des modèles d'évolution de l'auto-incompatibilité considèrent un système d'auto-incompatibilité gamétophytique. D'une part, ce système est beaucoup plus fréquent que l'auto-incompatibilité sporophytique (identifiés respectivement dans 17 et 6 familles d'Angiospermes ; Franklin-Tong and Franklin, 2003 ; Hiscock and Tabah, 2003). D'autre part, ce système est plus facile à modéliser en raison de l'absence de relations de dominance, ce qui permet notamment de considérer que tous les allèles S sont en même fréquence à l'équilibre. Les relations de dominance entre allèles S dans le système d'auto-incompatibilité sporophytique font que les croisements entre apparentés sont plus fréquents qu'en système gamétophytique et qu'on peut observer des individus homozygotes au locus S pour les allèles S récessifs. On peut donc s'attendre à observer une homozygotie à l'échelle du génome plus élevée qu'en gamétophytique, ce qui peut potentiellement mener à une purge plus efficace des mutations délétères, et donc à une dépression de consanguinité plus réduite. Le maintien du système d'auto-incompatibilité sporophytique pourrait alors nécessiter des conditions encore plus restreintes que le maintien de l'auto-incompatibilité gamétophytique, ce qui pourrait expliquer en partie le fait que l'auto-incompatibilité gamétophytique est maintenue dans un nombre de familles beaucoup plus important que l'auto-incompatibilité sporophytique. Néanmoins, on peut supposer que l'avantage d'une mutation conférant l'auto-compatibilité ne sera pas la même en fonction de la classe de dominance à laquelle appartient l'allèle touché. Ces relations étant très complexe, de mêmes que les mécanismes les contrôlant, il est très difficile d'intuiter sur les conditions menant à la perte ou au maintien de l'auto-incompatibilité sporophytique.





# Bibliographie

- Allen, A. and S. Hiscock, 2008. Evolution and phylogeny of self-incompatibility systems in angiosperms. Pp. 73–101, *in* Self-Incompatibility in Flowering Plants. Springer-Verlag Berlin Heidelberg.
- Barrett, S., 2002. The evolution of plant sexual diversity. *Nature Reviews Genetics* 3 :274–284.
- Barrett, S., L. Jesson, and A. Baker, 2000. The evolution and function of stylar polymorphisms in flowering plants. *Annals Of Botany* 85 :253–265.
- Barrett, S. and J. Shore, 2008. New Insights on Heterostyly : Comparative Biology, Ecology and Genetics. Pp. 3–32, *in* Self-Incompatibility in Flowering Plants. Springer-Verlag Berlin Heidelberg.
- Barton, N. and S. Otto, 2005. Evolution of recombination due to random drift. *Genetics* 169 :2353–2370.
- Bataillon, T. and M. Kirkpatrick, 2000. Inbreeding depression due to mildly deleterious mutations in finite populations : size does matter. *Genetical Research* 75 :75–81.
- Billiard, S., M. Lopez-Villavicencio, B. Devier, M. Hood, C. Fairhead, and T. Giraud, 2011. Having sex, yes, but with whom ? Inferences from fungi on the evolution of anisogamy and mating types. *Biol. Rev.* 86 :421–442.
- Busch, J., 2005. Inbreeding depression in self-incompatible and self-compatible populations of *Leavenworthia alabamica*. *Heredity* 94 :159–165.
- Busch, J. W., S. Joly, and D. J. Schoen, 2010. Does mate limitation in self-incompatible species promote the evolution of selfing ? The case of *Leavenworthia alabamica*. *Evolution* 64 :1657–1670.
- Castric, V., J. S. Bechsgaard, S. Grenier, R. Nouredine, M. H. Schierup, and X. Vekemans, 2010. Molecular evolution within and between self-incompatibility specificities. *Mol. Biol. Evol.* 27 :11–20.
- Castric, V. and X. Vekemans, 2007. Evolution under strong balancing selection : how many codons determine specificity at the female self-incompatibility gene SRK in Brassicaceae. *BMC Evol. Biol.* 7 :132.

