Ecole doctorale : Science de la Matière, du Rayonnement et de l'Environnement
Laboratoire : Evolution, Ecologie, Paléontologie, UMR CNRS 8198

# Processus de spéciation et impact des systèmes de reproduction dans le genre *Silene*

## Spéciation rapide chez l'espèce gynodioïque *Silene nutans* et labilité des chromosomes sexuels dans la section *Otites*

**Hélène Martin**

Thèse présentée et soutenue publiquement le 9 décembre 2016
pour l'obtention du grade de Docteur de l'Université de Lille - 1 Sciences et Technologies
dans la discipline Biologie Evolutive et Ecologie

| | |
|---|---|
| Rapporteurs | **Tatiana Giraud** *(CNRS, Université Paris-Sud XI, France)* |
| | **Pierre Boursot** *(CNRS, Université Montpellier 2, France)* |
| Examinateurs | **Jacqui Shykoff** *(CNRS, Université Paris-Sud XI, France)* |
| | **Maud Tenaillon** *(CNRS, Université Paris-Sud XI, France)* |
| | **Nicolas Bierne** *(CNRS, Université Montpellier 2, France)* |
| | **Xavier Vekemans** *(Université de Lille, France)* |
| Directeur de thèse | **Pascal Touzet** *(Université de Lille, France)* |
| Invitée | **Fabienne Van Rossum** *(Jardin Botanique de Meise, Belgique)* |

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

**Résumé** Le processus de spéciation reflète l'origine et la mise en place de barrières à la reproduction. Etudier ce processus revient donc à répondre aux questions suivantes : comment les barrières à la reproduction se sont-elles mises en place, est-ce qu'elles permettent le maintien de l'intégrité génétique des populations et quelle est l'architecture génétique de l'isolement reproducteur ? L'objectif de ma thèse est d'apporter des premières réponses à ces questions dans le genre *Silene* tout en prenant en compte l'impact potentiel des systèmes de reproduction sur le processus de spéciation. Dans une première partie, je me suis concentrée sur l'espèce gynodioïque *S. nutans*. Par une approche phylogéographique et de génétique des populations, je me suis intéressée à la structuration génétique de *S. nutans* sur une partie de l'Europe de l'Ouest. Deux lignées génétiques présentant des patrons de recolonisation post-glaciaire distincts ont été identifiés. Ces deux lignées sont en contact secondaire en Angleterre, en Belgique et en France, mais aucun flux de gènes n'a été détecté, suggérant la présence de barrières à la reproduction entre elles. J'ai donc testé la présence d'un isolement reproducteur postzygotique en utilisant des données de croisements contrôlés. Un important degré d'isolement reproducteur a été observé entre les lignées et se traduisait par une importante proportion d'individus chlorotiques, phénotype menant à la mort au stade juvénile. Enfin, en retraçant l'histoire démographique de ces deux lignées génétiques par une approche ABC, j'ai pu montrer que l'hétérogénéité de la sélection (sélection indirecte) le long du génome a façonné le patron de différenciation génétique entre les deux lignées. Ce résultat m'a empêché d'associer directement les régions présentant de fortes valeurs de différentiation génétique à l'isolement reproducteur postzygotique observé. Dans une seconde partie, je me suis intéressée à l'effet de la dioécie sur la divergence des populations, et en particulier l'impact des chromosomes sexuels. J'ai testé le déterminisme (XY ou ZW) du sexe chez *S. otites* et *S. pseudotites*, à l'aide de données RNAseq issues de familles intra-spécifiques. J'ai identifié pour *S. pseudotites* un déterminisme du sexe de type XY, et deux paires de chromosomes pour *S. otites*, l'une présentant une ségrégation de type XY et l'autre de type ZW. Ce déterminisme du sexe polygénique pourrait participer à l'isolement reproducteur des deux taxons. Enfin, je me suis intéressée à l'hypothèse du cul-de-sac évolutif associée à la dioécie. En contrastant les niveaux de polymorphisme et de divergence de plusieurs espèces présentant des systèmes de reproduction différents, j'ai constaté que les espèces dioïques étaient moins efficaces pour fixer des mutations bénéfiques que les espèces gynodioïques pour des niveaux de diversité équivalents. Mes travaux de thèse sur *S. nutans* ont permis de comprendre comment les barrières à la reproduction se sont mises en place entre les deux lignées génétiques et avec quelle intensité elles agissent, mais je n'ai pas pu déterminer l'architecture génétique de cet isolement reproducteur. Enfin, la dioécie dans la section *Otites*, via la labilité des chromosomes sexuels, pourrait être un moteur dans la divergence des taxons.

Mots clés : spéciation, système de reproduction, *Silene nutans*, section *Otites*, phylogéographie, isolement reproducteur postzygotique, selection indirecte, chromosomes sexuels, dioécie

**Abstract** Speciation reflects the origins and further development of reproductive barriers. The study of speciation thus requires answering three main questions : which processes lead to reproductive barriers ; how effective these barriers are to maintain the genetic integrity of populations ; and what are their genetic basis. The goal of my thesis is to attempt to answer these questions in the context of the *Silene* genus, and to evaluate the potential impact of breeding systems in the speciation process. In the first part, I focused on *S. nutans*, a gynodioecious species. Using a mixed phylogeographic and population genetic approach, I investigated the genetic structure of *S. nutans* in Western Europe. I identified two main lineages, with contrasting patterns of post-glacial recolonisation, that are found in secondary contact in England, Belgium, and France but no detected gene flow, suggesting the presence of reproductive barriers. Thus, I tested the presence of postzygotic reproductive isolation using data from cross experiments. I observed a high level of reproductive isolation that is mainly due to a high proportion of chlorotic seedlings (leading to lethality) in between-lineage crosses. Finally, I reconstructed the demographic history of the lineages using an ABC approach. Results showed that the heterogeneity of selection (linked selection) along the genome has shaped the heterogeneity of genetic differentiation between the lineages. Therefore, I was unable to linked highly differentiated regions to reproductive isolation. In a second part, I focused on the impact of dioecy and sex chromosomes on population divergence. I tested the sex determinism (XY or ZW) of *S. otites* and *S. pseudotites* using RNAseq data from intraspecific families. I found XY sex determinism in *S. pseudotites* and polygenic sex determinism in *S. otites*, with a XY pair and a ZW pair of chromosomes. This contrasting sex determinism could be involved in the reproductive isolation between both taxa. Finally, I tested the evolutionary dead-end hypothesis for dioecy. By comparing levels of genetic polymorphism and divergence of several species with contrasting breeding systems, I found that selection was less effective in dioecious taxa compared to gynodioecious ones for similar levels of genetic diversity. My work on *S. nutans* furthers the understanding how reproductive barriers developed, as well as how effective they are, but their genetic architecture could not be determined. Finally, in the *Otites* section, I found that dioecy could play an important role in taxa diversification due to sex chromosome lability.

Keywords : speciation, breeding systems, *Silene nutans*, *Otites* section, phylogeography, postzygotic reproductive isolation, linked selection, sex chromosomes, dioecy



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

# Remerciements

Ce travail est le fruit d'une confiance mutuelle, de discussions régulières et d'un soutien sans faille. Pascal Touzet, je te suis extrêmement reconnaissante de m'avoir laissé la liberté d'orienter mon sujet de thèse à ma façon [1], d'avoir toujours été présent pour répondre à mes interrogations et mes inquiétudes, et de m'avoir toujours fourni la possibilité de mener à bien mes projets (moyens humains, matériels et financiers). Fabienne Van Rossum, je tiens à te remercier tout particulièrement pour m'avoir initié à la problématique de la spéciation chez *Silene nutans* (dès 2012!), et d'avoir été présente tout au long de cette thèse pour discuter de mes travaux et les corriger, toujours avec beaucoup de minutie.

Autour de ce duo d'encadrants ont gravité Jean-François Arnaud, Mathilde Dufaÿ, Xavier Vekemans, Camille Roux, Aline Muyle et Gabriel Marais. Merci de m'avoir guidé dans les analyses, d'avoir partagé des réunions prise de tête avec mes résultats et d'avoir patiemment calmé mes inquiétudes. Merci aux membres extérieurs de mon jury d'avoir accepté d'évaluer mon travail et d'être venus (jusqu'à Lille!) discuter de mes recherches, un matin de décembre : Tatiana Giraud, Pierre Boursot, Jacqui Shykoff, Maud Tenaillon et Nicolas Bierne.

Entourée de toutes ces personnes, les idées ont fusé de toutes part, mais leur mise en pratique n'a été possible qu'avec le concours de Cécile Godé, Eric Schmidt, Sophie Gallina et Sandrine Belingheri. Merci de m'avoir accompagné dans les manips de biologie moléculaire, dans le relevé des plantes en serre, à travers l'écran noir du terminal et dans les méandres de l'administration. Merci également à Bilille et tout son personnel, sans qui les calculs présentés dans cette thèse auraient probablement duré beaucoup plus longtemps (rep("merci", 1000);D). Je n'oublie pas aussi les stagiaires qui ont subi mon apprentissage de l'encadrement : Fanjamalala Rajaonalison, Thomas Lesaffre, Henri Deveyer, Manon Cardon et Fantin Carpentier.

I am grateful to Lynda Delph for allowing me to come and work in her lab at Bloomington. Mandy and you were very helpful in pushing my way of thinking further and clearer.

Je tiens à remercier l'ensemble du laboratoire pour les moments de discussion et de détentes en salle café ... et ailleurs. En particulier, les (post-)doctorant(e)s (passé(e)s et présent(e)s), auprès desquels j'ai toujours trouvé un échappatoire et un exutoire. Et une pensée particulière pour Dialise, Leslizbeth et Glaire, les trois grâces. Merci également à Déborah, les plans classiques nous auront finalement bien servi ...

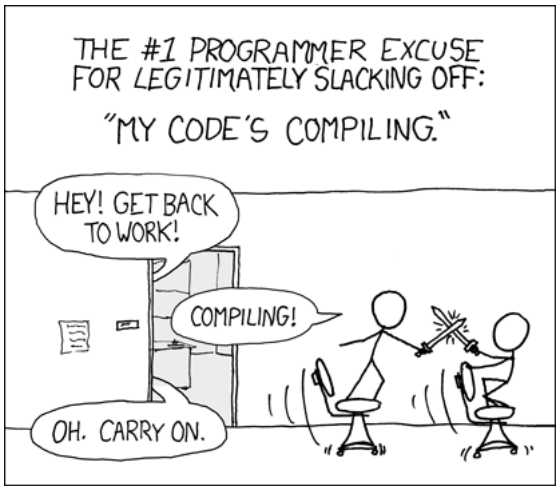
Enfin, merci à ma famille et belle-famille pour leur présence réconfortante. Et parce que tu aurais pu être co-auteur de cette thèse, merci Jon. *Dal mek creon te shree tal'ma.*

1. Rappelons ici le titre initial de ma thèse : determinants génétiques et impact de la transition vers la dioécie chez *Silene*, une approche de génomique des populations



© 2016 Tous droits réservés.

lilliad.univ-lille.fr





© 2016 Tous droits réservés.
lilliad.univ-lille.fr

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



# Table des matières





© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016







© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016





© 2016 Tous droits réservés.



lilliad.univ-lille.fr

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

# Introduction

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



## 1.1 Contexte

Après avoir longtemps été décrite comme une succession de séparations dichotomiques, la spéciation est actuellement considérée comme un continuum le long duquel se déplace une paire de populations, avec à une extrémité une population panmictique et à l'autre extrémité deux espèces irréversiblement isolées (Seehausen et al., 2014). En fonction de la présence de barrières à la reproduction et du degré d'isolement qu'elles génèrent, une paire de populations peut évoluer vers une population panmictique, en l'absence de barrières, ou vers deux espèces distinctes, lorsque l'isolement reproducteur est complet. Bien que le concept biologique de l'espèce stipule que les espèces sont des groupes de populations interfertiles mais isolés reproductivement d'autres groupes (Mayr and Ashlock, 1991), il n'est pas rare d'employer le terme espèce alors même que des événements d'hybridation sont observés. On fait alors le pari que les barrières à la reproduction entre les espèces sont suffisamment fortes pour maintenir l'intégrité génétique des groupes de populations jusqu'à ce que l'isolement reproducteur soit complet. Les bases génétiques de l'isolement reproducteur sont multiples (Nei and Nozawa, 2011) et s'accumulent au cours du temps, avec la divergence des populations (Orr, 1995), ce qui fait de la spéciation un processus temporel.

Un des patrons (peut-être le plus général) observés dans l'étude de la spéciation est la règle d'Haldane (Delph and Demuth, 2016). Elle stipule que *dans la première génération issue d'un croisement entre deux taxons, si un sexe est absent, rare ou stérile, il s'agit du sexe hétérogamétique* (Haldane, 1922). Toutefois, ce patron est restreint aux taxons à sexes séparés (dioïque chez les plantes, gonochorique chez les animaux) avec un déterminisme du sexe génétique. Or, il existe bien d'autres systèmes de reproduction, en particulier chez les plantes.

L'introduction de cette thèse se concentre sur le processus de spéciation, défini comme la mise en place de barrières à la reproduction entre deux populations. Après avoir exposé la diversité des barrières à la reproduction, nous présenterons le mécanisme génétique de l'isolement reproducteur postzygotique intrinsèque en nous concentrant sur le modèle d'incompatibilité génétique. Pour comprendre comment ces incompatibilités peuvent se mettre en place, nous détaillerons brièvement deux forces évolutives sous-jacentes (la dérive génétique et la sélection).

Parce qu'une troisième force évolutive, le flux de gènes, peut limiter la mise en place d'incompatibilités génétiques, nous ferons un bref point sur le contexte géographique de la divergence des populations et dans quelle mesure ce contexte influe sur le type de barrières impliqué dans l'isolement reproducteur. L'intensité de ces trois forces évolutives pouvant évoluer au cours

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



du temps, nous soulignerons l'importance de retracer l'histoire démographique des populations afin d'interpréter au mieux les patrons de différenciation génétiques observés. Puis, nous verrons comment les systèmes de reproduction (et en particulier la dioécie) peuvent être impliqués dans la divergence des populations en impactant l'intensité de ces trois forces évolutives.

C'est dans ce contexte général que s'inscrit mon travail de thèse en prenant comme modèle biologique le genre *Silene*; modèle en biologie évolutive (Bernasconi et al., 2009) en particulier dans l'étude de l'évolution des systèmes de reproduction (Touzet and Delph, 2009, Käfer et al., 2013, Lahiani et al., 2013, Dufay et al., 2014) du fait de la diversité de systèmes rencontrés (hermaphrodisme, gynodioécie, dioécie; Casimiro-Soriguer et al., 2015), et de l'existence d'au moins deux transitions indépendantes vers la dioécie (Desfeux et al., 1996, Marais et al., 2011). Dans une dernière partie de l'introduction, nous présenterons les modèles biologiques étudiés dans le cadre de cette thèse.

## 1.2 Les barrières à la reproduction sont nombreuses

L'isolement reproducteur est le résultat de la présence de barrières à la reproduction. En fonction du moment du cycle de vie au cours duquel elles interviennent et de leur interaction avec l'environnement, on distingue les barrières prézygotiques des barrières postzygotiques et les barrières intrinsèques des barrières extrinsèques (Coyne and Orr, 2004).

### Barrières prézygotiques

Les barrières prézygotiques empêchent (ou limitent) la fécondation entre les gamètes mâles et femelles. Elles peuvent intervenir avant même l'accouplement ou la pollinisation, par exemple par la présence de caractères attractifs qui diffèrent entre espèces ou un décalage de la période de floraison. Chez les poissons Cichlidae par exemple, la couleur du poisson mâle est un caractère utilisé par les femelles pour choisir leur partenaire, ce qui favorise la reproduction intraspécifique (Selz et al., 2014). Le décalage de la période de floraison entre *Mimulus nasutus* et *M. guttatus* peut atteindre jusqu'à 18 jours en moyenne dans certains sites en populations naturelles ; ce qui contribue à limiter la formation d'individus hybrides, bien que la distribution des phénologies de floraison des deux espèces se superpose entre la fin de floraison de *M. nasutus* et le début de floraison de *M. guttatus* (Martin and Willis, 2007). Si l'accouplement ou la pollinisation a lieu, il est tout de même encore possible de limiter la fécondation entre les gamètes. Chez les plantes,

© 2016 Tous droits réservés.



lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



une barrière prézygotique post-pollinisation peut impliquer l'interaction spécifique entre le pollen et le stigmate, par exemple, dans le genre *Pedicularis* (Tong and Huang, 2016), *Silene* (Van Rossum et al., 1996, Nista et al., 2015) ou encore *Orchis* (Luca et al., 2015).

### Barrières postzygotiques

Une fois que la fécondation a eu lieu, les barrières postzygotiques interviennent tout au long du cycle de vie de l'individu hybride et limitent sa survie. Elles peuvent intervenir tôt dans le stade de développement ; par exemple à l'état embryonnaire chez le poisson entre les deux écotypes de *Coregonus clupeaformis* (Lu and Bernatchez, 1998) ou au stade larvaire entre les moules *Mitylus edulis* et *M. galloprovincialis* (Bierne et al., 2002). Si l'individu hybride survit jusqu'à l'âge adulte, sa fertilité peut également être limitée comme c'est le cas des hybrides chez la souris entre *Mus musculus domesticus* et *Mus m. musculus* (Britton-Davidian et al., 2005) ou entre des populations éloignées de la plante *Draba nivalis* (Skrede et al., 2008).

### Barrières en interaction avec l'environnement

Outre le cycle de vie, la dimension environnementale est également utilisée pour classer les barrières à la reproduction (Coyne and Orr, 2004). Les barrières postzygotiques présentées plus haut sont intrinsèques : l'individu hybride présente des problèmes de développement sans lien avec l'environnement dans lequel il se trouve. Au contraire, les barrières postzygotiques sont extrinsèques si la survie ou le succès de reproduction d'un individu hybride dépend de l'environnement dans lequel il se trouve. Par exemple, les individus hybrides entre deux écotypes de *Mimulus guttatus* (un côtier, tolérant au sel, et un continental plus sensible à la salinité) présentent un phénotype intermédiaire et une relativement bonne tolérance au sel, ce qui semble le rendre mal adapté à l'habitat continental mais performant dans l'habitat côtier (Lowry et al., 2008b).

Il existe donc de nombreuses barrières à la reproduction qui limitent les flux de gènes entre les populations. Le degré de l'isolement reproducteur dépend de l'efficacité et du nombre de barrières. En comptabilisant 12 barrières à la reproduction chez 19 taxons végétaux, Lowry et al. (2008a) montrent que c'est la combinaison des différentes barrières qui rend l'isolement reproducteur complet, et suggèrent que les barrières prézygotiques sont plus importantes que les barrières postzygotiques. Toutefois, ce n'est que lorsque les barrières postzygotiques sont importantes que le processus de spéciation devient irréversible et aboutit à la formation d'espèces

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



complètement isolées (Seehausen et al., 2014).

## 1.3 Modèles d'incompatibilité génétique à l'origine des barrières postzygotiques

Les barrières postzygotiques intrinsèques sont attribuées à un dysfonctionnement entre les génomes parentaux qui se retrouvent ensemble au sein de l'individu hybride. Cela peut être dû, par exemple, à d'importants réarrangements chromosomiques, comme c'est le cas entre les deux espèces de tournesols *Helianthus petiolaris* et *H. annuus* (Rieseberg et al., 1999, Rieseberg, 2001). Mais le plus souvent, ce ne sont que quelques locus (entre deux et cinq) qui sont impliqués dans le dysfonctionnement intrinsèque de l'hybride (Fraïsse et al., 2014).

### Incompatibilité nucléaire : le modèle Bateson-Dobzhansky-Muller à deux locus

Le modèle d'incompatibilité génétique à deux locus nucléaires proposé par Bateson, Dobzhansky et Muller (modèle IBDM; Coyne and Orr, 2004) permet d'expliquer comment un dysfonctionnement peut se produire chez l'individu hybride sans avoir eu lieu chez les individus des populations parentales (Figure 1.1a). Une population diploïde ancestrale de génotype *aa* au locus 1 et *bb* au locus 2 se divise en deux populations filles. Dans la première population, l'allèle *A* apparaît par mutation au locus 1 et se fixe, alors que le locus 2 maintient le génotype ancestral (*bb*). Dans la seconde population, l'allèle *B* apparaît par mutation au locus 2 et se fixe alors que le locus 1 maintient le génotype ancestral (*aa*). Les populations filles sont donc maintenant respectivement *AAbb* et *aaBB* (Figure 1.1a). Les allèles dérivés *A* et *B* ne se sont jamais retrouvés ensemble. Il est donc possible qu'ils n'interagissent pas correctement entre eux si les deux populations s'hybrident. Cette incompatibilité génétique peut aboutir à la stérilité ou à la mort de l'individu qui porte les deux allèles *A* et *B*. La mortalité des individus hybrides entre les deux espèces de poisson *Xiphophorus maculatus* et *X. helleri* est un exemple de l'effet d'une incompatibilité génétique à deux locus (Orr and Presgraves, 2000).

### Incompatibilité nucléo-cytoplasmique

Lorsque l'incompatibilité implique deux locus provenant de génomes différents (nucléaire et cytoplasmique) on parle alors d'incompatibilité nucléo-cytoplasmique. Par exemple, une population ancestrale présente, à un locus nucléaire, le génotype *aa* et, à un locus cytoplasmique,

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



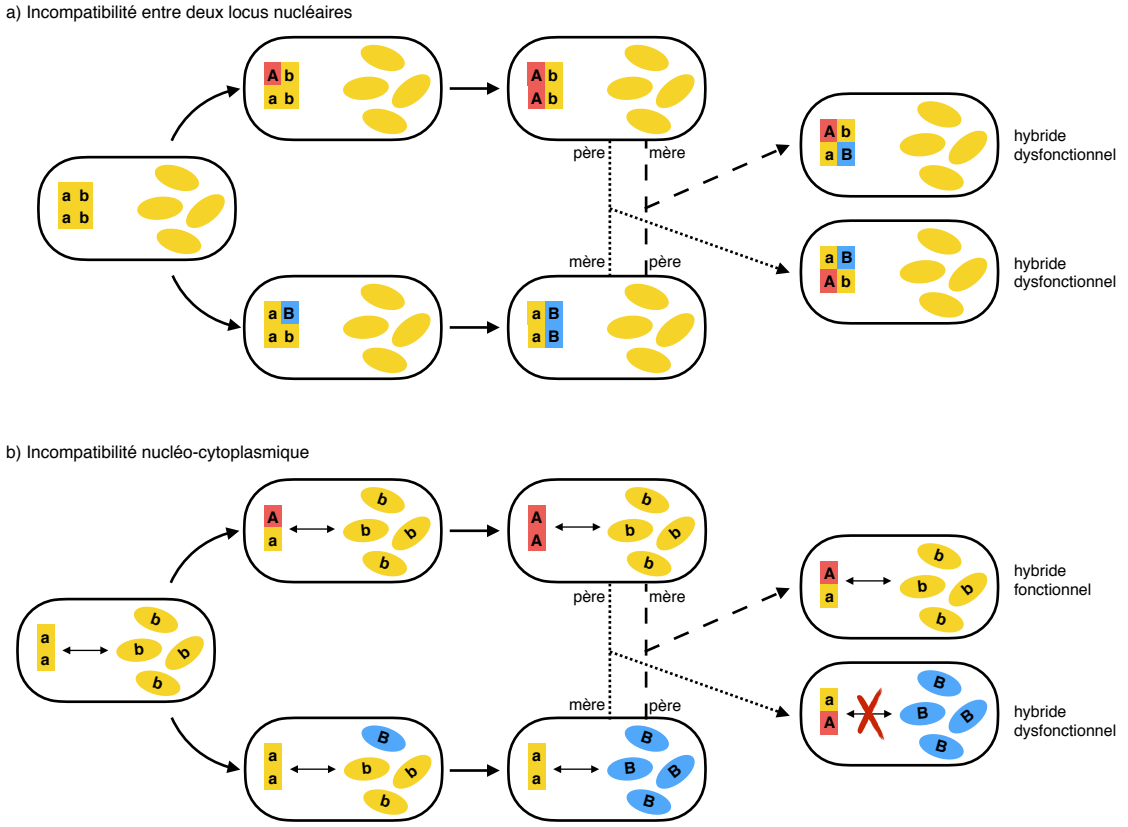



**Figure 1.1 :** Représentation schématique d'un exemple a) d'incompatibilité génétique au sein du génome nucléaire et b) d'incompatibilité nucléo-cytoplasmique.

l'allèle *b*, se divise en deux populations filles. Dans la population fille numéro 1, l'allèle *A* apparaît au locus nucléaire et se fixe dans la population, alors que dans la population fille numéro 2, l'allèle *B* apparaît au locus cytoplasmique et se fixe dans la population (Figure 1.1b). Si la transmission du génome nucléaire et des génomes cytoplasmiques n'est pas la même (chez les angiospermes, le génome nucléaire est transmis par les voies mâle et femelle alors que les génomes mitochondriaux et chloroplastiques sont transmis uniquement par la voie femelle), les génotypes des individus hybrides ne vont pas être les même en fonction de l'identité de la mère (Figure 1.1b). Si les allèles *A* et *B* n'interagissent pas correctement et génèrent un dysfonctionnement chez l'individu, seuls les individus hybrides issus des croisements ayant la populations fille numéro 2 comme parent maternel souffrent d'un dysfonctionnement (Figure 1.1b).