## Bibliographie

- Charlesworth, D., 2000. How can two-gene models of self-incompatibility generate new specificities? *Plant Cell* 12 :309–310.
- , 2002. Plant sex determination and sex chromosomes. *Heredity* 88 :94–101.
- Charlesworth, D. and B. Charlesworth, 1979. Evolution and breakdown of S-allele systems. *Heredity* 43 :41–55.
- , 1987. Inbreeding Depression And Its Evolutionary Consequences. *Annual Review Of Ecology And Systematics* 18 :237–268.
- Charlesworth, D. and J. H. Willis, 2009. Fundamental concepts in genetics. The genetics of inbreeding depression. *Nature Reviews Genetics* 10 :783–796.
- Chookajorn, T., A. Kachroo, D. Ripoll, A. Clark, and J. Nasrallah, 2004. Specificity determinants and diversification of the *Brassica* self-incompatibility pollen ligand. *Proc. Natl. Acad. Sci. U. S. A.* 101 :911–917.
- Fisher, R., 1941. Average excess and average effect of a gene substitution. *Annals of Eugenics* 11 :53–63.
- Fisher, R. A., 1961. A Model for the Generation of Self-Sterility Alleles. *J. Theor. Biol.* 1 :411–414.
- Fobis-Loisy, I., C. Miede, and T. Gaude, 2004. Molecular evolution of the S locus controlling mating in the Brassicaceae. *Plant Biology* 6 :109–118.
- Foxe, J. P., T. Slotte, E. A. Stahl, B. Neuffer, H. Hurka, and S. I. Wright, 2009. Recent speciation associated with the evolution of selfing in *Capsella*. *Proceedings of the National Academy of Sciences of the United States of America* 106 :5241–5245.
- Franklin-Tong, N. and F. C. H. Franklin, 2003. Gametophytic self-incompatibility inhibits pollen tube growth using different mechanisms. *Trends in Plant Science* 8 :598–605.
- Gervais, C. E., V. Castric, A. Ressayre, and S. Billiard, 2011. Origin and diversification dynamics of self-incompatibility haplotypes. *Genetics* 188 :625–U205.
- Glémin, S., 2003. How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution* 57 :2678–2687.
- Glémin, S., T. Bataillon, J. Ronfort, A. Mignot, and I. Olivieri, 2001. Inbreeding depression in small populations of self-incompatible plants. *Genetics* 159 :1217–1229.
- Glémin, S., J. Ronfort, and T. Bataillon, 2003. Patterns of inbreeding depression and architecture of the load in subdivided populations. *Genetics* 165 :2193+.
- Goodwillie, C., S. Kalisz, and C. Eckert, 2005. The evolutionary enigma of mixed mating systems in plants : Occurrence, theoretical explanations, and empirical evidence. *Annual Review Of Ecology Evolution And Systematics* 36 :47–79.