Chez les plantes, ce type d'incompatibilité, auquel on attribue un phénotype chlorotique (feuilles en partie ou totalement blanches probablement lié à une dysfonctionnement de la structure photosynthétique), est souvent observé (Greiner et al., 2011, a dénombré ce type d'incompatibilité dans neuf familles). Mais d'autres facteurs peuvent créer une asymétrie dans les croisements réciproques, tels que des incompatibilités entre des gènes portés par le chromosome sexuel X et des gènes autosomaux (Turelli and Moyle, 2007).

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



### Effet *boule de neige*

Le modèle IBDM prédit que les incompatibilités génétiques vont s'accumuler de plus en plus rapidement au cours du temps (selon une fonction quadratique ; Orr, 1995). Cet effet *boule de neige* a été démontré au sein du genre *Solanum* (Moyle and Nakazato, 2010). Les auteurs ont trouvé une relation quadratique entre le nombre de locus de caractères quantitatifs (Quantitative Trait Loci, QTL) associés à la stérilité des graines et le niveau de divergence entre trois paires d'espèces (Figure 1.2). Cet effet *boule de neige* est attendu pour des traits déterminés par des relations épistatiques (un locus influence l'expression d'un autre). Au contraire, parce que la différence dans la forme du fruit est plus probablement déterminée par des relations additives (en proportion du nombre de substitution affectant les locus concernés), la relation attendue entre le nombre de QTL et le niveau de divergence est linéaire (Figure 1.2; Moyle and Nakazato, 2010). En allant au bout du raisonnement, Orr (1995) suggère que l'isolement reproducteur postzygotique, parce que déterminé nécessairement par des relations épistatiques (sous le modèle IBDM), devrait évoluer plus vite que des différences morphologiques déterminées par des relations additives, entre les populations. Toutefois, le nombre d'incompatibilités génétiques ne renseigne pas forcément sur le degré de l'isolement reproducteur (Moyle and Nakazato, 2010).

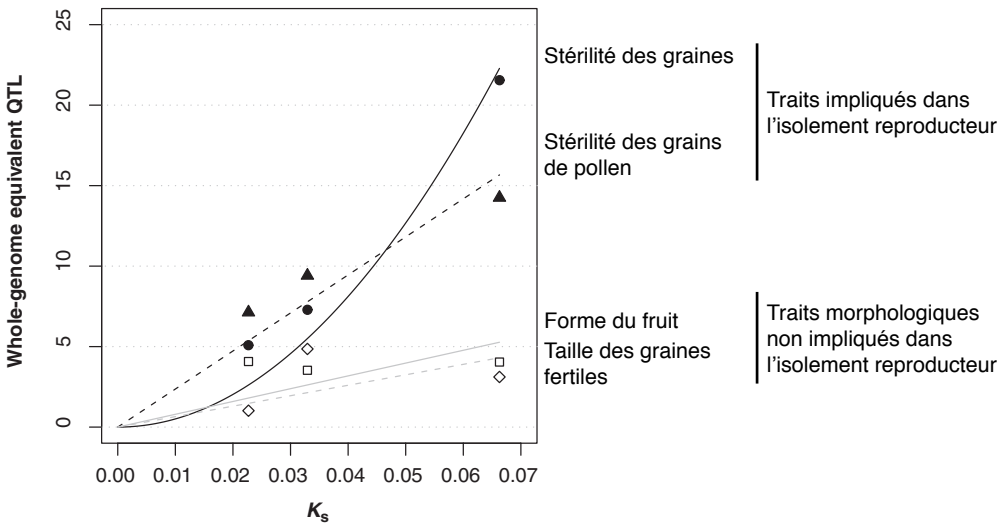



**Figure 1.2 :** Accumulation de QTL en fonction de la divergence ($K_S$) entre paires d'espèces de *Solanum*. L'accumulation de QTL impliqués dans la stérilité des graines augmente de façon non linéaire en fonction de la divergence contrairement aux autres traits mesurés (d'après Moyle and Nakazato, 2010)

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



## 1.4 Processus de divergence

### Dérive génétique

Si un nouvel allèle apparaît par mutation dans une population et qu'il ne confère pas d'avantage (ou de désavantage) sélectif, sa fréquence dans la population évolue de manière aléatoire, par dérive génétique. Le temps de fixation d'un allèle rare est d'environ *4Ne* générations, avec Ne la taille efficace de la population (Kimura and Ohta, 1969). Donc plus la taille efficace de la population est petite, plus le temps de fixation d'un allèle par dérive génétique est court. Grossièrement, si on estime un temps de spéciation de l'ordre de cinq millions d'années (Levin and Scarpino, 2016), cela représenterait, pour une espèce annuelle (une génération par an), une taille efficace de l'ordre de maximum 1.250.000 individus pour fixer un allèle rare dans ce laps de temps. Cette très grossière estimation de la taille efficace requise pour fixer des allèles par dérive en cinq millions d'années chez une espèce annuelle est supérieure à l'estimation de la taille efficace chez les plantes (par exemple, la taille efficace de *Arabidopsis thaliana* et *Helianthus annuus* est estimé à 130 000 et 830 000 respectivement, Gossmann et al., 2010). Même si pour des espèces pérennes (moins d'une génération par an), le temps de génération plus grand nécessite une taille efficace plus petite pour fixer des mutations par dérive en cinq million d'années, les tailles efficaces estimées sont suffisamment petites pour permettre la mise en place des incompatibilités par fixation d'allèles par dérive.

### Sélection et recombinaison

La divergence entre les populations peut être accélérée par sélection positive. Si la mutation qui apparaît dans une population confère un avantage sélectif, la sélection naturelle a pour effet d'accélérer le temps de fixation (Li and Nei, 1977). Quelque études de cas ont d'ailleurs montré des traces de sélection à des locus impliqués dans l'isolement reproducteur chez la mouche *Drosophila* (Presgraves et al., 2003, Brideau et al., 2006, Tang and Presgraves, 2009). En envahissant la population, l'allèle sélectionné emporte avec lui des locus qui lui sont liés génétiquement (Figure 1.3; Charlesworth et al., 1997). Cet auto-stop génétique est d'autant plus important que la recombinaison est faible, et permet de fixer des allèles indirectement sélectionnés (Kaplan et al., 1989). Si la mutation qui apparaît est délétère, l'allèle contre-sélectionné a un temps de fixation bien plus important (Kimura, 1980). Toutefois, la sélection négative va avoir pour effet de diminuer la variation génétique du locus lié génétiquement au locus sous sélection (Figure


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



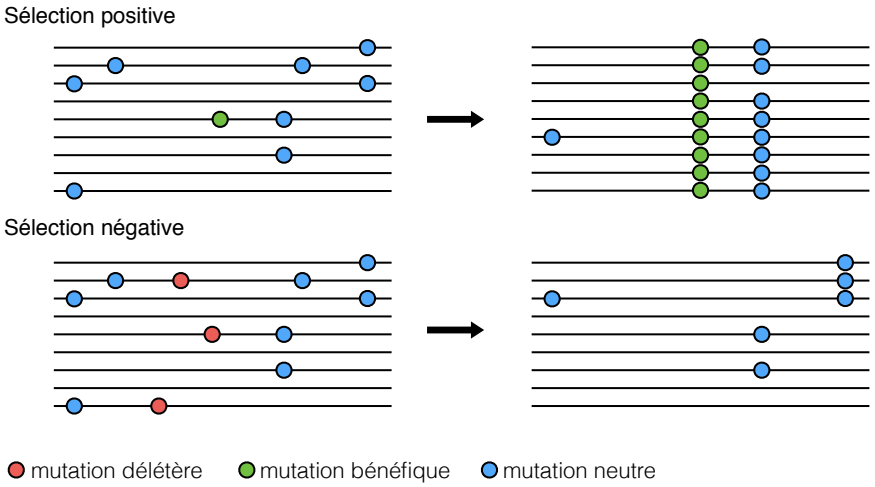



**Figure 1.3 :** Illustration de l'effet de la sélection indirecte sur la diversité aux locus neutres, dans un scénario de sélection positive et négative (d'après Josephs and Wright, 2016)

1.3; Charlesworth et al., 1997, 1993). La sélection indirecte qui agit aux locus liés diminue leur taille efficace, ce qui accélère le temps de fixation, par dérive génétique, des allèles neutres.

Comme pour la sélection positive, un faible taux de recombinaison peut amplifier le phénomène (Charlesworth et al., 1993). L'environnement génomique joue donc un rôle important dans le temps de fixation d'un allèle rare apparu par mutation. Chez le tournesol *Helianthus* (Renaut et al., 2013) ou l'oiseau gobe-mouche *Ficedula* (Burri et al., 2015) par exemple, les régions à fort niveau de $F_{ST}$ ntre les taxons sont également des régions à faible taux de recombinaison.

## Coévolution entre gènes

La divergence des lignées peut être favorisée par la coévolution entre locus au sein des lignées parentales, générant des incompatibilités génétiques chez l'individu hybride (Orr and Presgraves, 2000, Presgraves, 2010). La pression de mutation et les conflits génétiques sont deux moteurs de la coévolution entre gènes.

### Pression de mutation

La coévolution par pression de mutation est le résultat d'une mutation légèrement délétère à un locus qui se fixe dans une population et génère une pression de sélection en faveur d'une mutation compensatrice à un autre locus avec lequel il est en interaction (Presgraves, 2010). Si cette coévolution ne se produit que dans une seule lignée parentale, l'individu hybride peut souffrir de dysfonctionnement développemental parce que la mutation légèrement délétère n'est plus compensée.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



Ce mécanisme est proposé pour expliquer l'isolement postzygotique entre les deux levures *Saccharomyces bayanus* et *S. caravisiae* (Barbash, 2008). L'hypothèse est qu'un relâchement de la pression de sélection au niveau de la région 5'-UTR du gène mitochondrial *OLI1* aurait favorisé le maintien d'une mutation légèrement délétère chez *S. bayanus* et entraîné une sélection en faveur d'une mutation compensatrice du gène nucléaire *AEP2*, maintenant leur interaction fonctionnelle et aboutissant à la traduction de l'ARN messager de *OLI1* (Barbash, 2008, Presgraves, 2010). Dans plusieurs autres cas, la coévolution entre un gène nucléaire et un gène mitochondrial semble impliquée dans la divergence entre deux lignées : chez les anguilles *Anguilla anguilla* et *A. rostrata* (Gagnaire et al., 2012), les guêpes *Nasonia girault*i et *N. vitripennis* (Ellison et al., 2008) ou le copépode *Tigriopus californicus* (Harrison and Burton, 2006).

### Conflits génétiques

Les conflits génétiques entre un élément égoïste et un restaurateur favorisent également la divergence entre deux lignées (Presgraves, 2010, Werren, 2011). Un élément génétique égoïste possède des caractéristiques qui améliorent sa propre transmission par rapport au reste du génome, même si c'est au détriment de l'individu qui le porte. Parce que la meilleure transmission de l'élément égoïste se fait au détriment de la transmission du reste du génome, un conflit se met en place. Un élément restaurateur qui possède des caractéristiques pouvant réprimer l'élément égoïste est sélectionné ; s'en suit une course à l'armement entre les deux éléments (Werren, 2011). Si un conflit génétique a eu lieu dans une seule lignée, les individus hybrides entre les deux lignées peuvent exprimer l'élément égoïste car le restaurateur se trouve à l'état hétérozygote. Ce type de conflit semble être à l'origine de barrières postzygotiques chez plusieurs espèces, par exemple, chez la mouche *Drosophila* (Phadnis and Orr, 2009) ou chez la plante *Mimulus* (Case et al., 2016).

## 1.5 Contexte géographique de la mise en place des barrières à la reproduction

### Allopatrie versus sympatrie

Dans le modèle IBDM, la mise en place d'incompatibilités génétiques n'est possible qu'une fois la population ancestrale divisée en deux populations filles isolées. Selon Mayr et ses contemporains, c'est un isolement géographique qui permet la mise en place d'un isolement reproducteur


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



(Coyne, 1994). Dans ce scénario de spéciation allopatrique, deux populations d'une même espèce, isolées géographiquement, évoluent indépendamment sous l'effet de la dérive génétique et de la selection. Ici, l'isolement reproducteur est un sous-produit de la divergence entre deux populations en allopatrie (Coyne, 1994). Il n'est mis en évidence que lorsque les populations se retrouvent en contact et que les flux de gènes sont limités par des barrières postzygotiques intrinsèques.

Parce que les flux de gènes ont tendance à homogénéiser les populations et donc réduire leur divergence, il est plus facile de considérer la mise en place d'incompatibilités génétiques entre populations allopatriques qu'entre populations échangeant des flux de gènes. Toutefois, des modèles théoriques ont montré qu'un scénario de spéciation entre populations sympatrique est possible si un processus de sélection divergente se met en place (Gavrilets, 2004). Pour cela, il faut que l'environnement dans lequel se trouvent les individus soit hétérogène. Si deux allèles d'un locus sont différentiellement sélectionnés en fonction de l'habitat (allèle *a* sélectionné dans l'habitat 1 et l'allèle *A* sélectionné dans l'habitat 2), deux groupes génétiques peuvent alors apparaître et se maintenir (le génotype *aa* dans l'habitat 1 et le génotype *AA* dans l'habitat 2). Ainsi, les groupes génétiques présentes des airs de distribution disjointes mais contigües, on dit qu'ils sont en parapatrie (Coyne and Orr, 2004). Parce que les groupes génétiques présentent des caractéristiques écologiques différentes, on les définit alors comme des écotypes (Turesson, 1922). Chez les plantes plusieurs espèces présentes des écotypes distincts (chez *Silene nutans* De Bilde, 1973 ; chez *Festuca ovina*, Prentice et al., 1995 ; ou chez *Mimulus guttatus* Lowry et al., 2008b). Toutefois, pour que les écotype puisse diverger l'un de l'autre, cela nécessite de très forts coefficients de sélection et de faibles niveaux de migration (Gavrilets, 2004).

La sélection divergente peut également créer une situation favorable à l'évolution d'un isolement reproducteur prézygotique qui limiterai la fécondation inter-écotypes, comme par exemple l'homogamie (Gavrilets, 2004). Mais encore une fois, dans les modèles, la divergence des écotypes n'est possible que dans un espace de paramètre restreint (Felsenstein, 1981, Gavrilets, 2004) Mais l'homogamie n'est pas forcément nécessaire pour que le processus de spéciation puisse aboutir à la formation d'espèces distinctes en présence de flux de gènes (Feder et al., 2012). Le locus sélectionnée est lié physiquement à son environnement génomique local et par un phénomène d'autostop-moléculaire, la sélection va également agir sur les locus voisins (*divergence hitchhiking*). A terme, l'accumulation des locus impliqués dans l'adaptation va impacter tout le génome. Mais encore une fois, l'espace des paramètres pour que le processus de *divergence hitchhiking* puisse se faire sont limitées (peu de migration et petite taille efficace des populations, Feder and Nosil, 2010).

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



### Est-ce que le contexte géographique influe sur le type de barrière à la reproduction mise en place ?

Si deux populations divergentes sont en sympatrie, et si les descendants de croisements inter-population ont une moins bonne valeur sélective que les descendant de croisement intra-population, les croisements inter-population devrait alors être contre selectionnés. La sélection contre ces hybrides favoriserait alors un renforcement de l'isolement reproducteur par rapport à des situations en allopatrie où aucun hybride n'est formé (Noor, 1999, Hopkins, 2013). Parce que la sélection favoriserait toute stratégie diminuant le coût de l'hybridation, le renforcement impliquerait généralement des barrières prézygotiques (Hopkins, 2013). En estimant le degré d'isolement reproducteur généré par des barrières prézygotiques entre des paires d'espèces de mouche (*Drosophila*) de niveaux de divergence génétique différents, (Coyne and Orr, 1997) ont trouvé une évolution plus rapide des barrières prézygotiques entre des espèces sympatriques qu'entre des espèces allopatriques, en accord avec la notion de renforcement.

Toutefois, des modèles théoriques ont montré que le renforcement pouvait être limité et dépendait entre autres du nombre d'allèles impliqués dans l'isolement prézygotique, du degré de changement généré par la mutation ou encore du taux de recombinaison entre le locus impliqué dans l'isolement prézygotique et le ou les locus d'adaptation ou d'incompatibilité (Felsenstein, 1981, Bank et al., 2011). En comparant le degré d'isolement reproducteur généré par une barrière prézygotique entre espèces sympatriques et allopatriques, Moyle et al. (2004) ne distinguent pas de différence, et ce pour deux genres végétaux différents (*Glycine* et *Silene*). De même, aucun renforcement n'a été observé dans le genre végétal *Nolana* (Jewell et al., 2012).

## 1.6 Intensité des flux de gènes dans le temps et dans l'espace

### L'intensité des flux de gènes change au cours du temps

La notion d'allopatrie et de sympatrie est à prendre en compte d'un point de vue temporel. Des populations actuellement en sympatrie ne l'étaient peut-être pas dans le passé. Les cycles climatiques du Quaternaire (2,4 millions d'année à nos jours) ont largement contribué à la structuration géographique et génétique des populations actuelles (Hewitt, 1996, Waters et al., 2013). Lors des périodes glaciaires, la répartition géographique des espèces tempérées s'est rétractée, des populations d'une même espèce se retrouvant ainsi géographiquement isolées dans des zones dites refuges. Et lors des périodes interglaciaires, plus chaudes, les espèces ont recolonisé le territoire


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



à partir de ces refuges, remettant en contact géographique des populations précédemment isolées. En étudiant les fréquences alléliques de marqueurs chloroplastiques dans de nombreuses populations, les premières études phylogéographiques ont permis d'identifier des lignées génétiques divergentes, des refuges glaciaires à partir desquels les lignées se sont répandues et des zones de contact où deux lignées génétiques se rencontrent et peuvent s'hybrider (Taberlet et al., 1998, Hewitt, 1999, 2011). On retrouve de tels patrons phylogéographiques chez beaucoup d'espèces, animales (le campagnol *Microtus agrestis*, Jaarola and Searle, 2002 ; la grenouille *Rana arvalis*, Babik et al., 2004 ; ou encore la chouette *Strix aluco*, Brito, 2005) ou végétales (le pin *Pinus sylvestris*, Cheddadi et al., 2006 ; le pommier sauvage *Malus sylvestris*, Cornille et al., 2013 ; ou encore la bettrave *Beta maritima*, Leys et al., 2014).

## L'intensité des flux de gènes change au sein du génome

Les zones hybrides, issues soit d'une divergence en sympatrie (contact primaire), soit par la remise en contact de lignées qui ont divergé en allopatrie (contact secondaire), sont des lieux privilégiés pour étudier les bases génétiques des barrières à la reproduction (Barton and Bengtsson, 1986). Les locus liés à l'isolement reproducteur entre deux populations (impliqués dans l'adaptation à un habitat ou dans des incompatibilités génétiques) limitent les flux de gènes car les allèles qu'ils portent sont contre-sélectionnés en fonction de la population dans laquelle ils se trouvent. Au contraire, les locus neutres, non soumis à sélection, devraient migrer librement entre les populations. Mais parce qu'ils sont liés aux locus sous sélection (ils appartiennent au même fond génétique), les flux de gènes aux locus neutres peut être réduit si le taux de recombinaison entre les locus neutres et sélectionnés est faible (Barton and Bengtsson, 1986).

Le long du génome, il y a donc des régions qui font barrière aux flux de gènes, maintenant une différenciation génétique entre les populations, et des régions qui sont homogénéisées par les flux de gènes, gommant la différenciation génétique. Sur un scan génomique, les pics de $F_{ST}$ dénommés également ilôts de différenciation, ou ilôts de spéciation) sont alors interprétés comme des régions impliquées dans l'isolement reproducteur (Turner et al., 2005, Nadeau et al., 2012, Gagnaire et al., 2013). L'hétérogénéité du niveau de différenciation génétique peut donc être interprétée comme une hétérogénéité des flux de gènes (Figure 1.4; Cruickshank and Hahn, 2014).

Mais ce patron hétérogène du niveau de différenciation génétique peut également être observé entre des populations allopatriques, sous l'action hétérogène de la sélection le long du génome (Figure 1.4; Noor and Bennett, 2009, Cruickshank and Hahn, 2014). La sélection qui agit

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



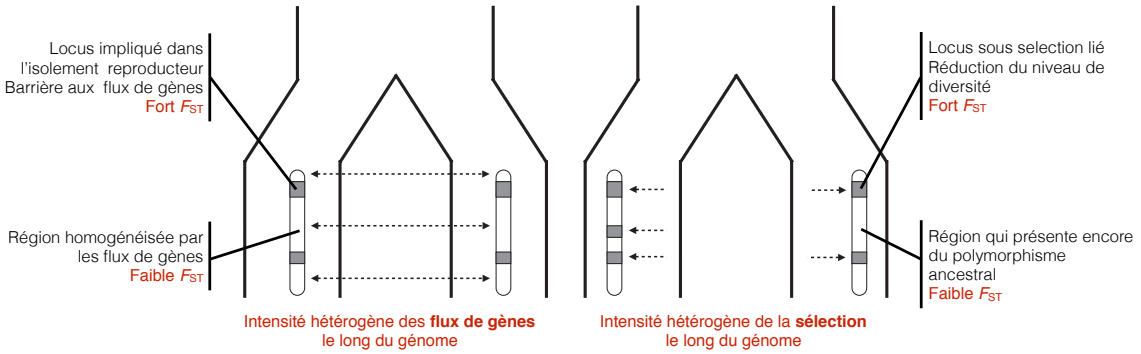



**Figure 1.4 :** Scénarios de spéciation, à gauche, avec une intensité des flux de gènes hétérogène le long du génome, et à droite, avec une intensité de la sélection hétérogène le long du génome. Les deux scénarios génèrent des variations du niveau de différenciation le long du génome (d'après Cruickshank and Hahn, 2014).

sur un locus a pour effet de diminuer sa diversité génétique (Charlesworth et al., 1997). Parce que la diversité est faible, toutes les différences entre populations sont mécaniquement importantes (Charlesworth et al., 1997). Dans ce scénario, les pics de $F_{ST}$ ont des régions sous sélection mais rien ne présume qu'elles interviennent dans l'isolement reproducteur.