- Hiscock, S., 2000. Genetic control of self-incompatibility in *Senecio squalidus* L. (Asteraceae) : a successful colonizing species. *Heredity* 85 :10–19.
- Hiscock, S. and D. Tabah, 2003. The different mechanisms of sporophytic self-incompatibility. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 358 :1037–1045.
- Hiscock, S. J., 2002. Pollen recognition during the self-incompatibility response in plants. *Genome Biology* 3.
- Igic, B., L. Bohs, and J. Kohn, 2004. Historical inferences from the self-incompatibility locus. *New Phytologist* 161 :97–105.
- Igic, B. and J. Kohn, 2001. Evolutionary relationships among self-incompatibility RNases. *PNAS* 98 :13167–13171.
- , 2006. The distribution of plant mating systems : Study bias against obligately out-crossing species. *Evolution* 60 :1098–1103.
- Igic, B., R. Lande, and J. R. Kohn, 2008. Loss of self-incompatibility and its evolutionary consequences. *Int. J. Plant Sci.* 169 :93–104.
- Igic, B., W. A. Smith, K. A. Robertson, B. A. Schaal, and J. R. Kohn, 2007. Studies of self-incompatibility in wild tomatoes : I. S-allele diversity in *Solanum chilense* (Dun.) Reiche (Solanaceae). *Heredity* 99 :562.
- Ikeda, K., B. Igic, K. Ushijima, H. Yamane, N. Hauck, R. Nakano, H. Sassa, A. Iezzoni, J. Kohn, and R. Tao, 2004. Primary structural features of the S haplotype-specific F-box protein, SFB, in *Prunus*. *Sexual Plant Reproduction* 16 :235–243.
- Ioerger, T., J. Gohlke, B. Xu, and T. Kao, 1991. Primary Structural Features Of The Self-Incompatibility Protein In Solanaceae. *Sexual Plant Reproduction* 4 :81–87.
- Ishimizu, T., T. Shinkawa, F. Sakiyama, and S. Norioka, 1998. Primary structural features of rosaceous S-RNases associated with gametophytic self-incompatibility. *Plant Molecular Biology* 37 :931–941.
- Kondo, K., M. Yamamoto, R. Itahashi, T. Sato, H. Egashira, T. Hattori, and Y. Kowayama, 2002. Insights into the evolution of self-compatibility in *Lycopersicon* from a study of stylar factors. *Plant Journal* 30 :143–153.
- Lande, R. and D. Schemske, 1985. The Evolution Of Self-Fertilization And Inbreeding Depression In Plants .1. Genetic Models. *Evolution* 39 :24–40.
- Lande, R., D. Schemske, and S. Schultz, 1994. High Inbreeding Depression, Selective Interference Among Loci, And The Threshold Selfing Rate For Purging Recessive Lethal Mutations. *Evolution* 48 :965–978.

## Bibliographie

- Langridge, P. and U. Baumann, 2008. Self-Incompatibility in the Grasses. Pp. 275–287, *in* S.-V. B. Heidelberg, ed. Self-Incompatibility in Flowering Plants.
- Lawrence, M., 2000. Population genetics of the homomorphic self-incompatibility polymorphisms in flowering plants. *Ann. Bot.* 85 :221–226.
- Lewis, D., 1949. Structure of the Incompatibility Gene. *Heredity* 3 :339–355.
- Llaurens, V., V. Castric, F. Austerlitz, and X. Vekemans, 2008. High paternal diversity in the self-incompatible herb *Arabidopsis halleri* despite clonal reproduction and spatially restricted pollen dispersal. *Molecular Ecology* 17 :1577–1588.
- Llaurens, V., L. Gonthier, and S. Billiard, 2009. The sheltered genetic load linked to the S locus in plants : new insights from theoretical and empirical approaches in sporophytic self-incompatibility. *Genetics* 183 :1105–1118.
- Matton, D., D. Luu, Q. Xike, G. Laublin, M. O’Brien, O. Maes, D. Morse, and M. Cappadocia, 1999. Production of an S-RNase with dual specificity suggests a novel hypothesis for the generation of new S alleles. *Plant Cell* 11 :2087–2097.
- Matton, D., C. Maes, G. Laublin, X. Qin, C. Bertrand, D. Morse, and M. Cappadocia, 1997. Hypervariable domains of self-incompatibility RNases mediate allele-specific pollen recognition. *Plant Cell* 9 :1757–1766.
- May, G., F. Shaw, H. Badrane, and X. Vekemans, 1999. The signature of balancing selection : Fungal mating compatibility gene evolution. *Proc. Natl. Acad. Sci. U. S. A.* 96 :9172–9177.
- Mena-Ali, J. I. and A. G. Stephenson, 2007. Segregation analyses of partial self-incompatibility in self and cross progeny of *Solanum carolinense* reveal a leaky S-allele. *Genetics* 177 :501–510.
- Miller, J. S., R. A. Levin, and N. M. Feliciano, 2008. A tale of two continents : Baker’s rule and the maintenance of self-incompatibility in *Lycium* (Solanaceae). *Evolution* 62 :1052–1065.
- Morlon, H., M. D. Potts, and J. B. Plotkin, 2010. Inferring the dynamics of diversification : A coalescent approach. *PLoS Biology* 8 :e1000493.
- Nasrallah, M., P. Liu, and J. Nasrallah, 2002. Generation of self-incompatible *Arabidopsis thaliana* by transfer of two S locus genes from *A. lyrata*. *Science* 297 :247–249.
- de Nettancourt, D., 2001. Incompatibility and incongruity in wild and cultivated plants. Springer-Verlag, Berlin.
- Nishio, T. and M. Kusaba, 2000. Sequence diversity of SLG and SRK in *Brassica oleracea* L. *Annals of Botany* 85 :141–146.
- Nunes, M. D. S., R. A. M. Santos, S. M. Ferreira, J. Vieira, and C. P. Vieira, 2006. Variability patterns and positively selected sites at the gametophytic self-incompatibility pollen SFB gene in a wild self-incompatible *Prunus spinosa* (Rosaceae) population. *New Phytol.* 172 :577–587.

- Ortega, E., R. I. Boskovic, D. J. Sargent, and K. R. Tobutt, 2006. Analysis of S-RNase alleles of almond (*Prunus dulcis*) : characterization of new sequences, resolution of synonyms and evidence of intragenic recombination. *Mol. Genet. Genomics* 276 :577.
- Otto, S. and T. Lenormand, 2002. Resolving the paradox of sex and recombination. *Nature Reviews Genetics* 3 :252–261.
- Paape, T., B. Igit, S. D. Smith, R. Olmstead, L. Bohs, and J. R. Kohn, 2008. A 15-Myr-old genetic bottleneck. *Mol. Biol. Evol.* 25 :655–663.
- Paape, T., T. Miyake, N. Takebayashi, D. Wolf, and J. R. Kohn, 2011. Evolutionary genetics of an S-like polymorphism in Papaveraceae with putative function in self-incompatibility. *PLoS ONE* 6.
- Pannell, J. and S. Barrett, 1998. Baker's law revisited : Reproductive assurance in a metapopulation. *Evolution* 52 :657–668.
- Phadke, S. S. and R. A. Zufall, 2009. Rapid diversification of mating systems in ciliates. *Biol. J. Linn. Soc.* 98 :187–197.
- Porcher, E. and R. Lande, 2005. Loss of gametophytic self-incompatibility with evolution of inbreeding depression. *Evolution* 59 :46–60.
- Quental, T. B. and C. R. Marshall, 2010. Diversity dynamics : molecular phylogenies need the fossil record. *Trends in Ecology and Evolution* 25 :434–441.
- Richards, A., 1986. *Plant breeding systems*. George Allen & Unwin.
- Ronce, O., 2007. How does it feel to be like a rolling stone? Ten questions about dispersal evolution. *Annual Review of Ecology Evolution and Systematics* 38 :231–253.
- Roze, D., 2009. Diploidy, population structure, and the evolution of recombination. *AmNat* 174 :S79–S94.
- Roze, D. and N. H. Barton, 2006. The Hill-Robertson effect and the evolution of recombination. *Genetics* 173 :1793–1811.
- Roze, D. and F. Rousset, 2004. Joint effects of self-fertilization and population structure on mutation load, inbreeding depression and heterosis. *Genetics* 167 :1001–1015.
- Sabelleil, M., S. Rivard, D. Morse, and M. Cappadocia, 1994. The S11 and S13 self-incompatibility alleles in *Solanum chacoense* are remarkably similar. *Plant Molecular Biology* 24 :571–583.
- Sainudiin, R., W. Wong, K. Yogeewaran, J. Nasrallah, Z. Yang, and R. Nielsen, 2005. Detecting site-specific physicochemical selective pressures : Applications to the class-I HLA of the human major histocompatibility complex and the SRK of the plant sporophytic self-incompatibility system. *J. Mol. Evol.* 60 :315–326.