## 1.7 Isolement reproducteur et systèmes de reproduction

Le système de reproduction chez les plantes est très varié. Chez les plantes à fleurs, il y a des espèces hermaphrodites (les organes mâles et femelles sont présents au sein de la même fleur) auto-compatibles ou auto-incompatibles, des espèces monoïques (chez un individu, les organes mâles et femelles sont séparés dans des fleurs différentes), gynodioïques (présence d'individus hermaphrodites et d'individus femelles), androdioïques (présence d'individus hermaphrodites et d'individus mâles) ou encore dioïques (présence d'individus femelles et d'individus mâles), pour ne citer que les plus fréquents (Barrett, 2002).

### Transition dans le régime de reproduction et isolement reproducteur

La transition de l'allofécondation vers l'autofécondation est une transition majeure chez les plantes à fleurs et peut favoriser des évènements de spéciation (Foxe et al., 2009, Wright et al., 2013). L'autofécondation limite les évènements d'hybridation en favorisant l'auto-pollen aux dépens de l'allo-pollen pour la reproduction. Cette barrière prézygotique s'est révélée très importante dans l'isolement reproducteur chez *Mimulus* (Martin and Willis, 2007) ou chez *Centaurium* (Brys et al., 2016). Mais parce que cette barrière s'applique tout aussi bien entre espèces (limitant l'hybridation) qu'au sein d'une même espèce (limitant l'allofécondation), Coyne and Orr (2004)


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



ne la considèrent pas comme une barrière impliquée dans la spéciation. Toutefois, l'autogamie est associé à un syndrome d'autofécondation qui comprend entre autres une réduction de la taille des pétales (Sicard and Lenhard, 2011). Ce changement dans la morphologie des fleurs associé aux préférences des pollinisateurs pourrait réduire les flux de pollen entre espèces majoritairement autogames et majoritairement allogames (Martin and Willis, 2007, Wright et al., 2013).

### Gynodioécie et conflits génétiques

La gynodioécie est un système de reproduction relativement rare chez les angiospermes, seulement présent dans 2% des genres (Dufay et al., 2014). Le déterminisme du sexe est généralement induit par un conflit génétique entre des locus mitochondriaux qui stérilisent les organes mâles (stérilité mâle cytoplasmique) et des allèles nucléaires qui restaurent la fonction mâle. Le polymorphisme sexuel (femelle et hermaphrodite) peut être maintenu sous l'action de la sélection et de la migration (Dufay and Pannell, 2010), comme c'est le cas chez la betterave (*Beta maritima*; Dufay et al., 2009). Toutefois, le polymorphisme sexuel peut disparaître suite à la fixation de la stérilité mâle cytoplasmique et de son restaurateur. L'espèce redevient alors hermaphrodite avec une stérilité mâle cytoplasmique cryptique. Lors de croisements hybrides entre deux lignées hermaphrodites, la stérilité mâle cytoplasmique peut alors être révélée si le conflit nucléo-cytoplasmique s'est produit dans une des deux lignées. Ce scénario est proposé pour expliquer la présence d'individus hybrides mâle-stériles en F2 dans les croisements entre *Mimulus guttatus* et *M. nasutus* (Fishman and Willis, 2006, Case et al., 2016).

La gynodioécie est aussi considérée comme une étape transitoire possible dans l'évolution vers la dioécie à partir d'un ancêtre hermaphrodite (Barrett, 2002). Bien qu'une transition vers la dioécie est possible à partir d'une gynodioécie nucléo-cytoplasmique (Maurice et al., 1994), c'est le modèle qui part d'une gynodioécie nucléaire-nucléaire (Charlesworth and Charlesworth, 1978) qui est le plus classiquement présenté dans l'évolution des chromosomes sexuels (Charlesworth et al., 2005).

### Dioécie et chromosomes sexuels

Parce que les chromosomes sexuels ont une évolution particulière (arrêt de la recombinaison et dégénérescence du chromosome Y, Charlesworth and Charlesworth, 2000 ; renouvellement des chromosomes sexuels, Bachtrog et al., 2014), il n'est probablement pas étonnant qu'ils prennent

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



une part important dans la divergence des lignées génétique et la spéciation (Demuth, 2014). La règle d'Haldane est sans doute le patron le plus général observé dans le cadre de la spéciation (Schilthuizen et al., 2011, Delph and Demuth, 2016). Elle stipule que *dans la première génération issue d'un croisement entre deux taxons animaux, si un sexe est absent, rare ou stérile, il s'agit du sexe hétérogamétique* (c'est-à-dire les mâle XY ou les femelles ZW Haldane, 1922). Cette théorie a depuis été étendue au règne végétal (Brothers and Delph, 2010). Ce phénomène peut être le résultat de différents mécanismes génétiques (Delph and Demuth, 2016).

Si on se place en présence de chromosomes sexuels hétéromorphiques (le chromosome X est morphologiquement différent du chromosome Y sur un caryotype), le chromosome X se retrouve à l'état hémizygote chez les mâles. Si une mutation adaptative récessive apparaît, elle sera plus rapidement fixée par sélection positive si elle se trouve sur le chromosome X que sur un autosome car elle est forcément exprimée chez les mâles ; le chromosome X évoluerait alors plus rapidement que les autosomes (*faster-X theory*; Charlesworth et al., 1987, Meisel and Connallon, 2013). Si ces substitutions sont impliquées dans une incompatibilité génétique et qu'elles sont récessives chez l'individu hybride, alors elles seront uniquement exprimées chez les mâles et la règle d'Haldane serait observé (*dominance theory*, Delph and Demuth, 2016). Le caractère récessif des incompatibilités a été montré chez *Drosophila* (Presgraves, 2003).

Toutefois, la règle d'Haldane ne nécessite pas forcément une hétéromorphie des chromosomes sexuels. Une hypothèse alternative est la théorie des conflits intra-génomiques (*meiotic drive theory*). Si un gène à l'origine d'une distorsion de ségrégation (un gène égoïste) se trouve sur un chromosome sexuel et que son restaurateur se situe sur l'autre chromosome sexuel (et/ou un autosome), des évènements d'hybridation peuvent générer un biais de sex-ratio ou de la stérilité , comme attendu dans la règle d'Haldane (Delph and Demuth, 2016). Cette théorie est mise en avant chez la mouche *Drosophila simulans* (Tao et al., 2007) ou chez la mouche *Teleopsis dalmanni* (Wilkinson et al., 2014).

Enfin, le renouvellement des chromosomes sexuels a également été suggéré comme impliqué dans l'isolement reproducteur postzygotique, chez de nombreuses espèces de poissons (Kitano and Peichel, 2012), mais aussi dans le genre *Silene* entre *S. latifolia* et *S. diclinis* (Demuth et al., 2014). Parce que les néo-chromosomes sexuels peuvent déjà porter des locus impliqués dans l'isolement reproducteur (Kitano et al., 2009, Yoshida et al., 2014) ou générer de l'aneuploïdie chez l'embryon (Demuth et al., 2014), le renouvellement rapide des chromosomes sexuels peut être impliqué dans la divergence et la spéciation entre deux taxons.


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



### La dioécie vue comme un cul-de-sac évolutif chez les plantes

Heilbuth (2000) a remarqué que chez les plantes la diversité spécifique des taxons dioïques était plus faible que celle de taxons hermaphrodites apparentés, ce qui l'a amené à considérer la dioécie comme un cul-de-sac évolutif. Deux hypothèses principales sont avancées pour expliquer ce phénomène. Parce que seules les femelles produisent des graines, la dispersion des graines est restreinte par rapport à des espèces hermaphrodites. Ce processus appelé *seed-shadow handicap* a pour effet d'augmenter la compétition locale entre les graines chez les taxons dioïques par rapport aux taxons hermaphrodites pour un même nombre de graines produites (Heilbuth et al., 2001). La seconde hypothèse propose que la sélection sexuelle peut conduire à un dimorphisme sexuel et jouer sur l'attractivité des pollinisateurs. Si les pollinisateurs sont rares, un seul des sexes peut être préférentiellement visité et donc aucune graine ne sera produite (Vamosi and Otto, 2002). Ces deux hypothèses conduisent à une réduction de la taille efficace des populations, ce qui réduit l'efficacité de la sélection, et donc leur potentiel adaptatif, pouvant mener à la disparition des populations. Cette hypothèse a été soutenue dans le genre *Silene* (Käfer et al., 2013). Toutefois, plus d'un faible taux de diversification, un faible taux de transition vers la dioécie (et un fort taux de perte) pourrait expliquer la faible diversité spécifique des taxons dioïques (Käfer et al., 2014, Käfer and Mousset, 2014).

## 1.8 Conclusion : le processus de spéciation est complexe

La spéciation est étroitement liée à l'histoire des populations. Pour avoir une vision d'ensemble de ce processus, il faut comprendre comment les populations ont divergé, dans quelle mesure la divergence est liée à des barrières à la reproduction et estimer si ces barrières à la reproduction sont suffisamment importantes pour maintenir les populations séparées et leur permettre d'évoluer en deux espèces distinctes. Cela demande donc une approche intégrée.

L'histoire des flux de gènes entre des populations a dans un premier temps été retracée par des approches phylogéographiques et de génétique des populations (Avise et al., 1987a,b). Les patrons de fréquences alléliques ont permis de proposer des hypothèses de divergence génétique qui peuvent alors être testés en les confrontant à d'autres scénarios. Les approches ABC (Approximate Bayesian Computation) se sont révélées particulièrement puissantes pour tester différents scénarios de flux de gènes au cours du temps (Duvaux et al., 2011, Roux et al., 2011), mais également le long du génome (Roux et al., 2013, 2014, Tine et al., 2014).

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



L'identification de régions associées à l'isolement reproducteur peut se faire par des approches QTL (Sweigart et al., 2006, Moyle and Nakazato, 2010), mais nécessite un grand nombre de descendants hybrides, ce qui n'est pas forcément réalisable lorsque l'isolement reproducteur postzygotique intrinsèque est important. Le scan génomique pourrait être utilisé (Turner et al., 2005, Gagnaire et al., 2013), mais son interprétation est étroitement liée à l'histoire des populations (Cruickshank and Hahn, 2014). Si des flux de gènes actuel sont détectés dans une zone hybride, un cline étroit de fréquence allélique d'un marqueur moléculaire peut être attribué à son implication dans l'isolement reproducteur (Teeter et al., 2008, Janoušek et al., 2012). Enfin, le système de reproduction peut impacter les forces évolutives impliquées dans la divergence des populations et des régions candidates impliquées dans l'isolement reproducteur peuvent être proposées en fonction du système de reproduction du modèle biologique (chromosomes sexuels, interactions nucléo-cytoplasmiques).

## 1.9 Le genre *Silene* comme modèle d'étude

Le genre *Silene* (Caryophyllaceae) est très étudié en biologie évolutive, notamment dans l'étude de l'évolution des systèmes de reproduction (Bernasconi et al., 2009). En effet, on y trouve des espèces hermaphrodites, gynodioïques et dioïques (Casimiro-Soriguer et al., 2015). Deux évolutions indépendantes de la dioécie ont été détectées (Figure 1.5; Desfeux et al., 1996, Marais et al., 2011), ce qui en fait un modèle d'étude de la transition de l'hermaphrodisme vers la dioécie par la gynodioécie (Dufay et al., 2014). Dans le cadre de cette thèse, nous allons surtout nous intéresser à *S. nutans*, une espèce gynodioique et *S. otites* et *S. pseudotites* qui sont dioïques.

### *Silene nutans*, une espèce gynodioïque

*Silene nutans* L. est une espèce pérenne à large distribution continentale (Figure 1.6a) que l'on retrouve dans des habitats xéro-thermophiles (pelouses, lisières forestières) sur affleurements rocheux ou sur sol sableux (Figure 1.6b). A la marge septentrionale et occidentale de sa distribution, c'est une espèce rare avec des populations isolées et souvent de petites tailles (Hepper, 1956, Van Rossum et al., 2003, Van Rossum and Prentice, 2004).

On retrouve des individus hermaphrodites et femelles dans les populations naturelles, ce qui est caractéristique d'un système gynodioïque. Il s'agit d'une gynodioécie cyto-nucléaire ; au moins deux types de cytoplasmes stérilisants ont été observés, qui ont des taux de restaurations


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



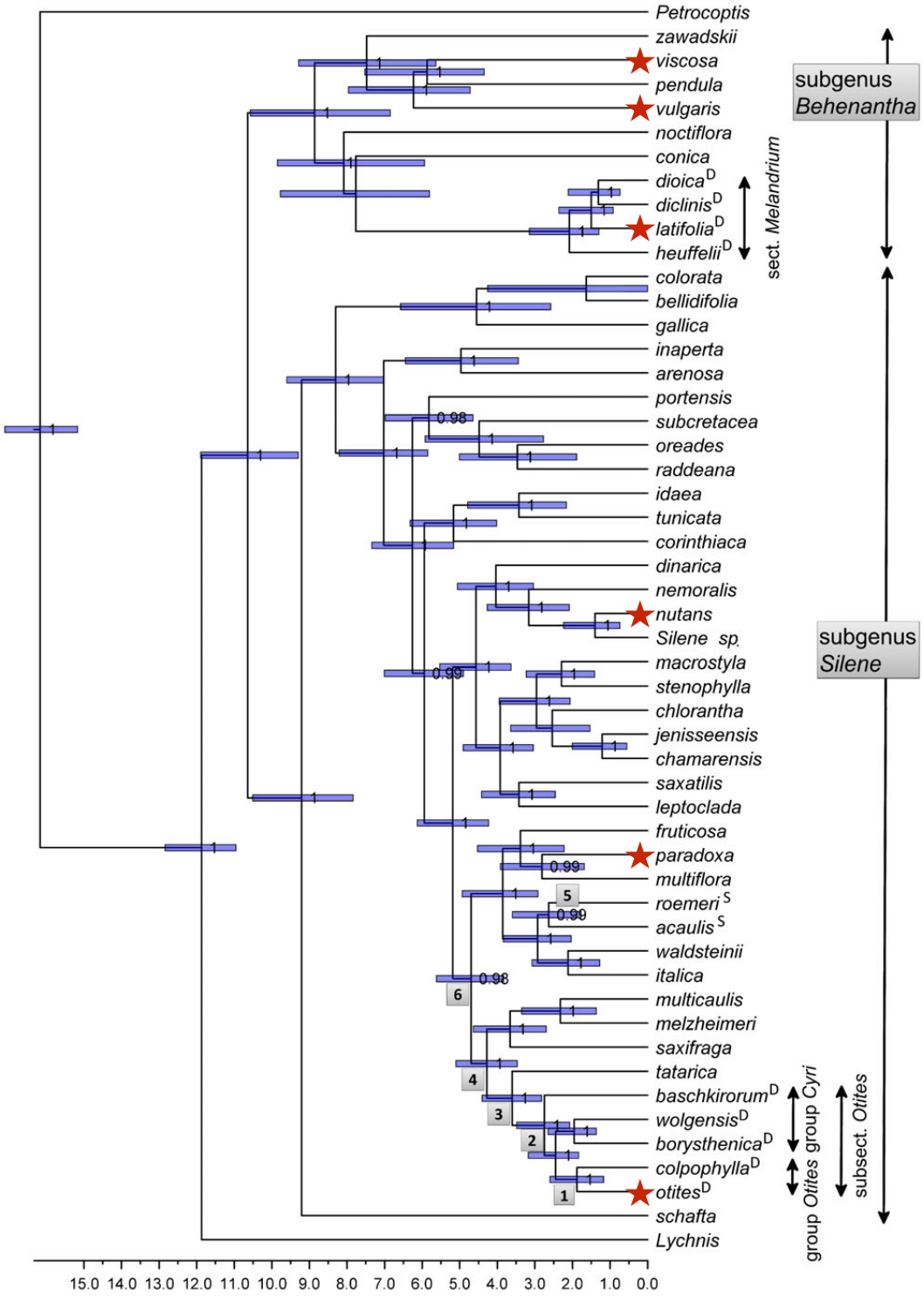



**Figure 1.5 :** Phylogénie (non exhaustive) du genre *Silene* d'après Slancarova et al. (2013). L'axe représente l'échelle de temps en millions d'années. Les étoiles rouges indiquent les espèces étudiées dans cette thèse.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



différents (Garraud et al., 2011). Le maintien de la gynodioécie chez *S. nutans* serait dû à de la sélection balancée fréquence-dépendante négative : la stérilité mâle cytoplasmique est seulement avantagée quand le restaurateur est rare et le restaurateur est avantagé quand la stérilité mâle cytoplasmique est fréquente (Gouyon et al., 1991). Il y aurait donc maintien d'un polymorphisme ancestral (Touzet and Delph, 2009, Lahiani et al., 2013).

Les individus hermaphrodites sont protandres (développement des étamines en premier, puis des stigmates), mais comme plusieurs fleurs peuvent être ouvertes sur un même individu à des stades différents, il peut y avoir autofécondation par géitonogamie (entre fleurs d'un même individu), s'accompagnant de dépression hybride (Hauser and Siegismund, 2000, Dufay et al., 2010). La pollinisation est principalement réalisée par des insectes nocturnes (par des papillons de nuit, en particulier des Noctuidae) et, à moindre échelle, par des insectes diurnes (par des abeilles sauvages à long proboscis et des syrphes; Jürgens et al., 1996, Lahiani, 2013).

Dans le sud de la Belgique, deux écotypes édaphiques (en fonction de la nature du sol) génétiquement et morphologiquement distincts ont été décrits (Figure 1.7; De Bilde, 1973, Van

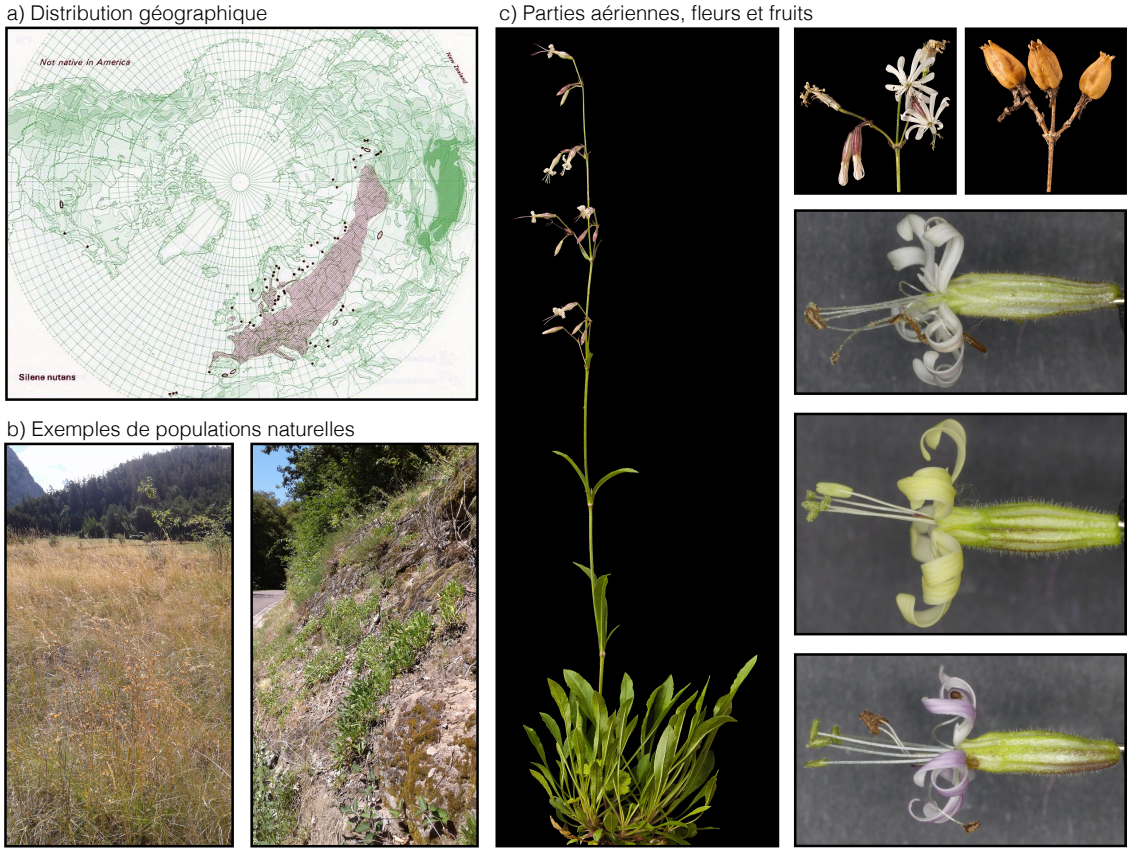



**Figure 1.6 :** a) Distribution géographique de *S. nutans* dans l'hémisphère nord (http ://linnaeus.nrm.se). b) Photos de populations naturelles, dans le Jura (Forêt de Serre, à gauche) et dans les Hautes Alpes (Siguret, à droite ; photos de Pascal Touzet). c) Photos de *S. nutans* en fleurs et en fruits (photos de Maarten Strack van Schijndel) et exemple de la diversité de couleur des pétales (photos de Daniel Parmentier).

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Rossum et al., 1997). L'écotype calcicole (Ca, sur sol à pH basique) comparé à l'écotype silicicole (Si, sur sol à pH acide à neutre) a des capsules plus petites et plus nombreuses et des feuilles spatulées (De Bilde, 1973, Van Rossum et al., 1999). Van Rossum et al. (1996) ont montré un fort isolement reproducteur pré- et postzygotique entre les deux écotypes : les plantules hybrides avaient en très large majorité un phénotype chlorotique, ce qui aboutissait à un for taux de mortalité juvénile.

En dehors de la Belgique, il existe également des variations morphologiques, mais aucune association entre la morphologie et le sol n'a été trouvée. En Angleterre par exemple, deux variétés morphologiques ont été décrites : *S. nutans* var. *salmoniana* (Hepper) et var. *smithiana* (Moss) (Hepper, 1951). La variété *salmoniana* a de grosses capsules et des feuilles fines (similaire à l'écotype belge Si) comparé à la variété *smithiana*, mais sans avoir de spécialisation édaphique (F. Van Rossum, données non publiées).

### *Silene otites* et *Silene pseudotites*, deux espèces dioïques

*Silene otites* et *S. pseudotites* sont deux taxons très proches, mais dont les relations phylogénétiques restent floues. Oxelman et al. (2013) considèrent que *S. pseudotites* est inclus dans l'espèce *S. otites* alors que la phylogénie présentée dans Desfeux et al. (1996) montre clairement une divergence entre les deux taxons. Enfin, Wrigley (1986) a suggéré que *S. pseudotites* est un hybride entre *S. otites* et *S. colpophylla* parce que *S. pseudotites* présente une morphologie intermédiaire entre ces deux espèces.

La répartition de *S. otites* couvre l'Europe de l'Ouest (de l'Espagne au Royaume-Uni) jusqu'en Europe centrale (de la Bulgarie à la Pologne), alors que *S. pseudotites* a une distribution

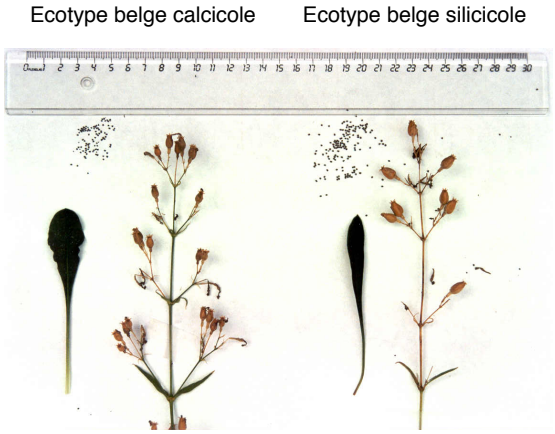



**Figure 1.7 :** Différences morphologiques entre les deux écotypes belges de *S. nutans* (photo de Fabienne Van Rossum).

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



restreinte au sud-est de la France et au nord de l'Italie (Tutin et al., 1964, Wrigley, 1986). Ces deux espèces présentent un syndrome de pollinisation anémophile (petites fleurs avec des petits pétales verdâtres et des organes sexuels à l'air libre ; Figure 1.8a), bien que de la pollinisation entomophile a été rapporté comme principal mode de pollinisation (Brantjes and Leemans, 1976). Globalement plus petit, *S. otites* a des feuilles lancéolées (fines et longues) comparé à *S. pseudotites*, plus grande, et qui présente des feuilles spatulées (plus larges) avec des bords ondulés (Figure 1.8b).

*Silene otites* et *S. pseudotites* appartiennent à la sous-section *Otites* qui ne comporte que des espèces dioïques (Figure 1.5) dont le déterminisme du sexe est encore très mal connu, notamment parce que les chromosomes sexuels, dans cette section, sont homomorphes (pas de différence de taille entre les chromosomes sexuels; Mrackova et al., 2008). En réalisant des croisements réciproques entre *S. otites* et *S. pseudotites*, Sansome (1938) a détecté un biais de sex-ratio en faveur des femelles lorsque *S. pseudotites* est le parent maternel et un sex-ratio équilibré lorsque le le parent maternel est *S. otites*. En plus d'un biais dans le sex-ratio, la descendance hybride présente des phénotypes chlorotiques. Etonnamment, pour expliquer le biais de sex-ratio dans les descendances hybrides, il ne fait pas appel à la règle d'Haldane, mais propose un déterminisme du sexe différent entre les deux espèces : un déterminisme de type XY pour *S. pseudotites* et ZW pour *S. otites*. Lors de croisements avec *S. pseudotites* comme parent maternel (XX) et *S. otites* comme parent paternel (ZZ), les descendants ont donc un unique génotype XZ, et donc un seul sexe serait observé, en l'occurrence femelle. Au contraire, lors de croisements avec *S. pseudotites* comme parent paternel (XY) et *S. otites* come parent maternel (ZW), les descendant ont donc les génotypes XZ ou XW ou YZ ou YW, et donc potentiellement un mélange d'individus mâles et femelles.

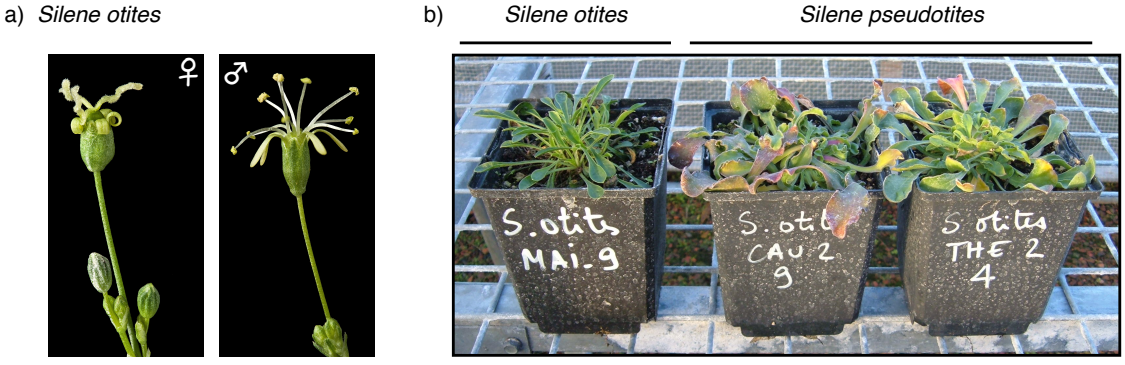



**Figure 1.8 :** a) Photo d'une fleur femelle et d'une fleur mâle de *S. otites* illustrant le syndrome de pollinisation anémophile (photos de Maarten Strack van Schijndel). b) Photo illustrant les différences morphologiques entre des rosettes d'un individu de *S. otites* et de deux individus de *S. pseudotites* provenant de trois populations différentes.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



## 1.10 Contenu de la thèse

Dans cette thèse, nous nous sommes intéressés au processus de spéciation et aux systèmes de reproduction dans le genre *Silene*. Dans une première partie, nous nous efforcerons de retracer l'histoire démographique de lignées génétiques chez *S. nutans*, une espèce gynodioïque (chapitres 1 à 3). Dans une seconde partie, nous nous intéresserons au cas d'espèces dioïques, en étudiant l'origine des chromosomes sexuels chez *S. otites* et *S. pseudotites* (chapitre 4). Enfin, en utilisant un échantillon d'espèces hermaphrodites, gynodioïques et dioïques, nous évaluerons l'effet du système de reproduction sur l'efficacité de la sélection (qu'elle soit négative ou positive), et testerons l'hypothèse de cul-de-sac évolutif de la dioécie (chapitre 5).

**Chapitre 1** Les écotypes belges de *S. nutans* sont génétiquement différenciés (Van Rossum et al., 1997). Dans ce chapitre, nous allons déterminer si cette différenciation génétique est le résultat d'une sélection disruptive liée à la nature du sol (contact primaire), ou d'une remise en contact de lignées qui ont divergé en allopatrie (contact secondaire). Nous utiliserons une approche phylogéographique et de génétique des populations avec un échantillonnage de populations couvrant cinq pays d'Europe occidentale et génotypé sur 13 marqueurs microsatellites et 6 SNP chloroplastiques pour savoir comment la structuration génétique en Belgique s'intègre dans une structuration génétique à plus large échelle géographique.

**Chapitre 2** L'isolement reproducteur entre les écotypes belges de *S. nutans* est caractérisé par une importante proportion d'individus hybrides chlorotiques dont la majorité meurt et les rares survivant sont stériles (Van Rossum et al., 1996). Dans ce chapitre, nous allons déterminer si cet isolement est associé à la sélection disruptive liée à la nature du sol (dans ce cas l'isolement reproducteur est limité aux écotypes belges), ou, si dans le cas d'un contact secondaire, il est lié à la divergence de lignées qui étaient en allopatrie (dans ce cas, l'isolement reproducteur n'est pas limité aux écotypes belges, mais peut être trouvé dans d'autres zones de contact). Nous avons eu à notre disposition une expérience de croisements contrôlés avec des populations provenant de zones de contacts entre lignées (identifiées grâce au chapitre 1) situées en Angleterre et en France, pour quantifier les barrières à la reproduction entre lignées et comparer les résultats avec ceux des écotypes belges.

**Chapitre 3** Si des flux de gènes ont (ou ont eu) lieu entre les lignées génétiques de *S. nutans*, un scan génomique pourrait nous permettre d'identifier des régions candidates impliquées dans l'isolement reproducteur. Dans ce chapitre, nous allons tester différents scénarios de divergence en tenant compte de la variation temporelle des flux de gènes, mais également de

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



la variation de leur intensité ainsi que de la sélection le long du génome. Nous utiliserons une approche ABC pour tester les différents scénarios et estimer les paramètres du meilleur scénario (par exemple le temps de divergence) afin d'interpréter au mieux le patron de différenciation génétique observé le long du transcriptome.

**Chapitre 4** Bien que les croisements entre *S. otites* et *S. pseudotites* suivent la règle d'Haldane, Sansome (1938) propose un déterminisme différent entre les deux taxons. Dans ce chapitre, nous allons tester son hypothèse. Nous utiliserons des données RNAseq d'une famille de *S. otites* et d'une famille de *S. pseudotites* pour tenter de déterminer la ségrégation des marqueurs et inférer le déterminisme du sexe de chacune des espèces.

**Chapitre 5** Dans ce chapitre, nous allons tester l'effet du système de reproduction sur l'efficacité de la sélection. Nous utilisons des données RNAseq de 8 espèces hermaphrodites, gynodioiques et dioiques, réparties dans deux sections du genre *Silene*, pour estimer les niveaux de polymorphisme et de divergence des espèces, en lien avec leur système de reproduction, afin de tester l'hypothèse de cul-de-sac évolutif de la dioécie.

## Bibliographie

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., and Saunders, N. C. (1987a). Intraspecific phylogeography : the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18 :489–522.

Avise, J. C., Reeb, C. A., and Saunders, N. C. (1987b). Geographic population structure and species differences in mitochondrial DNA of mouthbrooding marine catfishes (Ariidae) and demersal spawning toadfishes (Batrachoididae). *Evolution*, 41 :991–1002.

Babik, W., Branicki, W., Sandera, M., Litvinchuk, S., Borkin, L. J., Irwin, J. T., and Rafiński, J. (2004). Mitochondrial phylogeography of the moor frog, *Rana arvalis*. *Molecular Ecology*, 13 :1469–1480.

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J. C., and the tree of sex consortium (2014). Sex determination : why so many ways of doing it ? *PLoS biology*, 12 :e1001899.

Bank, C., Hermisson, J., and Kirkpatrick, M. (2011). Can reinforcement complete speciation ? *Evolution*, 66 :229–239.

Barbash, D. A. (2008). Clash of the genomes. *Cell*, 135 :1002–1003.

Barrett, S. C. H. (2002). The evolution of plant sexual diversity. *Nature Reviews Genetics*, 3 :274–284.

Barton, N. and Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity*, 56 :357–376.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Bernasconi, G., Antonovics, J., Biere, A., Charlesworth, D., Delph, L. F., Filatov, D., Giraud, T., Hood, M. E., Marais, G. A. B., McCauley, D., Pannell, J. R., Shykoff, J. A., Vyskot, B., Wolfe, L. M., and Widmer, A. (2009). *Silene* as a model system in ecology and evolution. *Heredity*, 103 :5–14.

Bierne, N., David, P., Boudry, P., and Bonhomme, F. (2002). Assortative fertilization and selection at larval stage in the mussels *Mytilus edulis* and *M. galloprovincialis*. *Evolution*, 56 :292–298.

Brantjes, N. B. M. and Leemans, J. A. A. M. (1976). *Silene otites* (Caryophyllaceae) pollinated by nocturnal lepidoptera and mosquitoes. *Acta Botanica Neerlandica*, 25 :281–295.

Brideau, N. J., Flores, H. A., Wang, J., Maheshwari, S., Wang, X., and Barbash, D. A. (2006). Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science*, 314 :1292–1295.

Brito, P. H. (2005). The influence of Pleistocene glacial refugia on tawny owl genetic diversity and phylogeography in western Europe. *Molecular Ecology*, 14 :3077–3094.

Britton-Davidian, J., Fel-Clair, F., Lopez, J., Alibert, P., and Boursot, P. (2005). Postzygotic isolation between the two European subspecies of the house mouse : estimates from fertility patterns in wild and laboratory-bred hybrids. *Biological Journal of the Linnean Society*, 84 :379–393.

Brothers, A. N. and Delph, L. F. (2010). Haldane's rule is extended to plants with sex chromosomes. *Evolution*, 64 :3643–3648.

Brys, R., van Cauwenberghe, J., and Jacquemyn, H. (2016). The importance of autonomous selfing in preventing hybridization in three closely related plant species. *Journal of Ecology*, 104 :601–610.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružic, M., Sæther, S. A., Sætre, G. P., Török, J., and Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, 25 :1656–1665.

Case, A. L., Finseth, F. R., Barr, C. M., and Fishman, L. (2016). Selfish evolution of cytonuclear hybrid incompatibility in *Mimulus*. *Proceedings of the Royal Society B*, 283 :20161493.

Casimiro-Soriguer, I., Buide, M. L., and Narbona, E. (2015). Diversity of sexual systems within different lineages of the genus *Silene*. *Annals of Botany Plants*, 7 :plv037.

Charlesworth, B. and Charlesworth, D. (1978). A model for the evolution of dioecy and gynodioecy. *The American Naturalist*, 112 :975–997.

Charlesworth, B. and Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London B : Biological Sciences*, 355 :1563–1572.

Charlesworth, B., Coyne, J. A., and Barton, N. H. (1987). The relative rates of evolution of sex chromosomes and autosomes. *The American Naturalist*, 130 :113–146.

Charlesworth, B., Morgan, M. T., and Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134 :1289–1303.

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



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Charlesworth, D., Charlesworth, B., and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, 95 :118–128.

Cheddadi, R., Vendramin, G. G., Litt, T., François, L., Kageyama, M., Lorentz, S., Laurent, J.-M., de Beaulieu, J.-L., Sadori, L., Jost, A., and Lunt, D. (2006). Imprints of glacial refugia in the modern genetic diversity of *Pinus sylvestris*. *Global Ecology and Biogeography*, 15 :271–282.

Cornille, A., Giraud, T., Bellard, C., Tellier, A., Le Cam, B., Smulders, M. J. M., Kleinschmit, J., Roldan-Ruiz, I., and Gladieux, P. (2013). Postglacial recolonization history of the European crabapple (*Malus sylvestris* Mill.), a wild contributor to the domesticated apple. *Molecular Ecology*, 22 :2249–2263.

Coyne, J. A. (1994). Ernst Mayr and the origin of species. *Evolution*, 48 :19–30.

Coyne, J. A. and Orr, H. A. (1997). "Patterns of speciation in *Drosophila*" revisited. *Evolution*, 51 :295–303.

Coyne, J. A. and Orr, H. A. (2004). *Speciation*. Sinauer associates, Sunderland, Massachusetts, USA, 1st edition.

Cruickshank, T. E. and Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23 :3133–3157.

De Bilde, J. (1973). Etude génécologique du *Silene nutans* L. en Belgique : populations du *Silene nutans* L. sur substrats siliceux et calcaires. *Revue Générale de Botanique*, 80 :161–176.

Delph, L. F. and Demuth, J. P. (2016). Haldane's rule : genetic bases and their empirical support. *Journal of Heredity*, page esw026.

Demuth, J. (2014). Do sex chromosomes affect speciation rate ? *Bioessays*, 36 :632.

Demuth, J. P., Flanagan, R. J., and Delph, L. F. (2014). Genetic architecture of isolation between two species of *Silene* with sex chromosomes and Haldane's rule. *Evolution*, 68 :332–342.

Desfeux, C., Maurice, S., Henry, J.-P., Lejeune, B., and Gouyon, P.-H. (1996). Evolution of reproductive systems in the genus *Silene*. *Proceedings of the Royal Society B : Biological Sciences*, 263 :409–414.

Dufay, M., Champelovier, P., Käfer, J., Henry, J.-P., Mousset, S., and Marais, G. A. B. (2014). An angiosperm-wide analysis of the gynodioecy-dioecy pathway. *Annals of Botany*, 114 :539–548.

Dufay, M., Cuguen, J., Arnaud, J. F., and Touzet, P. (2009). Sex ratio variation among gynodioecious populations of sea beet : can it be explained by negative frequency-dependent selection ? *Evolution*, 63 :1483–1497.

Dufay, M., Lahiani, E., and Brachi, B. (2010). Gender variation and inbreeding depression in gynodioecious-gynomonoecious *Silene nutans* (Caryophyllaceae). *International Journal of Plant Sciences*, 171 :53–62.

Dufay, M. and Pannell, J. R. (2010). The effect of pollen versus seed flow on the maintenance of nuclear-cytoplasmic gynodioecy. *Evolution*, 64 :772–784.

Duvaux, L., Belkhir, K., Boulesteix, M., and Boursot, P. (2011). Isolation and gene flow : inferring the speciation history of European house mice. *Molecular Ecology*, 20 :5248–5264.

Ellison, C. K., Niehuis, O., and Gadau, J. (2008). Hybrid breakdown and mitochondrial dysfunction in hybrids of *Nasonia* parasitoid wasps. *Journal of Evolutionary Biology*, 21 :1844–1851.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



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

Feder, J. L. and Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution*, 64 :1729–1747.

Felsenstein, J. (1981). Skepticism towards Santa Rosalia, or why are there so few kinds of animals ? *Evolution*, 35 :124–138.

Fishman, L. and Willis, J. H. (2006). A cytonuclear incompatibility causes anther sterility in *Mimulus* hybrids. *Evolution*, 60 :1372–1381.

Foxe, J. P., Slotte, T., Stahl, E. A., Neuffer, B., Hurka, H., and Wright, S. I. (2009). Recent speciation associated with the evolution of selfing in *Capsella*. *Proceedings of the National Academy of Sciences*, 106 :5241–5245.

Fraïsse, C., Elderfield, J. A. D., and Welch, J. J. (2014). The genetics of speciation : are complex incompatibilities easier to evolve ? *Journal of Evolutionary Biology*, 27 :688–699.

Gagnaire, P.-A., Normandeau, E., and Bernatchez, L. (2012). Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American Eels. *Molecular Biology and Evolution*, 29 :2909–2919.

Gagnaire, P.-A., Pavey, S. A., Normandeau, E., and Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by rad sequencing. *Evolution*, 67 :2483–2497.

Garraud, C., Brachi, B., Dufay, M., Touzet, P., and Shykoff, J. A. (2011). Genetic determination of male sterility in gynodioecious *Silene nutans*. *Heredity*, 106 :757–764.

Gavrilets, S. (2004). *Fitness landscapes and the origin of species*. Princeton University Press, Princeton, NJ.

Gossmann, T. I., Song, B. H., Windsor, A. J., Mitchell-Olds, T., Dixon, C. J., Kapralov, M. V., Filatov, D. A., and Eyre-Walker, A. (2010). Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Molecular Biology and Evolution*, 27 :1822–1832.

Gouyon, P.-H., Vichot, F., and Van Damme, J. M. M. (1991). Nuclear-cytoplasmic male sterility : single-point equilibria versus limit cycles. *The American Naturalist*, 137 :498–514.

Greiner, S., Rauwolf, U., Meurer, J., and Herrmann, R. G. (2011). The role of plastids in plant speciation. *Molecular Ecology*, 20 :671–691.

Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of genetics*, 12 :101–109.

Harrison, J. S. and Burton, R. S. (2006). Tracing hybrid incompatibilities to single amino acid substitutions. *Molecular Biology and Evolution*, 23 :559–564.

Hauser, T. P. and Siegismund, H. R. (2000). Inbreeding and outbreeding effects on pollen fitness and zygote survival in *Silene nutans* (Caryophyllaceae). *Journal of Evolutionary Biology*, 13 :446–454.

Heilbuth, J. C. (2000). Lower species richness in dioecious clades. *The American Naturalist*, 156 :221–241.

Heilbuth, J. C., Ilves, K. L., and Otto, S. P. (2001). The consequences of dioecy for seed dispersal : modeling the seed-shadow handicap. *Evolution*, 55 :880–888.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Hepper, F. N. (1951). The variations of *Silene nutans* L. in Great Britain. *Watsonia*, 2 :80–90.

Hepper, F. N. (1956). Biological flora of the British Isles : *Silene nutans* L. *Journal of Ecology*, 44 :693–700.

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

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

Hewitt, G. M. (2011). Quaternary phylogeography : the roots of hybrid zones. *Genetica*, 139 :617–638.

Hopkins, R. (2013). Reinforcement in plants. *New Phytologist*, 197 :1095–1103.

Jaarola, M. and Searle, J. B. (2002). Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology*, 11 :2613–2621.

Janoušek, V., Wang, L., Luzynski, K., Dufková, P., Vyskočilová, M. M., Nachman, M. W., Munclinger, P., Macholán, M., Piálek, J., and Tucker, P. K. (2012). Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. *Molecular Ecology*, 21 :3032–3047.

Jewell, C., Papineau, A. D., Freyre, R., and Moyle, L. C. (2012). Patterns of reproductive isolation in *Nolana* (chilean bellflower). *Evolution*, 66 :2628–2636.

Josephs, E. B. and Wright, S. I. (2016). On the trail of linked selection. *PLOS Genetics*, 12 :e1006240.

Jürgens, A., Witt, T., and Gottsberger, G. (1996). Reproduction and pollination in central European populations of *Silene* and *Saponaria* species. *Botanica Acta*, 109 :316–324.

Käfer, J., de Boer, H. J., Mousset, S., Kool, A., Dufay, M., and Marais, G. A. B. (2014). Dioecy is associated with higher diversification rates in flowering plants. *Journal of Evolutionary Biology*, 27 :1478–1490.

Käfer, J. and Mousset, S. (2014). Standard sister clade comparison fails when testing derived character states. *Systematic Biology*, 63 :601–609.

Käfer, J., Talianová, M., Bigot, T., Michu, E., Guéguen, L., Widmer, A., Žlůvová, J., Glémin, S., and Marais, G. A. B. (2013). Patterns of molecular evolution in dioecious and non-dioecious *Silene*. *Journal of Evolutionary Biology*, 26 :335–346.

Kaplan, N. L., Hudson, R. R., and Langley, C. H. (1989). The 'hitchhiking effect' revisited. *Genetics*, 123 :887–899.

Kimura, M. (1980). Average time until fixation of a mutant allele in a finite population under continued mutation pressure : studies by analytical, numerical, and pseudo-sampling methods. *Proceedings of the National Academy of Sciences*, 77 :522–526.

Kimura, M. and Ohta, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61 :763–771.

Kitano, J. and Peichel, C. L. (2012). Turnover of sex chromosomes and speciation in fishes. *Environmental Biology of Fishes*, 94 :549–558.

Kitano, J., Ross, J. A., Mori, S., Kume, M., Jones, F. C., Chan, Y. F., Absher, D. M., Grimwood, J., Schmutz, J., Myers, R. M., Kingsley, D. M., and Peichel, C. L. (2009). A role for a neo-sex chromosome in stickleback speciation. *Nature*, 461 :1079–1083.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Lahiani, E. (2013). *Dynamique évolutive de la gynodioécie chez* Silene nutans *et conditions de son maintien en populations*. Phd thesis, Université de Lille 1, France.

Lahiani, E., Dufaÿ, M., Castric, V., Le Cadre, S., Charlesworth, D., Van Rossum, F., and Touzet, P. (2013). Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity*, 111 :157–164.

Levin, D. A. and Scarpino, S. V. (2016). On the young age of intraspecific herbaceous taxa. *New Phytologist*, doi : 10.1111/nph.14224.

Leys, M., Petit, E. J., El-Bahloul, Y., Liso, C., Fournet, S., and Arnaud, J.-F. (2014). Spatial genetic structure in *Beta vulgaris* subsp. *maritima* and *Beta macrocarpa* reveals the effect of contrasting mating system, influence of marine currents, and footprints of postglacial recolonization routes. *Ecology and Evolution*, 4 :1828–1852.

Li, W.-H. and Nei, M. (1977). Persistence of common alleles in two related populations or species. *Genetics*, 86 :901–914.

Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A., and Willis, J. H. (2008a). The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society of London B : Biological Sciences*, 363 :3009–3021.

Lowry, D. B., Rockwood, R. C., and Willis, J. H. (2008b). Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution*, 62 :2196–2214.