## Bibliographie

- Savage, A. and J. Miller, 2006. Gametophytic self-incompatibility in *Lycium parishii* (Solanaceae) : allelic diversity, genealogical structure, and patterns of molecular evolution at the S-RNase locus. *Heredity* 96 :434–444.
- Schierup, M., A. Mikkelsen, and J. Hein, 2001. Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes. *Genetics* 159 :1833–1844.
- Shimizu, K. K., R. Shimizu-Inatsugi, T. Tsuchimatsu, and M. D. Purugganan, 2008. Independent origins of self-compatibility in *Arabidopsis thaliana*. *Molecular Ecology* 17 :704–714.
- Silva, N. and D. Goring, 2001. Mechanisms of self-incompatibility in flowering plants. *Cellular and Molecular Life Sciences* 58 :1988–2007.
- Solberg, O. D., S. J. Mack, A. K. Lancaster, R. M. Single, Y. Tsai, A. Sanchez-Mazas, and G. Thomson, 2008. Balancing selection and heterogeneity across the classical human leukocyte antigen loci : A meta-analytic review of 497 population studies. *Hum. Immunol.* 69 :443–464.
- Stebbins, G., 1957. Self fertilization and population variability in the higher plants. *American Naturalist* 91 :337–354.
- Steinbachs, J. and K. Holsinger, 2002. S-RNase-mediated gametophytic self-incompatibility is ancestral in eudicots. *Molecular Biology and Evolution* 19 :825–829.
- Stone, J., 2002. Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility. *Quarterly Review Of Biology* 77 :17–32.
- , 2004. Sheltered load associated with S-alleles in *Solanum carolinense*. *Heredity* 92 :335–342.
- Takahata, N., 1990. A simple genealogical structure of strongly balanced allelic lines and transspecies evolution of polymorphism. *Proc. Natl. Acad. Sci. U. S. A.* 87 :2419–2423.
- Takayama, S. and A. Isogai, 2005. Self-incompatibility in plants. *Annu. Rev. Plant Biol.* 56 :467–489.
- Takebayashi, N., P. Brewer, E. Newbigin, and M. Uyenoyama, 2003. Patterns of variation within self-incompatibility loci. *Mol. Biol. Evol.* 20 :1778–1794.
- Tsuchimatsu, T., K. Suwabe, R. Shimizu-Inatsugi, S. Isokawa, P. Pavlidis, G. Städler, T. Suzuki, S. Takayama, M. Watanabe, and K. Shimizu, 2010. Evolution of self-compatibility by a mutation in the male specificity gene. *Nature* 464 :1342–1346.
- Uyenoyama, M., 1988a. On the evolution of genetic incompatibility systems 2. Initial increase of strong gametophytic self-incompatibility under partial selfing and half-sib mating. *Am. Nat.* 131 :700–722.

- , 1988b. On the evolution of genetic incompatibility systems 3. Introduction of weak gametophytic self-incompatibility self-incompatibility under partial inbreeding. *Theor. Popul. Biol.* 34 :47–91.
- , 1997. Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics* 147 :1389–1400.
- Uyenoyama, M. and E. Newbigin, 2000. Evolutionary dynamics of dual-specificity self-incompatibility alleles. *Plant Cell* 12 :310–312.
- Uyenoyama, M., Y. Zhang, and E. Newbigin, 2001. On the origin of self-incompatibility haplotypes : Transition through self-compatible intermediates. *Genetics* 157 :1805–1817.
- Vekemans, X. and M. Slatkin, 1994. Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* 137 :1157–1165.
- Vieira, C., D. Charlesworth, and J. Vieira, 2003. Evidence for rare recombination at the gametophytic self-incompatibility locus. *Heredity* 91 :262–267.
- Wang, X., A. Hughes, T. Tsukamoto, T. Ando, and T. Kao, 2001. Evidence that intragenic recombination contributes to allelic diversity of the S-RNase gene at the self-incompatibility (S) locus in *Petunia inflata*. *Plant Physiol.* 125 :1012–1022.
- Wheeler, M. J., B. H. J. de Graaf, N. Hadjiosif, R. M. Perry, N. S. Poulter, K. Osman, S. Vavtovic, A. Harper, F. C. H. Franklin, and V. E. Franklin-Tong, 2009. Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. *Nature* 459 :992–U118.
- Willi, Y. and K. Mäeettaenen, 2010. Evolutionary dynamics of mating system shifts in *Arabidopsis lyrata*. *Journal of Evolutionary Biology* 23 :2123–2131.
- Wolfram Research, I., 2008. Mathematica Edition : Version 7.0. Wolfram Research, Inc.
- Wright, S., 1939. The distribution of self-sterility alleles in populations. *Genetics* 24 :538–552.
- Wright, S. I. and S. C. H. Barrett, 2010. The Long-Term Benefits of Self-Rejection. *Science* 330 :459–460.
- Yokoyama, S. and L. Hetherington, 1982. The expected number of self-incompatibility alleles in finite plant-populations. *Heredity* 48 :299–303.