Lu, G. and Bernatchez, L. (1998). Experimental evidence for reduced hybrid viability between dwarf and normal ecotypes of lake whitefish (*Coregonus clupeaformis* Mitchill). *Proceedings of the Royal Society B : Biological Sciences*, 265 :1025–1030.

Luca, A., Palermo, A. M., Bellusci, F., and Pellegrino, G. (2015). Pollen competition between two sympatric *Orchis* species (Orchidaceae) : the overtaking of conspecific of heterospecific pollen as a reproductive barrier. *Plant Biology*, 17 :219–225.

Marais, G. A. B., Forrest, A., Kamau, E., Käfer, J., Daubin, V., and Charlesworth, D. (2011). Multiple nuclear gene phylogenetic analysis of the evolution of dioecy and sex chromosomes in the genus *Silene*. *PloS One*, 6 :e21915.

Martin, N. H. and Willis, J. H. (2007). Ecological divergence associated with mating system causes nearly complete reproductive isolation between sympatric *Mimulus* species. *Evolution*, 61 :68–82.

Maurice, S., Belhassen, E., Couvet, D., and Gouyon, P.-H. (1994). Evolution of dioecy : can nuclear-cytoplasmic interactions select for maleness ? *Heredity*, 73 :346–354.

Mayr, E. and Ashlock, P. D. (1991). *Principles of systematic zoology*. 2nd edition.

Meisel, R. P. and Connallon, T. (2013). The faster-X effect : integrating theory and data. *Trends in Genetics*, 29 :537–544.

Moyle, L. C. and Nakazato, T. (2010). Hybrid incompatibility "snowballs" between *Solanum* species. *Science*, 329 :1521–1523.

Moyle, L. C., Olson, M. S., and Tiffin, P. (2004). Patterns of reproductive isolation in three angiosperm genera. *Evolution*, 58 :1195–1208.

Mrackova, M., Nicolas, M., Hobza, R., Negrutiu, I., Monéger, F., Widmer, A., Vyskot, B., and Janousek, B. (2008). Independent origin of sex chromosomes in two species of the genus *Silene*. *Genetics*, 179 :1129–1133.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., Quail, M. A., Joron, M., Ffrench-Constant, R. H., Blaxter, M. L., Mallet, J., and Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 367 :343–353.

Nei, M. and Nozawa, M. (2011). Roles of mutation and selection in speciation : from Hugo de Vries to the modern genomic era. *Genome Biology and Evolution*, 3 :812–829.

Nista, P., Brothers, A. N., and Delph, L. F. (2015). Differences in style length confer prezygotic isolation between two dioecious species of *Silene* in sympatry. *Ecology and Evolution*, 5 :2703–2711.

Noor, M. A. F. (1999). Reinforcement and other consequences of sympatry. *Heredity*, 83 :503–508.

Noor, M. A. F. and Bennett, S. M. (2009). Islands of speciation or mirages in the desert ? Examining the role of restricted recombination in maintaining species. *Heredity*, 103 :439–444.

Orr, H. A. (1995). Population genetics of speciation : the evolution of hybrid incompatibilities. *Genetics*, 139 :1805–1813.

Orr, H. A. and Presgraves, D. C. (2000). Speciation by postzygotic isolation : forces, genes and molecules. *BioEssays*, 22 :1085–1094.

Oxelman, B., Rautenberg, A., Thollesson, M., Larsson, A., Frajman, B., Eggens, F., Petri, A., Aydin, Z., Töpel, M., and Brandtberg-Falkman, A. (2013). *Sileneae* taxonomy and systematics. *http ://www.sileneae.info*.

Phadnis, N. and Orr, H. A. (2009). A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science*, 323 :376–379.

Prentice, H. C., Lönn, M., Lefkovitch, L. P., and Runyeon, H. (1995). Associations between allele frequencies in *Festuca ovina* and habitat variation in the alvar grasslands on the Baltic island of Oland. *Journal of Ecology*, 83 :391–402.

Presgraves, D. C. (2003). A fine-scale genetic analysis of hybrid incompatibilities in *Drosophila*. *Genetics*, 163 :955–972.

Presgraves, D. C. (2010). The molecular evolutionary basis of species formation. *Nature Reviews Genetics*, 11 :175–180.

Presgraves, D. C., Balagopalan, L., Abmayr, S. M., and Orr, H. A. (2003). Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature*, 423 :715–719.

Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., Bowers, J. E., Burke, J. M., and Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature communications*, 4 :1827.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, 16 :351–358.

Rieseberg, L. H., Whitton, J., and Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, 152 :713–727.

Roux, C., Castric, V., Pauwels, M., Wright, S. I., Saumitou-Laprade, P., and Vekemans, X. (2011). Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation ? *PLoS One*, 6 :e26872.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Roux, C., Fraïsse, C., Castric, V., Vekemans, X., Pogson, G. H., and Bierne, N. (2014). Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation ? A case study in a *Mytilus* hybrid zone. *Journal of Evolutionary Biology*, 27 :1662–1675.

Roux, C., Tsagkogeorga, G., Bierne, N., and Galtier, N. (2013). Crossing the species barrier : genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, 30 :1574–1587.

Sansome, F. W. (1938). Sex determination in *Silene otites* and related sepecies. *Journal of Genetics*, 35 :387–396.

Schilthuizen, M., Giesbers, M. C. W. G., and Beukeboom, L. W. (2011). Haldane's rule in the 21st century. *Heredity*, 107 :95–102.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, A., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., Lindholm, A. K., Lucek, K., Maan, M. E., Marques, D. A., Martin, S. H., Matthews, B., Meier, J. I., Möst, M., Nachman, M. W., Nonaka, E., Rennison, D. J., Schwarzer, J., Watson, E. T., Westram, A. M., and Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15 :176–192.

Selz, O. M., Pierotti, M. E. R., Maan, M. E., Schmid, C., and Seehausen, O. (2014). Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. *Behavioral Ecology*, 25 :612–626.

Sicard, A. and Lenhard, M. (2011). The selfing syndrome : a model for studying the genetic and evolutionary basis of morphological adaptation in plants. *Annals of Botany*, 107 :1433–1443.

Skrede, I., Brochmann, C., Borgen, L., and Rieseberg, L. H. (2008). Genetics of intrinsic postzygotic isolation in a circumpolar plant species, *Draba nivalis* (Brassicaceae). *Evolution*, 62 :1840–1851.

Slancarova, V., Zdanska, J., Janousek, B., Talianova, M., Zschach, C., Zluvova, J., Siroky, J., Kovacova, V., Blavet, H., Danihelka, J., Oxelman, B., Widmer, A., and Vyskot, B. (2013). Evolution of sex determination systems with heterogametic males and females in *Silene*. *Evolution*, 67 :3669–3677.

Sweigart, A. L., Fishman, L., and Willis, J. H. (2006). A simple genetic incompatibility causes hybrid male sterility in *Mimulus*. *Genetics*, 172 :2465–2479.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G., and Cosson, J.-F. (1998). Comparative phylogeography and postglacial colonization. *Molecular Ecology*, 7 :453–464.

Tang, S. and Presgraves, D. C. (2009). Evolution of the *Drosophila* nuclear pore comples results in multiple hybrid incompatibilities. *Science*, 323 :779–782.

Tao, Y., Araripe, L., Kingan, S. B., Ke, Y., Xiao, H., and Hartl, D. L. (2007). A sex-ratio meiotic drive system in *Drosophila simulans*. II : an X-linked distorter. *PLoS Biology*, 5 :2576–2588.

Teeter, K. C., Payseur, B. A., Harris, L. W., Bakewell, M. A., Thibodeau, L. M., O'Brien, J. E., Krenz, J. G., Sans-Fuentes, M. A., and Nachman, M. W. (2008). Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Research*, 18 :67–76.

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert,



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



F. A. M., Bargelloni, L., Power, D. M., Bonhomme, F., Canario, A. V. M., and Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, 5 :5770.

Tong, Z.-Y. and Huang, S.-Q. (2016). Pre- and post-pollination interaction between six co-flowering *Pedicularis* species via heterospecific pollen transfer. *New Phytologist*, 211 :1452–1461.

Touzet, P. and Delph, L. F. (2009). The effect of breeding system on polymorphism in mitochondrial genes of *Silene*. *Genetics*, 181 :631–644.

Turelli, M. and Moyle, L. C. (2007). Asymmetric postmating isolation : Darwin's corollary to Haldane's rule. *Genetics*, 176 :1059–1088.

Turesson, G. (1922). The species and the variety as ecological units. *Hereditas*, 3 :100–113.

Turner, T. L., Hahn, M. W., and Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, 3 :1572–1578.

Tutin, T., Heywood, V., Burges, N., Valentine, D., Walters, S., and Webb, D., editors (1964). *Flora europaea*. Cambridge University Press, Cambridge.

Vamosi, J. C. and Otto, S. P. (2002). When looks can kill : the evolution of sexually dimorphic floral display and the extinction of dioecious plants. *Proceedings of the Royal Society B : Biological Sciences*, 269 :1187–1194.

Van Rossum, F., De Bilde, J., and Lefèbvre, C. (1996). Barriers to hybridization in calcicolous and silicolous populations of *Silene nutans* from Belgium. *Belgian Journal of Botany*, 129 :13–18.

Van Rossum, F., Meerts, P., Gratia, E., and Tanghe, M. (1999). Ecological amplitude in *Silene nutans* in relation to allozyme variation at the western margin of its distribution. *Journal of Vegetation Science*, 10 :253–260.

Van Rossum, F. and Prentice, H. C. (2004). Structure of allozyme variation in Nordic *Silene nutans* (Caryophyllaceae) : population size, geographical position and immigration history. *Biological Journal of the Linnean Society*, 81 :357–371.

Van Rossum, F., Vekemans, X., Gratia, E., and Meerts, P. (2003). A comparative study of allozyme variation of peripheral and central populations of *Silene nutans* L. (Caryophyllaceae) from Western Europe : implications for conservation. *Plant Systematics and Evolution*, 242 :49–61.

Van Rossum, F., Vekemans, X., Meerts, P., Gratia, E., and Lefèbvre, C. (1997). Allozyme variation in relation to ecotypic differentiation and population size in marginal populations of *Silene nutans*. *Heredity*, 78 :552–560.

Waters, J. M., Fraser, C. I., and Hewitt, G. M. (2013). Founder takes all : density-dependent processes structure biodiversity. *Trends in Ecology and Evolution*, 28 :78–85.

Werren, J. H. (2011). Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proceedings of the National Academy of Sciences*, 108 :10863–10870.

Wilkinson, G. S., Christianson, S. J., Brand, C. L., Ru, G., and Shell, W. (2014). Haldane's rule is linked to extraordinary sex ratios and sperm length in stalk-eyed flies. *Genetics*, 198 :1167–1181.

Wright, S. I., Kalisz, S., and Slotte, T. (2013). Evolutionary consequences of self-fertilization in plants. *Proceedings of the Royal Society B : Biological Sciences*, 280 :20130133.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Wrigley, F. (1986). Taxonomy and chorology of *Silene* section *Otites* (Caryophyllaceae). *Annales Botanici Fennici*, 23 :69–81.

Yoshida, K., Makino, T., Yamaguchi, K., Shigenobu, S., Hasebe, M., Kawata, M., Kume, M., Mori, S., Peichel, C. L., Toyoda, A., Fujiyama, A., and Kitano, J. (2014). Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLoS Genetics*, 10 :e1004223.

© 2016 Tous droits réservés.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

# **Chapitre 1** : Phylogeographic pattern of *Silene nutans*

**Hélène Martin, Pascal Touzet, Fabienne Van Rossum, Damien Delalande et Jean-François Arnaud**

Ce chapitre a été publié dans *Heredity* [1]

1. Martin H, Touzet P, Van Rossum F, Delalande D, Arnaud J-F (2016) Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity*, 116, 286–294.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



L'objectif de ce chapitre était de voir comment la structuration génétique de *Silene nutans* en Belgique s'intègre dans une structuration génétique à plus large échelle géographique. Nous avons utilisé une approche phylogéographique et de génétique des populations avec un échantillonnage de 111 populations génotypé sur 13 marqueurs microsatellites et 6 SNP chloroplastiques. Sur la zone d'étude, nous avons décrit deux lignées nucléo-cytoplasmiques géographiquement structurées : une lignée *Ouest* sous-structuré en trois groupes génétiques et une lignée *Est*. Ces lignées présentent des patrons distincts de recolonisation post-glaciaire. Elles sont actuellement en contact en Angleterre, en Belgique ou en France, mais aucun flux de gènes n'a été détecté dans ces trois régions. Nous avons donc conclu que la différenciation génétique des écotypes belges était le résultat d'un contact secondaire entre deux lignées génétiques.

Les donnée génétiques utilisées dans ce chapitre ont été générées lors du stage de M2 de Damien Delalande (2012) et du mien (2013) avant d'être complétées pendant cette thèse. J'ai réalisé toutes les analyses et écrit l'article en incorporant les corrections et remarques des co-auteurs.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



# Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe

## 2.1 Introduction

Most plant species in Europe have been extensively inventoried, described and documented. However, some of them show high intraspecific genetic, phenotypic and/or ecological variation leading to among-population differentiation that might ultimately result in speciation (Coyne and Orr, 2004, Gavin et al., 2014). We may wonder whether one single species might consist of several Evolutionary Significant Units (ESUs; Fraser and Bernatchez, 2001) evolving in relative isolation from each other, or even corresponding to cryptic species, as a result of recent or past evolutionary processes. Identifying these ESUs has implications in terms of sustainable plant diversity conservation. Depending on their distribution (wide or restricted), abundance (common or rare), habitat requirements (more or less specialized), and patterns of genetic diversity and structure, the conservation status of ESUs may differ, and in situ and ex situ conservation strategies should be evaluated for each ESU separately (Palsbøll et al., 2007). Besides, plant communities are facing strong and fast environmental changes due to human activities (e.g. climate change, habitat fragmentation), threatening their survival on the long term (Hautekèete et al., 2015). Recent studies of the effects of contemporary climate warming on plant species distribution and migration have emphasized the need for a better understanding of species response to past climatic changes (Petit et al., 2008).

The study of population genetic structure and phylogeographic patterns across a species range, using a combination of plastid and nuclear molecular markers, allows retracing past and present species spread and identifying past and recent evolutionary processes responsible for the current patterns of genetic variation (Avise, 2000, Gavin et al., 2014). Past climate events, especially the series of major ice ages that occurred during the Quaternary (2.4 million years ago (Mya) to the present), have driven the evolution of divergent genetic lineages by isolating populations in distant refugia, for instance in the Iberian peninsula, Italy or the Balkans in Europe (Hewitt, 2000). Ultimately, these historical refugia may have promoted allopatric genetic divergence leading to different ESUs or to sibling species that may not be morphologically distinguishable, i.e. cryptic species (Hewitt, 2000, Gavin et al., 2014). Subsequently, postglacial

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



migration and species range expansion from these different refugia have allowed divergent lineages to come into secondary contact in suture zones (e.g. Taberlet et al., 1998, Hewitt, 2011). Contemporary processes, such as gene flow, genetic drift and spatially heterogeneous selective pressures, in combination with species life-history traits, in particular the breeding system and seed and pollen dispersal capabilities, can also contribute to shape species' distribution, genetic diversity and population differentiation (Hamrick and Godt, 1996, Duminil et al., 2007).

In this paper, we examine the genetic structure of the perennial herb *Silene nutans* in Western Europe. In Great Britain, northern France and Belgium, this species exhibits a disjunct distribution, is locally rare and vulnerable, and populations show allozyme and/or morphological divergence (e.g. Hepper, 1956, Fitter, 1978, Van Rossum et al., 1997, 2003). In Great Britain, two varieties, *S. nutans* var. *salmoniana* Hepper and var. *smithiana* Moss, have been differentiated based on reproductive traits (Hepper, 1951) but this distinction is no longer considered in recent taxonomy (e.g. Stace, 2010). In Belgium, two parapatric edaphic ecotypes (calcicolous *Ca* and silicicolous *Si*), with contrasting morphological traits and substantial genetic divergence have been described. Isolating reproductive barriers have been identified between them, suggesting incipient speciation processes (De Bilde, 1973, Van Rossum et al., 1996, 1997). Our hypothesis is that populations of *S. nutans* in Western Europe may correspond to several ESUs or to cryptic biological species.

Using plastid SNPs and nuclear microsatellite markers, we investigated the large-scale phylogeographic and population genetic structure patterns on a comprehensive sampling of populations of *S. nutans* from Western Europe. Using *F*-statistics, Bayesian clustering, and spatial multivariate methods, we addressed the following questions:

1. What are the range-wide levels of genetic diversity and genetic differentiation in *S. nutans*? Can the patterns be related to historical gene flow, mating system and/or seed and pollen dispersal abilities?

2. If spatial genetic analyses reveal genetically differentiated lineages in *S. nutans* in Western Europe, what is the most plausible evolutionary process involved in the formation of these distinct lineages? We hypothesized that range shifts driven by the last Quaternary glaciations and leading-edge expansions associated with founding events may be the main processes explaining the large-scale patterns of genetic variability we observed.

3. Is there evidence for suture zones among genetically divergent lineages, and if so, can we detect hybridisation events mediated by seed and/or pollen flow in the contact zones?


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Our results will be discussed in the light of biogeographical scenarios, climate refugia and past historical migration events since the Last Glacial Maxima, leading to divergent ESUs and possible cryptic speciation.

## 2.2 Material and Methods

### 2.2.1 Studied species and plant material

*Silene nutans* L. (Caryophyllaceae) is a diploid, herbaceous, long-lived perennial rosette-forming plant species, occurring in dry habitats, on rock outcrops, sand or shingle, such as xerothermophilous grasslands, open forests and forest edges. Its wide continental distribution range extends from western Europe to central Siberia and South Caucasus (Hepper, 1956, Fitter, 1978). Flowers are protandrous and insect-pollinated, mainly by moths (Jürgens et al., 1996). *S. nutans* is self-compatible and expected to exhibit a mixed-mating system. Inbreeding depression in selfed progeny has been reported (Dufay et al., 2010).

Leaf material was sampled from a total of 1979 individuals from 111 populations located at the western border of *S. nutans* geographic range (Table S2.1 in the appendices, Figure 2.1.a). The sampling covered six countries, including 16 populations from the UK (populations 1 to 11 and 16 for *S. nutans* var. *smithiana*, and populations 12 to 15 for var. *salmoniana*; population assignment to varieties was based on fruit measurements; Van Rossum, unpubl. data), 50 populations originating from France, 10 populations from Luxemburg, 13 populations from Belgium (for which five have been identified as belonging to the *Ca* ecotype labelled 77, 78, 80, 87, and 88, and five to the *Si* ecotype labelled 79, 81, 82, 86, and 89; population assignment to ecotypes was based on De Bilde, 1973, Van Rossum et al., 1997), three populations from The Netherlands, and 19 populations from Germany. Leaf samples were dried and conserved in silica gel until DNA extraction. Sample sizes ranged from 2 to 133, with a mean of 17.8 individuals per population ($\pm$14.8).

### 2.2.2 Molecular analyses

All sampled individuals were genotyped using both plastid and nuclear DNA markers. DNA was extracted from 15-20 mg of leaf tissue using Macherey-Nagel (Düren, Germany) NucleoSpin ® 96 Plant II kits following the standard protocol outlined in the manufacturer's handbook. Multilocus nuclear DNA (nDNA) genotypes were characterised at 13 recently isolated microsatellite loci named B09, E08, G01, H07, D10, Sil16, Sil19, Sil24, Sil31, Sil35, Sil36, Sil37 and Sil42 and described

© 2016 Tous droits réservés.



lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



in Godé et al. (2014). Amplification procedures, multiplexing and genotyping were carried out following the standard protocols found in Godé et al. (2014).

Plastid diversity was investigated using SNPs recently developed from plastid sequences (pDNA) of *S. nutans* representative samples (Lahiani et al., 2013). Plastid SNPs were defined based on six polymorphic pDNA nucleotides in the intergenic spacer sequences psbA-trnH, named Cp 42, and the matK gene fragment, named Cp 397, Cp 540, Cp 656, Cp 730 and Cp 804. The Kompetitive Allele Specific PCR genotyping assays (KASP$^{tm}$) was used to detect the SNPs. This method is based on competitive allele-specific PCR and enables bi-allelic scoring of single nucleotide polymorphisms. The SNP-specific KASP Assay mix (designed by LGC group, www.lgcgroup.com) and the universal KASP Master mix are added to DNA samples, a thermal cycling reaction is then performed, followed by an end-point fluorescent read. Bi-allelic discrimination is achieved through the competitive binding of two allele-specific forward primers, each with a unique tail sequence that corresponds with two universal FRET (fluorescence resonant energy transfer) cassettes; one labelled with FAM$^{tm}$ dye and the other with HEX$^{tm}$ dye. The sequences flanking the SNPs that were used to define the primers are given in Table S2.2 in the appendices. Amplification reactions were carried out by PCR with 30 ng of DNA in 8.11 µl reaction volume containing 4 µl of universal KASP Master mix (2x) and 0.11 µl of the SNP-specific KASP Assay mix (composed of the two forward primers and the reverse primer). Cycling conditions for PCR amplification were 15 min at 94℃, 10 cycles of 20 s at 94℃ and 60 s at 65℃ with a step-down of 0.8℃ per cycle, 26 cycles of 20 s at 94℃ and 60 s at 57℃, and finally 12℃ for 10 min. Fluorescence was detected using a LightCycler® 480 (Roche Diagnostics, Basel, Switzerland). Individual haplotypes were defined as a combination of allelic states for all six SNPs.

### 2.2.3 Nuclear genetic variation within populations

Prior to further analyses, genotypic linkage disequilibrium (LD) among all pairwise locus combinations was checked with a log-likelihood ratio test, implemented in FSTAT version 2.9.3.2. (Goudet, 1995), using 10 000 randomisations of two-locus genotypes and Bonferroni correction. The following measures of genetic variation were calculated within and over all populations: allelic richness (*Ar*, El Mousadik and Petit, 1996) based on a minimum sample size of eight individuals (six populations with sample size below this threshold were not included, see Table S2.1 in the appendices), observed heterozygosity ($H_O$), and expected heterozygosity ($H_E$) corrected for sample size. Some populations failed to be genotyped for the B09 locus (populations 59, 102 and 103) and the D10 locus (population 21); allelic richness was thus only esti-


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



mated for the 11 remaining loci. Using GENEPOP version 4.3 (Rousset, 2008), deviations from Hardy-Weinberg equilibrium were also tested by estimating the intra-population fixation index ($F_{IS}$) following Weir and Cockerham (1984). Significance of Fis estimates was tested using exact probability tests across loci and populations. Markov chain method provided unbiased estimates of the Fisher's exact test probability using the following parameters: 100 000 dememorizations, 1 000 batches, 100 000 iterations per batch. To obtain corrected $F_{IS}$ estimates ($F_{IS\_corr}$) accounting for null allele occurrence, we used the Bayesian individual inbreeding model described in Chybicki and Burczyk (2009) and implemented in the software INEST version 2.0 available at http://genetyka.ukw.edu.pl. Using SPAGeDi V1.5 (Hardy and Vekemans, 2002), indirect and independent estimates of selfing rates ($s$) were derived from the multilocus correlation structure of population samples, following a multilocus estimate of the standardized identity disequilibrium and described in David et al. (2007). To examine geographic trends in the distribution of large-scale genetic diversity, we tested the relationship between allelic richness, longitude and latitude using a linear multiple regression model. Spatially interpolated values of allelic richness were then generated for a better visualisation using the thin plate spline method implemented in R version 3.0.2 (R Core Team Development, 2014) and represented on a geographical heat map.

### 2.2.4 Lineage distinctiveness and spatial patterns of genetic structure

Three complementary approaches were used on nuclear microsatellite data to investigate the genetic affinities among populations. Firstly, genetic distance-based population clustering was performed to infer the genetic relationships among *S. nutans* populations using an unrooted neighbour-joining (NJ) tree. It was constructed using Cavalli-Sforza and Edwards (1967) genetic chord distance (DCE) based on allele frequencies. Bootstrapped confidence values on obtained branches were determined using random resampling replications over loci (1000 replicates) implemented in POPULATIONS software 1.2.32.

Secondly, a non-spatially explicit Bayesian clustering was carried out to infer the number of distinct gene pools in the whole data set, using the model-based Bayesian algorithm implemented in STRUCTURE software (Pritchard et al., 2000). The number of potential K clusters was assessed from 10 different runs of K ranging from 1 to 80. To ascertain adequate convergence of the Markov Chain Monte Carlo (MCMC) model, we allowed a burn-in of 100 000 iterations, followed by $2.10^6$ MCMC replications without any prior geographic information on the putative affiliation of individuals. To identify the most likely number of $K$ clusters, the ad hoc statistic $\Delta K$ was calculated as described in Evanno et al. (2005). CLUMPP version 1.1.2. (Jakobsson and

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



Rosenberg, 2007) accounted for label switching and was subsequently used to find the optimal alignment of independent runs by averaging the top runs. The resulting most likely grouping of populations was plotted using DISTRUCT version 1.1. (Rosenberg, 2004). This method was applied (i) to the whole data set to identify potential lineages, (ii) to each identified lineage separately to investigate more subtle lineage specific genetic structure, and (iii) to populations from areas where different haplotypes (potential distinct lineages) co-occurred (potential contact zones), to detect potential admixture events. This kind of model-based Bayesian algorithm infers subtle population structure by minimising linkage disequilibrium and departures from Hardy-Weinberg equilibrium within each inferred cluster (Pritchard et al., 2000). However, the assumption of panmixia does not necessarily hold in *S. nutans* which can exhibit a mixed-mating system in natural populations (Dufay et al., 2010).

Thirdly, to avoid this potential bias, we performed a multivariate ordination analysis that does not require any genetic assumptions: a spatial principal component analysis (sPCA, reviewed in Jombart et al., 2009). We calculated independent synthetic variables that maximise the product of the genetic variance among populations and their spatial autocorrelation (based on Moran's *I*). The computation of Moran's *I* was done based on a Delaunay graph. Genetically distinguishable groups, clines in allele frequency and intermediate states can lead to what is called "global structure" and are identified by positive components, i.e. showing positive spatial autocorrelations (e.g. DiLeo et al., 2010). sPCA and tests for statistical significance of global structure were performed using the R ADEGENET package (Jombart, 2008). The main results of sPCA can be depicted on maps of population scores, allowing a visual assessment of spatial genetic structuring (Jombart et al., 2008). Using the same strategy as that used for Bayesian clustering, we constructed maps for (i) the whole data set to investigate potential genetically distinct lineages, and (ii) each inferred lineage separately to identify specific evolutionary and phylogeographic processes.

### 2.2.5 Genetic differentiation among nuclear-plastid lineages

Between-population genetic differentiation was investigated for nuclear data (i) over all populations for each locus and at the multilocus level, (ii) within the identified nuclear-plastid Western and Eastern lineages and (iii) between populations (over all loci) using Weir and Cockerham (1984) ANOVA procedure. Pairwise $F_{ST}$ values and their significance (G test) were calculated using GENEPOP. The partitioning of nuclear genetic diversity within and among genetically distinct lineages inferred from the spatial genetic structure analyses, was estimated following


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



the procedure proposed by Yang (1998) and implemented in the R package HIERFSTAT (Goudet, 2005). The among-plastid haplotype component is given by $F_{H/T}$ and its significance was tested by permuting individuals of each population across plastid haplotypes. The between-nuclear-plastid Western-Eastern lineage component is given by $F_{L/T}$ and its significance was tested by permuting individuals of each haplotype across nuclear-plastid lineages. Genetic differentiation among each pairwise comparison of plastid haplotypes was also performed. Variance in microsatellite allele size due to stepwise changes of repeat number can also account for the level of genetic differentiation and may give insights on the relative contribution of mutation versus migration rates in genetic divergence. Using SPAGeDi, we thus also calculated the microsatellite allele size-based estimates of genetic differentiation $R_{ST}$ (Slatkin, 1995) and tested them for significance by allele-size permutations following Hardy et al. (2003). Finally, we estimated the proportion of specific alleles of each identified lineage for nuclear microsatellite loci.

### 2.2.6 Spatial genetic structure within nuclear-plastid lineages

To detect more subtle cryptic genetic structure, we reanalysed the nuclear data set by carrying out Bayesian clustering and sPCA separately for the Eastern and Western lineages. We also tested for isolation by distance (IBD) by performing a spatial autocorrelation analysis on (i) the whole plastid data set, and (ii) for the nuclear data within the two identified nuclear-plastid lineages. For plastid data, we used classic one-dimensional spatial correlograms based on Moran's *I*. Random permutations of population locations were carried out to test the statistical significance of each Moran's *I*. For nuclear data, one-dimensional Mantel correlograms were designed according to Oden and Sokal (1986). The normalised Mantel statistic *rz* (Smouse et al., 1986) was used as described in Oden and Sokal (1986) to estimate the relationships between pairwise genetic distances (DCE distance) and pairwise geographical distances among populations. All calculations were carried out using PASSAGE version 2 (Rosenberg and Anderson, 2011).

## 2.3 Results

### 2.3.1 Nuclear and plastid genetic variation within populations

Exact tests for LD between microsatellite loci only showed 38 significant P-values out of 7690 comparisons (0.49%, 384 expected from type 1 error at $\alpha$ = 0.05). When analysed over all populations, 11 out of 13 loci showed significant deviation from Hardy-Weinberg (HW) expectations,

© 2016 Tous droits réservés.



lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



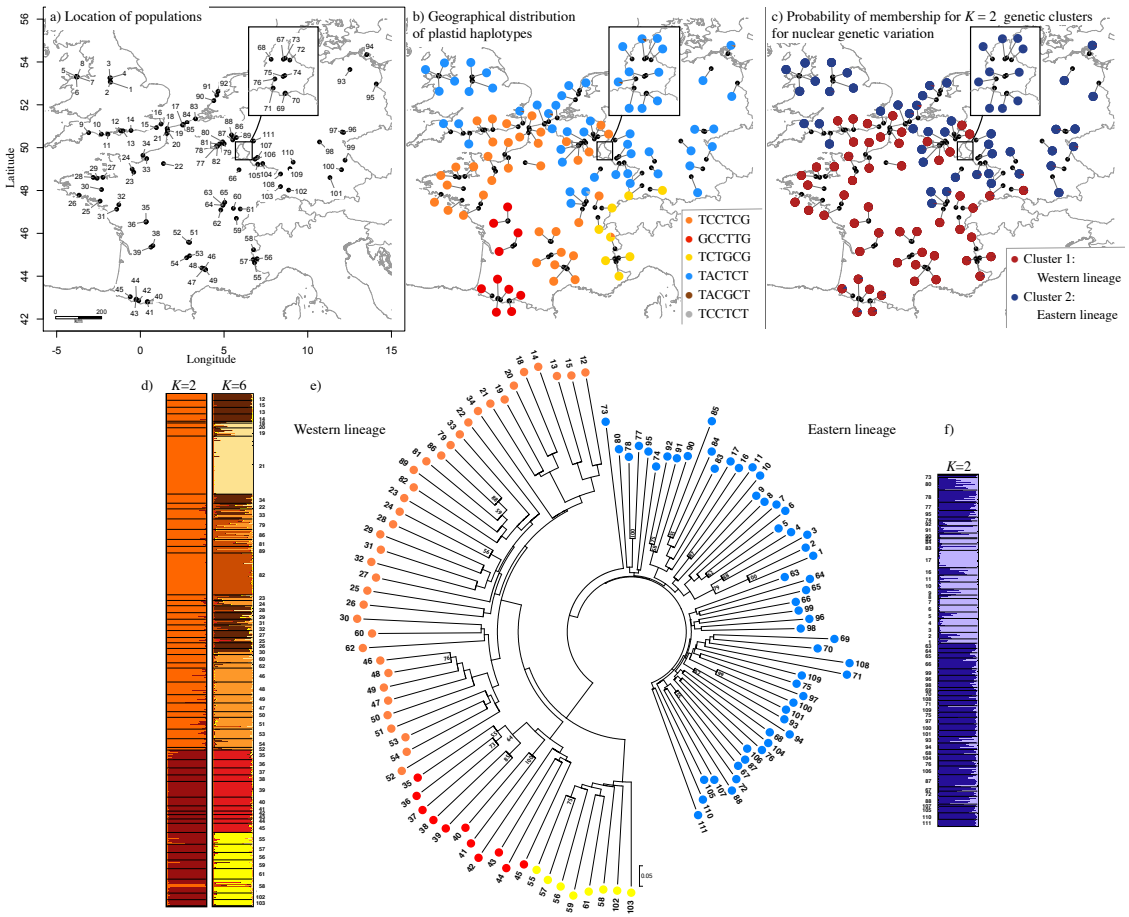



**Figure 2.1:** Map of studied populations of *Silene nutans* in Western Europe. (a) Geographical location and population numbers, (b) Geographical distribution and within-population proportion of plastid haplotypes based on six SNPs, (c) Map of mean population membership probabilities of belonging to the two modal clusters based on nuclear microsatellite loci, (d) Assignment probabilities of individual membership into $K$ = 2 and $K$ = 6 clusters for Western lineage. Each individual is represented by a thin line partitioned into $K$ colored segments displaying the individual's estimated membership fractions in $K$ clusters (e) NJ tree based on Cavalli-Sforza and Edwards' distance (1967), using nuclear data; only bootstraps > 50% are indicated. Colored circles next to population codes referred to plastid haplotype found in populations. (f) Assignment probabilities of individual membership into $K$ = 2 clusters for Eastern lineage. Each individual is represented by a thin line partitioned into two colored segments displaying the individual's estimated membership fractions in two clusters. The inset box in (a), (b) and (c) zoom on a densely sampled area in greater detail.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



**Table 2.1:** Estimates of nuclear genetic variation for 13 microsatellite loci over 111 populations of *Silene nutans* from Western Europe: total number of alleles (*An*), the allelic richness (*Ar*), the observed heterozygosity ($H_O$), expected heterozygosity ($H_E$) and the mean fixation indices [$F_{IS}$, $F_{IT}$, $F_{ST}$, following Weir & Cockerham (1984)].

| Locus | Size range | $A_n$ | $A_r$ | $H_O$ | $H_E$ | $F_{IS}$ | $F_{IT}$ | $F_{ST}$ | % E-private | % W-private | % common |
|---|---|---|---|---|---|---|---|---|---|---|---|
| B09 | 162-198 | 29 | 6.1 | 0.487 | 0.585 | 0.140*** | 0.387*** | 0.287*** | 0 | 65.5 | 34.5 |
| D10 | 175-328 | 38 | 6.3 | 0.450 | 0.594 | 0.246*** | 0.461*** | 0.285*** | 10.5 | 42.1 | 47.4 |
| E08 | 214-370 | 72 | 6.3 | 0.437 | 0.512 | 0.150*** | 0.463*** | 0.369*** | 11.1 | 65.3 | 23.6 |
| G01 | 233-357 | 79 | 8.7 | 0.779 | 0.854 | 0.099*** | 0.227*** | 0.141*** | 6.3 | 17.7 | 75.9 |
| H07 | 158-313 | 46 | 6.4 | 0.551 | 0.685 | 0.206*** | 0.379*** | 0.218*** | 6.5 | 52.3 | 41.3 |
| Sil16 | 114-204 | 29 | 6.5 | 0.524 | 0.666 | 0.227*** | 0.415*** | 0.244*** | 13.8 | 37.9 | 48.3 |
| Sil31 | 129-234 | 49 | 5.3 | 0.447 | 0.547 | 0.208*** | 0.436*** | 0.287*** | 22.4 | 44.9 | 32.6 |
| Sil35 | 84-225 | 49 | 7.8 | 0.701 | 0.757 | 0.094*** | 0.267*** | 0.191*** | 28.6 | 22.4 | 49.0 |
| Sil37 | 111-252 | 47 | 6.9 | 0.709 | 0.725 | 0.028** | 0.226*** | 0.204*** | 4.2 | 51.1 | 44.7 |
| Sil19 | 111-165 | 26 | 6.0 | 0.577 | 0.580 | 0.012 | 0.365*** | 0.357*** | 23.1 | 26.9 | 50.0 |
| Sil24 | 128-247 | 86 | 8.3 | 0.768 | 0.767 | -0.007 | 0.181*** | 0.188*** | 8.1 | 54.6 | 37.2 |
| Sil36 | 70-158 | 33 | 6.0 | 0.449 | 0.612 | 0.284*** | 0.494*** | 0.293*** | 42.4 | 21.2 | 36.4 |
| Sil42 | 97-202 | 40 | 7.5 | 0.550 | 0.752 | 0.278*** | 0.406*** | 0.177*** | 35.0 | 7.5 | 57.5 |
| Multilocus | - | 47.9 | 6.8 | 0.571 | 0.664 | 0.146*** | 0.357*** | 0.247*** | 14.8 | 40.4 | 43.2 |

The proportion of diagnostic alleles of Eastern lineage (% E-private), Western lineage (% W-private) and common alleles (% common) in both nuclear-plastid lineages are also given. ** $P < 0.01$, *** $P < 0.001$.

with mean single-locus $F_{IS}$ values ranging from -0.007 to 0.284 (Table 2.1). The mean multilocus $F_{IS}$ value was 0.146. At population level, $F_{IS}$ values ranged from -0.230 to 0.513, with 89 out of 111 populations (80.1%) showing significantly positive multilocus $F_{IS}$ values (Table S2.1 in the appendices). Corrected within-population multilocus $F_{IS}$ estimates ($F_{IS\_corr}$) accounting for null allele occurrence always yielded lower values with an arithmetic mean of 0.049. Using the classical relationship ($s = 2F_{IS}/(1+F_{IS})$ expected at genetic equilibrium for a mixed-mating system, estimated mean selfing rates were of 0.254 for $F_{IS}$ and of 0.093 for $F_{IS\_corr}$, respectively. Estimations of selfing rates based on standardized identity disequilibrium ranged from 0 to 0.379 with an arithmetic mean of 0.060 (Table S2.1 in the appendices). It should be kept in mind that selfing rate estimations should be interpreted with caution. Indeed, populations with small sample sizes lead to large variance in s estimates and an upward bias is expected when propensity for selfing is low.

Nuclear microsatellite loci displayed high levels of polymorphism, with the number of alleles ranging from 26 (Sil19) to 86 (Sil24), for an average number of 47.9 alleles over all populations (Table 2.1). Mean $H_O$ and $H_E$ ranged from 0.437 to 0.779 and from 0.512 to 0.854, respectively (Table 2.1). The map interpolating levels of allelic richness (Figure 2.2) and the multiple regression analysis (adjusted $R^2 = 0.422$, $F_{(2,102)} = 38.9$, $P < 0.001$, Variance Inflation Factor = 1.003; Pearson's correlation coefficient between latitude and longitude $r = 0.054$) indicated a significant decrease in allelic richness along a southern-northern ($\beta = -0.150$, $P < 0.001$) and an eastern-western gradient ($\beta = 0.073$, $P < 0.001$).

For plastid haplotype genetic diversity, we found a total of six different haplotypes, of

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



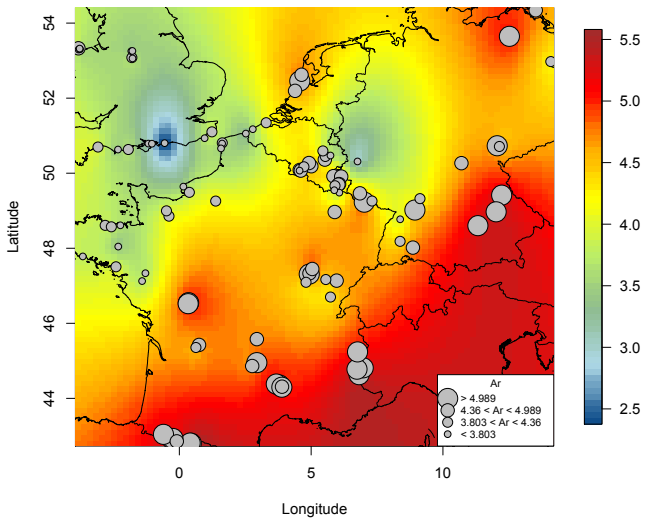



**Figure 2.2:** Map of observed level of allelic richness (*Ar*) found in each population of *Silene nutans* (displayed as proportional circles). Spatially interpolated levels of allelic richness can also be visualized as color plot gradient (from dark blue for low values to dark red for high values).

which four corresponded to 99.5% of the overall plastid diversity observed (Figure 2.1.b, Table S2.1 in the appendices): the TACTCT SNP combination (hereafter called the blue haplotype), TCCTCG SNP combination (orange haplotype), GCCTTG SNP combination (red haplotype) and TCTGCG SNP combination (yellow haplotype). All populations were monomorphic, except seven populations that showed intra-population plastid polymorphism, with the second haplotype occurring at very low frequency (Table S2.1 in the appendices). The Belgian *Ca* ecotype and the British var. *smithiana* showed the blue haplotype, whereas the *Si* ecotype and the British var. *salmoniana* showed the orange haplotype (Figure S2.1.b,c, Table S2.1 in the appendices).

### 2.3.2 Lineage distinctiveness and spatial patterns of genetic structure

The four main plastid haplotypes showed geographic structuring (Figures 2.1.b, S2.2) and several geographical areas in which two haplotypes co-occurred (Figures 2.1.b, S2.1.b/d). Given the lack of intra-population polymorphism and the strong spatial structure observed, these plastid haplotypes were used as a diagnostic criterion for population lineage identity. Based on nuclear microsatellite data, and taking plastid identity into account, a clear phylogeographic pattern and genetically differentiated nuclear-plastid lineages emerged from (i) the NJ tree based on allele frequencies using DCE (Figure 2.1.d); (ii) the non-spatially explicit Bayesian clustering method (Figure 2.1.c, e-h); and (iii) the sPCA (Figure 2.3). Populations clearly clustered according to their plastid haplotype identity, with a Western lineage containing the red, orange and yellow plastid


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



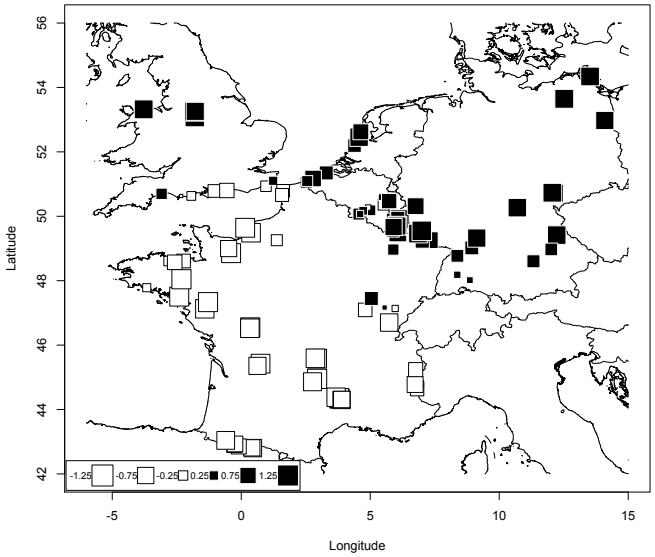



**Figure 2.3:** Geographic map of the first global scores of a spatial principal component analyse (sPCA) performed on the whole dataset in *Silene nutans* ($I$ = 0.631, var = 0.851, $P$ < 0.001). Each square represents a population. Populations that are more closely related in multivariate space share the same colour of squares. Sizes of squares indicate the magnitude of spatial positive autocorrelation.

haplotypes, and an Eastern lineage characterised by the blue plastid haplotype. The highest likelihood found using the $\Delta K$ statistic corroborated the occurrence of the two main nuclear-plastid lineages (Figure S2.3). The sPCA performed on the whole nuclear data set revealed significant global structure ($P$ < 0.001). The first two sPCA axes accounted for most of the spatial genetic structure, covering a high proportion of the spatial autocorrelation and most of the variance in the genetic data ($I$ = 0.631, var = 0.851 and $I$ = 0.610, var = 0.223, respectively). The first two eigenvalues were 0.537 and 0.136, while the remaining eigenvalues were < 0.090. The first axis in the sPCA revealed a striking longitudinal cline, differentiating a Western and an Eastern nuclear-plastid lineage (Figure 2.3).

### 2.3.3 Genetic differentiation among nuclear-plastid lineages

The overall and pairwise genetic differentiation for nuclear data were high and highly significant, with single-locus $F_{ST}$ estimates ranging from 0.177 to 0.369, and a mean multilocus $F_{ST}$ value of 0.247 (all significant at $P$ < 0.001, see Table 2.1). Pairwise $F_{ST}$ estimates between the 111 studied populations ranged from 0 to 0.608, with 52.8% of significant estimates (Figure S2.4).

The overall genetic distinctiveness among plastid haplotypes was supported by a significant hierarchical nuclear genetic differentiation ($F_{H/T}$ = 0.143, $P$ < 0.001). The highest levels of nuclear genetic differentiation among haplotypes were found between blue and orange haplo-

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



types ($F_{BO}$ = 0.157, $P$ < 0.001) and between blue and red haplotypes ($F_{BR}$ = 0.146, $P$ < 0.001). The differentiation was weaker, but still significant between orange and red haplotypes ($F_{OR}$ = 0.078, $P$ < 0.010). We found no significant nuclear genetic differentiation between the yellow and any other haplotype, likely due to small sample size (the yellow haplotype occurred in only eight populations). Nuclear genetic differentiation between the Eastern and Western lineages was highly significant ($F_{L/T}$ = 0.126, $P$ < 0.001). Mean genetic differentiation (as measured by multilocus $F_{ST}$ estimates) within these two main lineages was 0.187 and 0.195 for Eastern and Western lineages, respectively (all at $P$ < 0.001). Mean multi-locus allele size-based estimate of genetic differentiation $R_{ST}$ were (i) of 0.357 ($P$ < 0.01) among the 111 sampled populations, (ii) of 0.173 ($P$ > 0.05) between the western and eastern lineages, (iii) of 0.014 ($P$ > 0.05) between the two main genetic clusters within the eastern lineage and (iv) of 0.274 ($P$ < 0.01) between the two main genetic clusters within the western lineage. Over all loci, only 43% of the alleles were shared between the Eastern and Western nuclear-plastid lineages (Table 2.1).