## Abstract

Self-incompatibility is a widespread genetic system, which enables hermaphrodite plants to avoid self-fertilization and mating with close relatives. It is based on the pistil's capacity to recognize and reject pollen when they express cognate specificities. Specificities are encoded by alleles at the self-incompatibility gene complex (known as the S-locus), composed of two linked genes, one expressed in pollen and the other expressed in pistils. During my PhD, I studied the maintenance and evolution of self-incompatibility from a theoretical standpoint, using a modeling approach. The first part of my thesis examined the co-evolution of self-incompatibility and inbreeding depression in finite populations, focusing on the conditions for maintenance of self-incompatibility when self-compatible mutants were repeatedly introduced in the population by recurring mutations. Our results showed that the maintenance of self-incompatibility is associated with high inbreeding depression and is facilitated by high rates of self-pollination. The second part of my thesis explored the conditions for evolution of novel self-incompatibility alleles (S alleles), which we have studied both analytically in infinite populations and in finite populations *via* computer simulations. Our results showed that the conditions for diversification at the S locus are much less stringent in finite than in infinite populations and that there is more diversification at this locus when few S alleles are present in the population.

**Keywords :** Self-incompatibility , modeling, mating system, inbreeding depression.

## Résumé

L'auto-incompatibilité est un mécanisme génétique très répandu chez les Angiospermes. Elle permet aux plantes hermaphrodites d'éviter l'autofécondation et les croisements entre individus apparentés par un mécanisme de reconnaissance et de rejet du pollen par le pistil, lorsque ceux-ci expriment la même spécificité. Ces spécificités dépendent des allèles au locus d'auto-incompatibilité (le locus S) qui est composé de deux gènes strictement liés, l'un s'exprimant dans le pollen, l'autre dans le pistil. Dans le cadre de cette thèse, je me suis intéressée au maintien et à l'évolution de l'auto-incompatibilité, que j'ai étudiés par une approche de modélisation. La première partie de ma thèse traite de l'évolution conjointe de l'auto-incompatibilité et de la dépression de consanguinité, et s'attache notamment à étudier les conditions permettant le maintien de l'auto-incompatibilité. Cette étude est basée sur des simulations en population finie. Nos résultats montrent que le maintien de l'auto-incompatibilité est associé à une forte dépression de consanguinité, et qu'il est facilité par un taux élevé d'auto-pollinisation. La seconde partie traite des conditions permettant l'évolution de nouveaux allèles d'auto-incompatibilité (allèles S), ce que nous avons étudié d'une part analytiquement en population infinie, et d'autre part par des simulations en population finie. Nos résultats montrent que les conditions permettant la diversification au locus S sont beaucoup moins stringentes en population finie, et que le processus de diversification est d'autant plus rapide qu'il y a peu d'allèles S présents dans la population.

**Mots clefs :** Auto-incompatibilité, modélisation, système de reproduction, dépression de consanguinité.