### 2.3.4 Spatial genetic structure within nuclear-plastid lineages

When focusing on the Western lineage, the best mode suggested by the Bayesian clustering was $K$ = 2 (Figure S2.3), mirroring the nuclear genetic divergence between the red and the yellow plastid haplotype populations and the orange haplotype populations (Figure 2.1.f). Additional hierarchical structure was detected at $K$ = 6 (Figure S2.3): red and yellow haplotype populations were distinct and the orange haplotype was partitioned into four further population clusters, corresponding to four geographical regions: southern Belgium (*Si* ecotype), northern France, western France together with southern England (var. *salmoniana*), and south-central France (Figure 2.1.a,g, Table S2.1 in the appendices). sPCA revealed a latitudinal (north-south) cline regardless of haplotype identity (Figure 2.4.a). This global structure ($P$ = 0.001) showed a strong signal of positive spatial autocorrelation ($I$ = 0.694) and represented a high proportion of the genetic variability (var = 0.471). The Mantel correlogram displayed a clear continuous decrease in genetic similarity with increasing geographic distance (Figure 2.4.a). When analysing the Eastern nuclear-plastid lineage, a partition between (i) inland continental populations (Germany, Luxemburg, eastern France, and the *Ca* ecotype from southern Belgium) and (ii) populations located near the Channel coast (northern France, Belgium and The Netherlands) and in the UK (var. *smithiana*) was underlined by both Bayesian clustering with the highest likelihood for $K$ = 2 (Figures 2.1.h, S2.3) and the plot of scores from the first global component of the sPCA (Figure 2.4.b). The global structure ($P$ = 0.034) showed a strong signal of positive spatial

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



autocorrelation ($I$ = 0.784). The Mantel correlogram pictured a typical pattern of long-distance differentiation due to a substantial geographic break in allele frequencies across the range of the Eastern lineage (Figure 2.4.b).

### 2.3.5 Detection of hybridisation events in contact zones

Although most populations were fixed for a single plastid haplotype, we tested for the possibility of hybridisation events mediated by pollen in contact zones. We focused on three areas where populations of different haplotypes co-occurred, and ran further Bayesian assignments on these subsets of populations (see Figure S2.1). Admixture was absent: groups identified by Bayesian clustering always referred clearly to haplotype identity with no signature of admixed nuclear genotypes or evidence of discordance between nuclear membership probability and plastid haplotype identity (Figure S2.1).

## 2.4 Discussion

### 2.4.1 Broad patterns of nuclear and plastid genetic diversity

*S. nutans* has been described as a self-compatible, gynomonoecious-gynodioecious species with a censer mechanism of seed dispersal (Hepper, 1956). Hermaphroditic plants are able to self by geitonogamy, with an extreme variation in selfing rates among individuals (Dufay et al., 2010). $F_{IS}$ values were positive and significant for several populations and most loci, suggesting propensity for selfing even when accounting for null allele occurrence. Assuming that populations are at genetic equilibrium, the classic relationship between the intra-population fixation index and the selfing rate [$(s)$=$2F_{IS}/(1+F_{IS})$] gave a mean selfing rate of 25.4% and 9.3% for $F_{IS}$ and $F_{IS\_corr}$, respectively. However, positive significant $F_{IS}$ values may also result from biparental inbreeding due to restricted pollen and/or seed dispersal. As a consequence, mean within-population selfing rate based on identity disequilibrium and less sensitive to biparental inbreeding (David et al., 2007) gave a much lower rate of 6%. Overall, biparental inbreeding due to mating between relatives, along with a low rate of self-fertilization, is expected to amplify the potential for IBD pattern developing kin-structure at a local scale (Heywood, 1991, Vekemans and Hardy, 2004).

Inbreeding, along with restricted gene flow, leads to high genetic differentiation among populations. The mean level of genetic differentiation within Western ($F_{ST}$ = 0.195) and Eastern ($F_{ST}$ = 0.187) lineages of *S. nutans* was concordant with a mixed-mating system and moderate

© 2016 Tous droits réservés.



lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



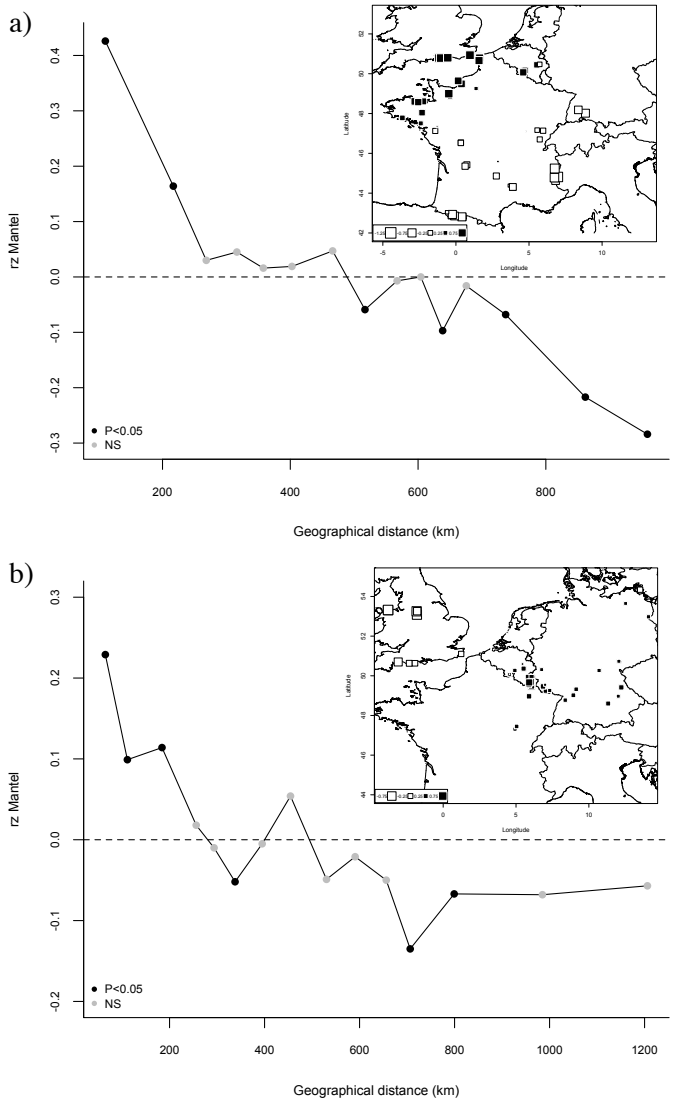



**Figure 2.4:** Spatial genetic structure within the two main Western and Eastern nuclear-plastid lineages in *Silene nutans*. One dimensional Mantel correlograms of (a) Western and (b) Eastern lineage. To reduce the random effects of sampling and ensure statistical supports for calculations, correlograms were designed with a minimum number of respectively 99 and 102 pairs of populations in each distance class. *rz*: normalized Mantel statistics. For each correlogram plot, the geographical map of the first global scores of sPCAs is also given. Each square represents a population. Populations that are more closely related in multivariate space share the same colour of squares. Sizes of squares indicate the magnitude of spatial positive autocorrelation


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



levels of gene flow among populations (Duminil et al., 2009). The nearly complete fixation of plastid haplotypes and their strong geographical clustering may either confirm a spatially restricted seed dispersal and/or a lack of within-lineages plastid diversity with density-dependent processes or ecological factors that constraint lineages' distribution (Waters et al., 2013). Moreover, a continuous decrease in genetic similarity with increasing geographical distance was shown, which suggested limited pollen/seed dispersal. Therefore, to conserve substantial levels of genetic diversity, in situ as well as ex situ, the preserved populations should reflect the nuclear-plastid differentiation patterns. Attention should particularly be paid to the southern populations which have conserved the highest allelic richness.

### 2.4.2 Phylogeographic patterns in *Silene nutans*

At a continental scale, leading-edge expansions following the Last Glacial Maximum may result in multiple phylogeographic patterns that are species-specific since each taxon responded independently to Quaternary cold periods (Taberlet et al., 1998). Rapid postglacial recolonization through repeated population bottlenecks of founding lineages, blocking establishment of latecomers, is expected to reduce within-lineage genetic diversity with increasing distance from its refugium (Waters et al., 2013).

In Western Europe, *S. nutans* showed a phylogeographic pattern with two main, strongly differentiated Western and Eastern nuclear-plastid lineages ($F_{L/T}$ = 0.126). Patterns of long-distance differentiation and spatial trends in allelic richness suggested a Western lineage originating from the western Mediterranean Basin and an Eastern lineage, probably originating from Eastern Europe. The Eastern lineage exhibited a north-westward expansion, from eastern Germany to northern Wales with a genetic break found along the continental European coastline. The Western lineage harboured three plastid haplotypes, spatially structured, consistent with a northward expansion from distinct potential refugia. These hypotheses are supported by the asymmetrical patterns of private alleles between the two main nuclear-plastid lineages. Most private alleles were found in the Western lineage. Moreover, allele-sized based measure of genetic differentiation yielded a highly significant $R_{ST}$ estimate of 0.274 between the two main genetic clusters found in this Western lineage. Overall, this could suggest (i) a predominant role of stepwise-like mutation processes over migration and relative cluster isolation, and (ii) different glacial refugia composed of distinct genetic pools. We should keep in mind that our microsatellite loci may not follow a strict stepwise mutation process and that homoplasy is likely to occur, which may explain why no significant $R_{ST}$ estimate was detected between western and eastern

© 2016 Tous droits réservés.

 lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



lineages.

### 2.4.3 Secondary contacts: from diverging lineages to cryptic species

Beyond shaping the distribution of present genetic diversity, Quaternary climate oscillations have been shown to favour speciation processes (Hewitt, 2000). Genetic drift and selection operating in different, isolated glacial refugia can lead to genetic divergence. Patterns of postglacial expansion from different refugia found in animals, plants, fungal taxa, and described in *S. nutans* here, produce suture zones in which lineages come into secondary contact (Taberlet et al., 1998, Hewitt, 2000, 2011). Many of these suture zones have been described across Europe, such as in the Alps, the Pyrenees or in eastern Central Europe along the eastern borders of Germany (Hewitt, 2000). Accordingly, our results in Western Europe identified three suture zones where the Eastern and Western nuclear-plastid lineages of *S. nutans* co-occur: in southern Belgium, eastern France and, surprisingly, in southern England. While previous phylogeographic studies have reported that only one lineage colonised Great Britain (e.g. oaks expanding from Spain or grasshoppers from the Balkans, reviewed in Hewitt, 2011), our study on *S. nutans* is, to our knowledge, the first that shows the existence of a contact zone in Great Britain shaped by postglacial expansion.

Our study shows that the two morphological varieties previously described in Hepper (1951) in Great Britain, *S. nutans* var. *salmoniana* and var. *smithiana*, belong to separate nuclear-plastid lineages, Western and Eastern, respectively. Morphological differences, e.g. differences in fruit size, have been also found elsewhere, in particular between blue and orange/red haplotype populations, while the continental populations belonging to the yellow and blue haplotypes did not show such contrasting differences (Van Rossum, unpubl. data). Moreover, the two reproductively isolated Belgian *Ca* and *Si* parapatric edaphic ecotypes also correspond to the Eastern and Western nuclear-plastid lineages, respectively. Bayesian clustering performed on the three contact zones identified in *S. nutans* did not reveal any signatures of admixture between haplotype lineages, suggesting that no hybridisation events occur in natural populations. Such result was somewhat unexpected given that range shifts and expansions of a colonising species into the range of a local species may lead to haplotype sharing due to longer pollen than seed dispersal Petit and Excoffier (e.g. 2009). Altogether, this strict association between nuclear gene pools and plastid haplotype suggests cryptic speciation in *S. nutans*, at least for Eastern and Western nuclear-plastid lineages for which a strong reproductive barrier can be suspected. From an evolutionary point of view, suture zones resulting from endogenous selection are not stabilised geographically and their dynamics are thought to be maintained by a balance between disper-

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



sal and selection against hybrids (Barton, 1979). These tension zones may be associated with an environmental transition where different genotypic combinations allow differential adaptive values, resulting in a genetic-environment association and ecotypic differentiation (Barton and Hewitt, 1985, Hewitt, 1988, Bierne et al., 2011). While in southern Belgium, the genetic lineages are clearly associated with edaphic properties (e.g. Van Rossum et al., 1997), to date, no evidence of such association between genetic variation and soil characteristics have been found elsewhere (Van Rossum et al., 2003, Van Rossum and Prentice, 2004). This inconsistency suggests that the divergence of the two main lineages precedes the specific adaptation to the nature of soil and indicates locally restricted association between edaphic adaptation and lineage identity.

## 2.5 Conclusion

Based on our findings on genetic structure in *S. nutans*, two strongly differentiated evolutionary lineages were identified with no hybridisation events detected using nuclear markers. Besides, no plastid haplotypes were shared between these two lineages. This suggests the occurrence of at least two cryptic species within *S. nutans*, with no introgression events affecting species integrity. According to Mayr's biological species concept (Coyne and Orr, 2004), two main ESUs can therefore be defined in Western Europe, corresponding to Eastern and Western nuclear-plastid lineages. Moreover, within the Western lineage, three potential ESUs might also be considered, corresponding to the three sub-lineages. Given the differences in their distribution range, the conservation status of these distinct ESUs should be evaluated separately. In situ preservation of populations and genetic rescue implying ex situ conservation techniques should take the lineage identity into account. Applying such conservation approaches are particularly true in Great Britain, northern France and Belgium, where *S. nutans* is rare and where distinct lineages co-occur in close contact.

## Acknowledgements

We would like to thank three anonymous referees and the associate editor for helpful comments that greatly improved the manuscript. This work was supported by a grant from the "Region Nord-Pas de Calais" (Genefrag project) to JFA and PT, a grant from the French Agence Nationale de la Recherche (ANR-11-BSV7-013-03, TRANS) to PT, and a PhD fellowship from the French Research Ministry to HM. We thank S. Le Cadre (Postdoctoral Fellowship from National

© 2016 Tous droits réservés.



lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Fund for Research Luxembourg), B. Brachi, E. Schmitt O. Raspé and P. Meerts for their contribution to plant sampling; the Conservatoire Botanique National de Rennes, the BSBI (D.A. Pearman, R.E.N. Smith, M.J. Hawksford, S. Wild, W. McCarthy, A. Willmot, M.W. Rand, A. Knapp, J. Knight, Q. Groom), MNHN Luxemburg (G. Colling), Flo.Wer (W. Van Landuyt), Floron (W. van der Slikke) and Bundesamt für Naturschutz (e.g. R. May, R. Hand, M. Engelhart, G. Müller) for providing location of populations; and local managers (e.g. L. Woué - Ardenne & Gaume, G. Bryant - Natural England, M. Bartlett - Browndown Training Camp, E.J. van Nieukerken - Natuurmonumenten, H. Hagen - DZH, H. Kivit- PWN) for access to natural sites. Numerical results presented in this paper were carried out using the European Grid Infrastructure (http://www.egi.eu) with the Biomed virtual organization (http://lsgc.org/en/Biomed:home) via DIRAC portal (http://diracgrid.org) supported by France Grille (http://www.france-grilles.fr/), thanks to the expertise of Sophie Gallina. We thank the technical staff of the European Grid Infrastructure and the supporting National Grid Initiatives for providing technical support and infrastructure.

## 2.6 Bibliography

Avise, J. C. (2000). *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA.

Barton, N. H. (1979). The dynamics of hybrid zones. *Heredity*, 43:341–359.

Barton, N. H. and Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, 16:113–148.

Bierne, N., Welch, J., Loire, E., Bonhomme, F., and David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, 20:2044–2072.

Cavalli-Sforza, L. L. and Edwards, A. W. F. (1967). Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*, 19:233–257.

Chybicki, I. J. and Burczyk, J. (2009). Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity*, 100:106–113.

Coyne, J. A. and Orr, H. A. (2004). *Speciation*. Sinauer associates, Sunderland, Massachusetts, USA, 1st edition.

David, P., Pujol, B., Viard, F., Castella, V., and Goudet, J. (2007). Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology*, 16:2474–2487.

De Bilde, J. (1973). Etude génécologique du *Silene nutans* L. en Belgique: populations du *Silene nutans* L. sur substrats siliceux et calcaires. *Revue Générale de Botanique*, 80:161–176.

DiLeo, M. F., Row, J. R., and Lougheed, S. C. (2010). Discordant patterns of population structure for two co-distributed snake species across a fragmented Ontario landscape. *Diversity and Distributions*, 16:571–581.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Dufay, M., Lahiani, E., and Brachi, B. (2010). Gender variation and inbreeding depression in gynodioecious-gynomonoecious *Silene nutans* (Caryophyllaceae). *International Journal of Plant Sciences*, 171:53–62.

Duminil, J., Fineschi, S., Hampe, A., Jordano, P., Salvini, D., Vendramin, G. G., and Petit, R. J. (2007). Can population genetic structure be predicted from life-history traits? *The American Naturalist*, 169:662–672.

Duminil, J., Hardy, O. J., and Petit, R. J. (2009). Plant traits correlated with generation time directly affect inbreeding depression and mating system and indirectly genetic structure. *BMC Evolutionary Biology*, 9:177–190.

El Mousadik, A. and Petit, R. J. (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinos* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, 92:832–839.

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

Fitter, A. (1978). *An atlas of the wild flowers of Britain and Northern Europe*. Collins, London, UK.

Fraser, D. J. and Bernatchez, L. (2001). Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, 10:2741–2752.

Gavin, D. G., Fitzpatrick, M. C., Gugger, P. F., Heath, K. D., Rodrìguez-Sànchez, F., Dobrowski, S. Z., Hampe, A., Hu, F. S., Ashcroft, M. B., Bartlein, P. J., Blois, J. L., Carstens, B. C., Davis, E. B., de Lafontaine, G., Edwards, M. E., Fernandez, M., Henne, P. D., Herring, E. M., Holden, Z. A., Kong, W.-s., Liu, J., Magri, D., Matzke, N. J., McGlone, M., Saltré, F., Stigall, A. L., Tsai, Y.-h. E., and Williams, J. W. (2014). Climate refugia : joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, 204:37–54.

Godé, C., Touzet, P., Martin, H., Lahiani, E., Delph, L. F., and Arnaud, J.-F. (2014). Characterization of 24 polymorphic microsatellite markers for *Silene nutans*, a gynodioecious-gynomonoecious species, and cross-species amplification in other *Silene* species. *Conservation Genetics Resources*, 6:915–918.

Goudet, J. (1995). FSTAT (Version 1.2): a computer program to calculate *F*-statistics. *The Journal of Heredity*, 86:485–486.

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes*, 2:184–186.

Hamrick, J. L. and Godt, M. J. W. (1996). Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 351:1291–1298.

Hardy, O. J., Charbonnel, N., Fréville, H., and Heuertz, M. (2003). Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics*, 163:1467–1482.

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

Hautekèete, N.-C., Frachon, L., Luczak, C., Toussaint, B., Van Landuyt, W., Van Rossum, F., and Piquot, Y. (2015). Habitat type shapes long-term plant biodiversity budgets in two densely populated regions in north-western Europe. *Diversity and Distributions*, 21:631–642.

Hepper, F. N. (1951). The variations of *Silene nutans* L. in Great Britain. *Watsonia*, 2:80–90.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Hepper, F. N. (1956). Biological flora of the British Isles: *Silene nutans* L. *Journal of Ecology*, 44:693–700.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405:907–913.

Hewitt, G. M. (1988). Hybrid zones - natural laboratories for evolutionary studies. *Trends in Ecology and Evolution*, 3:158–167.

Hewitt, G. M. (2011). Quaternary phylogeography: the roots of hybrid zones. *Genetica*, 139:617–638.

Heywood, J. S. (1991). Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics*, 22:335–355.

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

Jombart, T. (2008). *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24:1403–1405.

Jombart, T., Devillard, S., Dufour, A.-B., and Pontier, D. (2008). Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity*, 101:92–103.

Jombart, T., Pontier, D., and Dufour, A.-B. (2009). Genetic markers in the playground of multivariate analysis. *Heredity*, 102:330–341.

Jürgens, A., Witt, T., and Gottsberger, G. (1996). Reproduction and pollination in central European populations of *Silene* and *Saponaria* species. *Botanica Acta*, 109:316–324.

Lahiani, E., Dufaÿ, M., Castric, V., Le Cadre, S., Charlesworth, D., Van Rossum, F., and Touzet, P. (2013). Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity*, 111:157–164.

Oden, N. L. and Sokal, R. R. (1986). Directional autocorrelation : an extension of spatial correlograms to two dimensions. *Systematic Zoology*, 35:608–617.

Palsbøll, P. J., Bérubé, M., and Allendorf, F. W. (2007). Identification of management units using population genetic data. *Trends in Ecology and Evolution*, 22:11–16.

Petit, R. J. and Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology and Evolution*, 24:386–393.

Petit, R. J., Hu, F. S., and Dick, C. W. (2008). Forests of the past: a window to future changes. *Science*, 320:1450–1452.

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

R Core Team Development (2014). R: a language and environment for statistical computing. *R Foundation for Statistical Computing Vienna, Austria*, https://www.r–project.org/.

Rosenberg, M. S. and Anderson, C. D. (2011). PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. *Methods in Ecology and Evolution*, 2:229–232.

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4:137–138.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



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

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 139:457–462.

Smouse, P. E., Long, J. C., and Sokal, R. R. (1986). Multiple regression and correlation mantel test of matrix correspondence. *Systematic Zoology*, 35:627–632.

Stace, C. (2010). *New flora of the British Isles*. Cambridge University Press, Cambridge, UK, 3rd edition.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.-G., and Cosson, J.-F. (1998). Comparative phylogeography and postglacial colonization. *Molecular Ecology*, 7:453–464.

Van Rossum, F., De Bilde, J., and Lefèbvre, C. (1996). Barriers to hybridization in calcicolous and silicolous populations of *Silene nutans* from Belgium. *Belgian Journal of Botany*, 129:13–18.

Van Rossum, F. and Prentice, H. C. (2004). Structure of allozyme variation in Nordic *Silene nutans* (Caryophyllaceae): population size, geographical position and immigration history. *Biological Journal of the Linnean Society*, 81:357–371.

Van Rossum, F., Vekemans, X., Gratia, E., and Meerts, P. (2003). A comparative study of allozyme variation of peripheral and central populations of *Silene nutans* L. (Caryophyllaceae) from Western Europe: implications for conservation. *Plant Systematics and Evolution*, 242:49–61.

Van Rossum, F., Vekemans, X., Meerts, P., Gratia, E., and Lefèbvre, C. (1997). Allozyme variation in relation to ecotypic differentiation and population size in marginal populations of *Silene nutans*. *Heredity*, 78:552–560.

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

Waters, J. M., Fraser, C. I., and Hewitt, G. M. (2013). Founder takes all: density-dependent processes structure biodiversity. *Trends in Ecology and Evolution*, 28:78–85.

Weir, B. S. and Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, 38:1358–1370.

Yang, R.-C. (1998). Estimating hierarchical *F*-statistics. *Evolution*, 52:950–956.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



## 2.7 Appendices

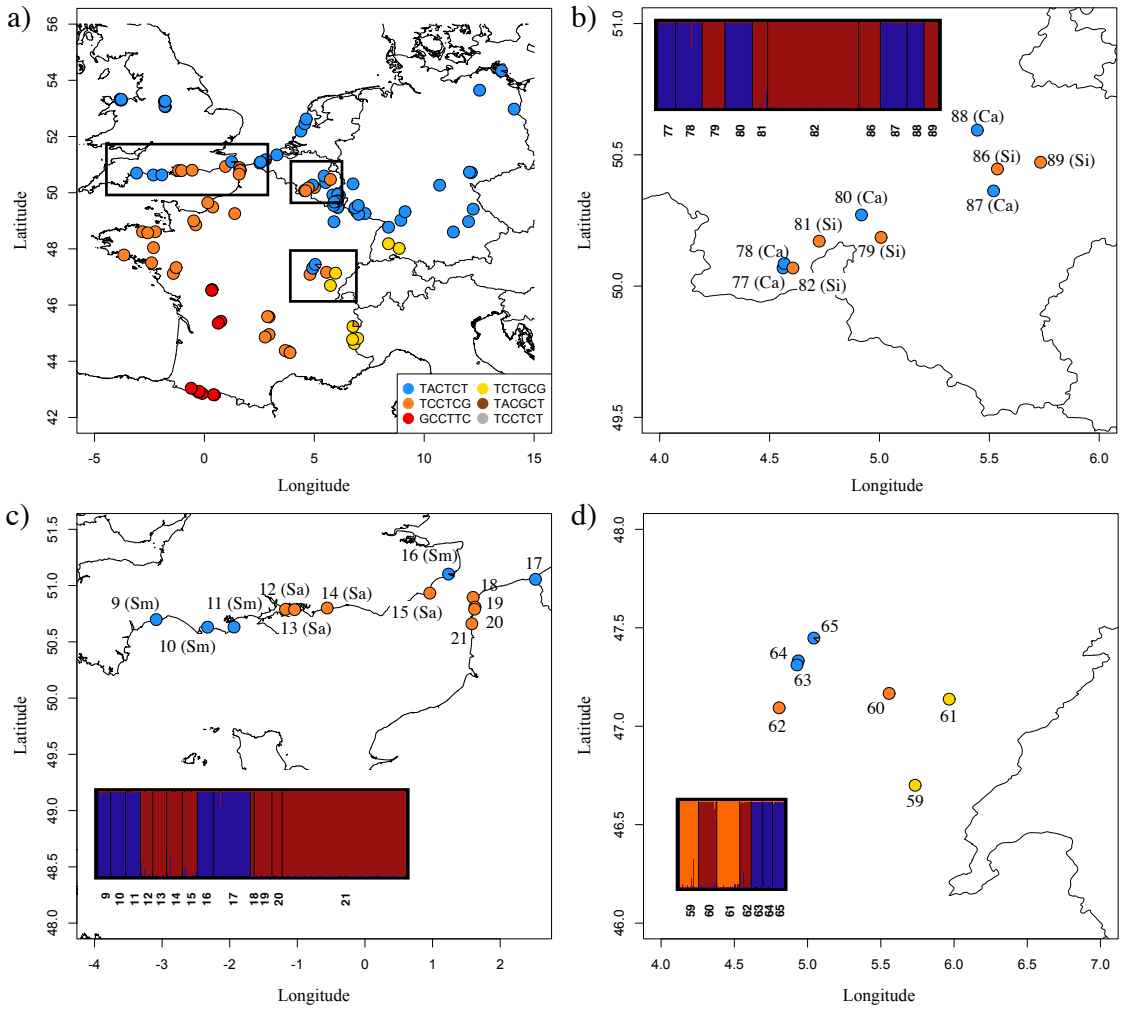



**Figure S2.1:** Map of plastid haplotype geographic distribution in *Silene nutans* over the whole studied area (a) and in potential contact zones where different plastid lineages co-occur (b-d). Results of Bayesian clustering on nuclear microsatellite loci, are displayed for (b) southern Belgium populations (modal $K$ = 2), (c) southern England and northern France populations (modal $K$ = 2) and (d) eastern France populations (modal $K$ = 3). Each individual is represented by a vertical bar, partitioned into $K$ segments corresponding to membership probability into the modal $K$ clusters. Belgian ecotype and UK varieties are indicated between brackets: Ca (calicolous ecotype), Si (silicicolous ecotype), Sm (*Silene nutans* var. *smithtiana*) and Sa (*Silene nutans* var. *salmoniana*).

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

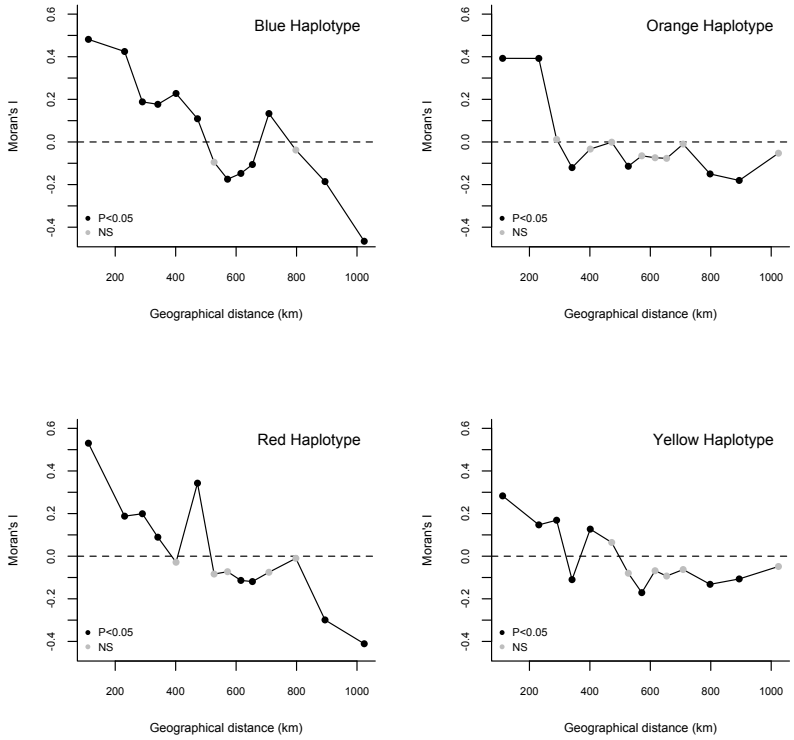



**Figure S2.2:** Spatial correlograms based on Moran's *I* for patterns of variation of plastid haplotype frequency against geographic distances (km) among populations.

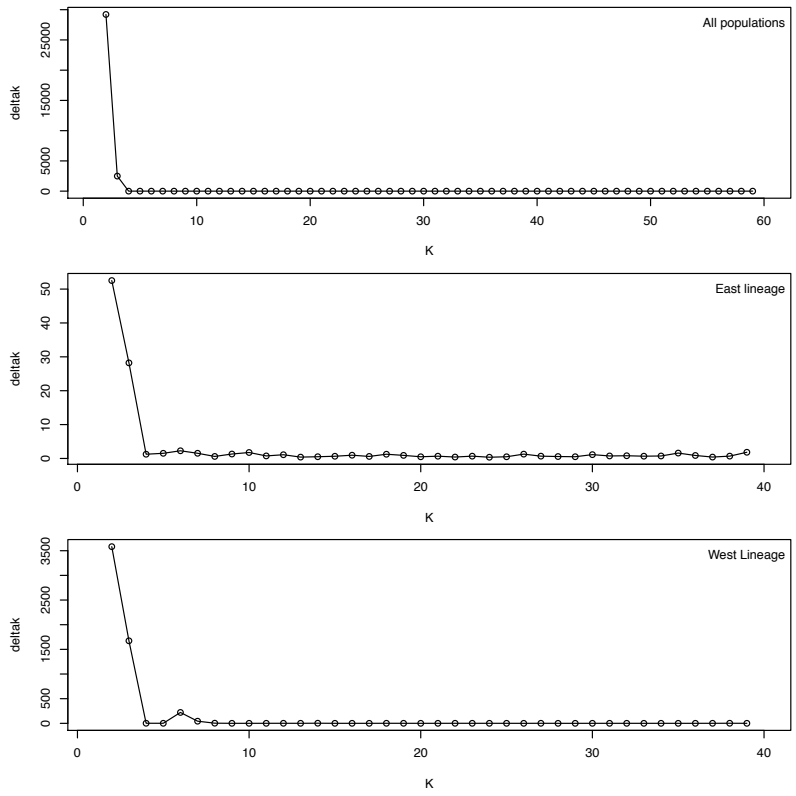



**Figure S2.3:** The relationship between $K$ (number of inferred clusters) and $\Delta K$ for the whole dataset, the Eastern lineage and the Western lineage.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



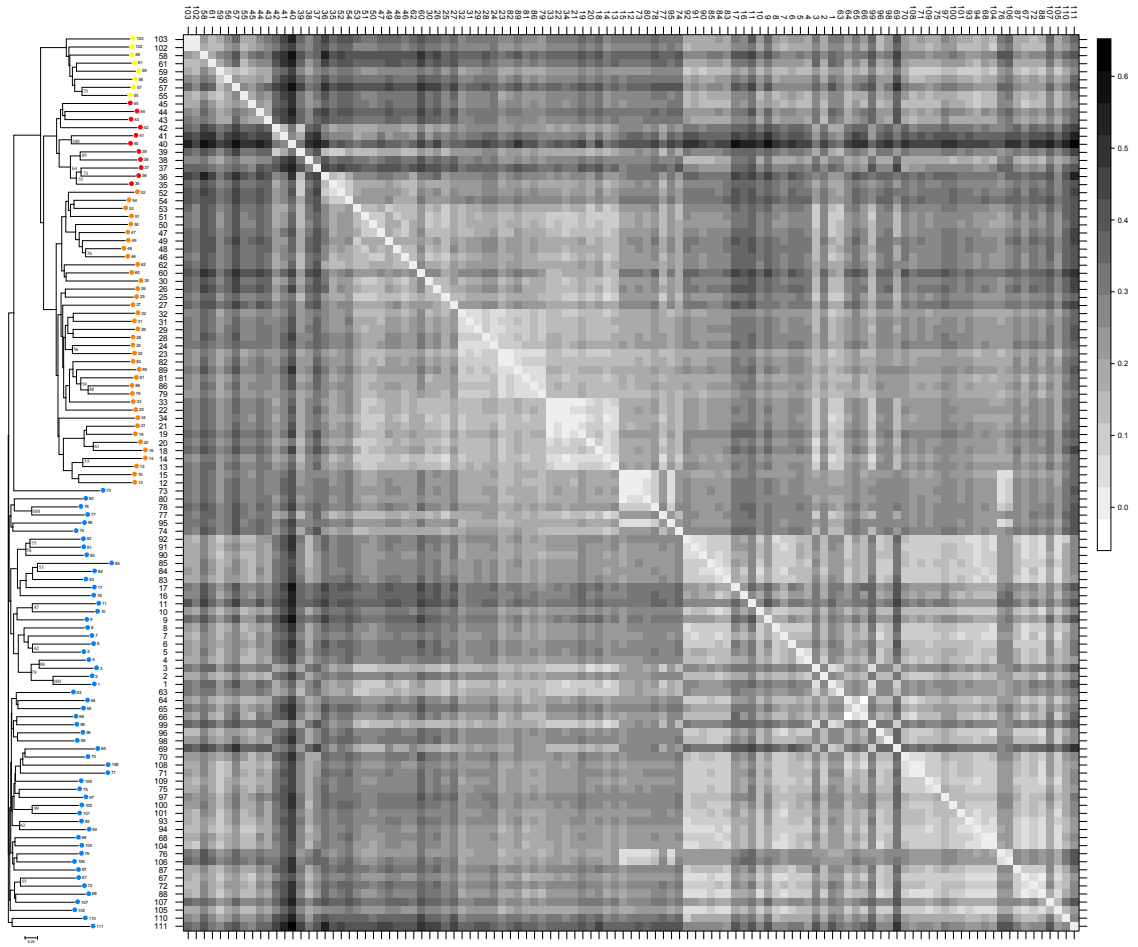


**Figure S2.4:** Matrix of pairwise $F_{ST}$ estimates (increasing with color intensity) between 111 populations of *Silene nutans* together with the NJ tree based on Cavalli-Sforza and Edwards's distance (DCE).

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



**Table S2.1:** Population details and mean estimates of genetic variation for 111 populations of *Silene nutans* from western Europe: number, country, locality, geographic coordinates, sample sizes (*n*); for 13 nuclear microsatellites loci: number of alleles (*An*), allelic richness (*Ar*), observed heterozygosity ($H_O$), expected heterozygosity ($H_E$), mean multilocus intra-population fixation index ($F_{IS}$) value, mean corrected multilocus intra-population fixation index accounting for null allele ($F_{IS_corr}$) and mean multilocus selfing-rate based on identity disequilibrium (*s*).

| | | | | | | Nuclear genetic variation | | | | | | Plastid diversity | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | Eastern Linage Haplotype | | Western Lineage Haplotype | | |
| Pop Num | Country | Locality | Latitude N | Longitude | *n* | *Ar* | $H_O$ | $H_E$ | $F_{IS}$ | $F_{IS}$–corr | *s* | Blue | Other | Orange | Red | Yellow |
| 1 | UK - Sm | Ilam (Derbyshire) | 53.059 | 1.777 W | 10 | 3.5 | 0.474 (±0.256) | 0.537 (±0.246) | 0.114** | 0.054 | 0 | 1 | 0 | 0 | 0 | 0 |
| 2 | UK – Sm | Stanshope (Derbyshire) | 53.080 | 1.802 W | 15 | 4.2 | 0.568 (±0.276) | 0.626 (±0.258) | 0.085* | 0.025 | 0.029 | 1 | 0 | 0 | 0 | 0 |
| 3 | UK – Sm | Buxton (Derbyshire) | 53.252 | 1.822 W | 15 | 3.1 | 0.426 (±0.235) | 0.456 (±0.241) | 0.052 | 0.035 | 0.159 | 1 | 0 | 0 | 0 | 0 |
| 4 | UK – Sm | Cressbrook (Derbyshire) | 53.258 | 1.782 W | 17 | 3.5 | 0.466 (±0.242) | 0.550 (±0.244) | 0.150** | 0.033 | 0 | 1 | 0 | 0 | 0 | 0 |
| 5 | UK – Sm | Llandudno (Caernarvonshire) | 53.329 | 3.829 W | 15 | 4.6 | 0.649 (±0.209) | 0.715 (±0.148) | 0.095** | 0.038 | 0.069 | 1 | 0 | 0 | 0 | 0 |
| 6 | UK – Sm | Deganwy (Caernarvonshire) | 53.295 | 3.822 W | 16 | 3.7 | 0.554 (±0.266) | 0.606 (±0.236) | 0.088* | 0.028 | 0 | 1 | 0 | 0 | 0 | 0 |
| 7 | UK – Sm | Llangwstenin (Caernarvonshire) | 53.303 | 3.769 W | 15 | 2.7 | 0.533 (±0.214) | 0.462 (±0.173) | -0.161 | 0.012 | 0.264 | 1 | 0 | 0 | 0 | 0 |
| 8 | UK – Sm | Penrhynside (Caernarvonshire) | 53.321 | 3.783 W | 8 | 3.7 | 0.559 (±0.251) | 0.626 (±0.212) | 0.107* | 0.030 | 0 | 1 | 0 | 0 | 0 | 0 |
| 9 | UK – Sm | Beer (Devon) | 50.698 | 3.087 W | 14 | 4.1 | 0.544 (±0.213) | 0.636 (±0.226) | 0.150** | 0.104 | 0.160 | 1 | 0 | 0 | 0 | 0 |
| 10 | UK – Sm | Holworth (Dorset) | 50.630 | 2.327 W | 16 | 3.2 | 0.526 (±0.192) | 0.544 (±0.177) | 0.035 | 0.040 | 0.133 | 1 | 0 | 0 | 0 | 0 |
| 11 | UK – Sm | Swanage (Dorset) | 50.632 | 1.936 W | 16 | 4.0 | 0.554 (±0.270) | 0.619 (±0.236) | 0.104* | 0.018 | 0.049 | 1 | 0 | 0 | 0 | 0 |
| 12 | UK - Sa | Gosport (South Hampshire) | 50.789 | 1.175 W | 13 | 3.2 | 0.509 (±0.299) | 0.581 (±0.221) | 0.122** | 0.030 | 0.121 | 0 | 0 | 1 | 0 | 0 |
| 13 | UK – Sa | Porthmouth (South Hampshire) | 50.786 | 1.041 W | 15 | 2.9 | 0.481 (±0.292) | 0.464 (±0.252) | -0.041 | 0.027 | 0.237 | 0 | 0 | 1 | 0 | 0 |
| 14 | UK - Sa | Littlehampton (West Sussex) | 50.801 | 0.558 W | 17 | 1.8 | 0.277 (±0.224) | 0.335 (±0.227) | 0.173* | 0.050 | 0 | 0 | 0 | 1 | 0 | 0 |
| 15 | UK - Sa | Dungeness (Kent) | 50.933 | 0.959 E | 16 | 3.8 | 0.580 (±0.225) | 0.657 (±0.156) | 0.120** | 0.037 | 0.044 | 0 | 0 | 1 | 0 | 0 |
| 16 | UK – Sm | Folkestone (Kent) | 51.102 | 1.236 E | 17 | 4.1 | 0.502 (±0.238) | 0.631 (±0.214) | 0.209** | 0.150 | 0.124 | 0.941 | 0.059 | 0 | 0 | 0 |
| 17 | France | Ghyvelde (Nord) | 51.056 | 2.522 E | 40 | 2.9 | 0.350 (±0.254) | 0.508 (±0.219) | 0.298** | 0.018 | 0 | 1 | 0 | 0 | 0 | 0 |
| 18 | France | La Slack (Boulonnais) | 50.806 | 1.604 E | 4 | - | - | - | 0.098 | - | - | 0 | 0 | 1 | 0 | 0 |
| 19 | France | Ambleteuse (Boulonnais) | 50.808 | 1.628 E | 19 | 4.0 | 0.632 (±0.197) | 0.660 (±0.157) | 0.044 | 0.015 | 0 | 0 | 0 | 1 | 0 | 0 |
| 20 | France | Wimereux (Boulonnais) | 50.790 | 1.622 E | 11 | 3.4 | 0.664 (±0.293) | 0.615 (±0.237) | -0.084 | 0.011 | 0.023 | 0 | 0 | 1 | 0 | 0 |
| 21 | France | Ecault (Boulonnais) | 50.661 | 1.581 E | 133 | 3.1 | 0.509 (±0.196) | 0.583 (±0.178) | 0.126** | 0.056 | 0.082 | 0 | 0 | 1 | 0 | 0 |
| 22 | France | Val St-Martin (Normandie) | 49.257 | 1.379 E | 13 | 4.3 | 0.587 (±0.223) | 0.670 (±0.212) | 0.119** | 0.067 | 0.073 | 0 | 0 | 1 | 0 | 0 |
| 23 | France | Ménil-Villement (Normandie) | 48.855 | 0.394 W | 13 | 4.3 | 0.627 (±0.278) | 0.711 (±0.129) | 0.104** | 0.020 | 0.033 | 0 | 0 | 1 | 0 | 0 |
| 24 | France | Thury-Harcourt (Normandie) | 48.997 | 0.489 W | 12 | 4.0 | 0.577 (±0.275) | 0.701 (±0.148) | 0.150** | 0.029 | 0 | 0 | 0 | 1 | 0 | 0 |
| 25 | France | Arzal (Bretagne) | 47.506 | 2.404 W | 12 | 4.2 | 0.597 (±0.248) | 0.667 (±0.199) | 0.104* | 0.023 | 0.087 | 0 | 0 | 1 | 0 | 0 |
| 26 | France | Moëlan-sur-mer (Bretagne) | 47.782 | 3.663 W | 13 | 3.7 | 0.533 (±0.269) | 0.627 (±0.154) | 0.162** | 0.052 | 0 | 0 | 0 | 1 | 0 | 0 |
| 27 | France | Saint-Cast-le-Guildo | 48.611 | 2.237 W | 16 | 3.6 | 0.455 (±0.239) | 0.620 (±0.185) | 0.264** | 0.150 | 0.259 | 0 | 0 | 1 | 0 | 0 |
| 28 | France | Binic (Bretagne) | 48.610 | 2.815 W | 16 | 3.8 | 0.549 (±0.355) | 0.654 (±0.222) | 0.148** | 0.012 | 0 | 0 | 0 | 1 | 0 | 0 |
| 29 | France | Planguenoual (Bretagne) | 48.572 | 2.580 W | 15 | 3.9 | 0.590 (±0.274) | 0.681 (±0.131) | 0.124** | 0.035 | 0 | 0 | 0 | 1 | 0 | 0 |
| 30 | France | Néant-sur-Yvel (Bretagne) | 48.045 | 2.318 W | 14 | 3.0 | 0.535 (±0.343) | 0.537 (±0.280) | -0.004 | 0.035 | 0.070 | 0 | 0 | 1 | 0 | 0 |
| 31 | France | Château-Thébaud (Bretagne) | 47.126 | 1.413 W | 13 | 3.7 | 0.526 (±0.244) | 0.641 (±0.182) | 0.179** | 0.071 | 0.082 | 0 | 0 | 1 | 0 | 0 |
| 32 | France | Champtoceaux (Bretagne) | 47.335 | 1.287 W | 13 | 3.6 | 0.540 (±0.300) | 0.605 (±0.220) | 0.104** | 0.045 | 0.054 | 0 | 0 | 1 | 0 | 0 |

-: not applicable because of too small sample size.
Plastid diversity is provided as haplotype frequencies found in each population.
Morphological varieties in Great Britain are indicated as "UK – Sm" for *Silene nutans* var. *smitiana* and "UK – Sa" for var. *Salmoniana* and Belgium ecotypes as "Belgium – Ca" for the calcicolous ecotype and "Belgium – Si" for the silicicolous ecotype.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



Population details and mean estimates of genetic variation for 111 populations of *Silene nutans* from western Europe. (continued)

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 33 | France | Saint Vigor (Normandie) | 49.490 | 0.379 E | 23 | 4.0 | 0.589 (±0.245) | 0.696 (±0.186) | 0.153** | 0.017 | 0 | 0 | 0 | 1 | 0 | 0 |
| 34 | France | Saint Jouin (Normandie) | 49.643 | 0.152 E | 21 | 3.3 | 0.448 (±0.200) | 0.603 (±0.214) | 0.261** | 0.211 | 0.228 | 0 | 0 | 1 | 0 | 0 |
| 35 | France | Saint Benoit (Poitou) | 46.542 | 0.341 E | 22 | 5.3 | 0.654 (±0.215) | 0.781 (±0.139) | 0.165** | 0.017 | 0 | 0 | 0 | 0 | 1 | 0 |
| 36 | France | Saint Benoit (Poitou) | 46.548 | 0.352 E | 19 | 5.1 | 0.541 (±0.254) | 0.754 (±0.212) | 0.288** | 0.058 | 0 | 0 | 0 | 0 | 1 | 0 |
| 37 | France | Ligugé (Poitou) | 46.523 | 0.334 E | 16 | 5.0 | 0.644 (±0.300) | 0.747 (±0.224) | 0.142** | 0.018 | 0 | 0 | 0 | 0 | 1 | 0 |
| 38 | France | Puiguilhem (Dordogne) | 45.424 | 0.748 E | 15 | 4.8 | 0.575 (±0.226) | 0.722 (±0.214) | 0.205** | 0.068 | 0 | 0 | 0 | 0 | 1 | 0 |
| 39 | France | Bourdeilles (Dordogne) | 45.355 | 0.628 E | 36 | 4.4 | 0.555 (±0.244) | 0.708 (±0.207) | 0.217** | 0.072 | 0.106 | 0 | 0 | 0 | 1 | 0 |
| 40 | France | Peyresourde (Pyrénées) | 42.802 | 0.453 E | 20 | 5.8 | 0.710 (±0.205) | 0.835 (±0.093) | 0.142** | 0.021 | 0.030 | 0 | 0 | 0 | 1 | 0 |
| 41 | France | Aranvielle (Pyrénées) | 42.810 | 0.409 E | 12 | 5.5 | 0.681 (±0.243) | 0.804 (±0.135) | 0.158** | 0.031 | 0.089 | 0 | 0 | 0 | 1 | 0 |
| 42 | France | Cauterets (Pyrénées) | 42.854 | 0.098 W | 8 | 4.5 | 0.609 (±0.248) | 0.752 (±0.141) | 0.188** | 0.035 | 0 | 0 | 0 | 0 | 1 | 0 |
| 43 | France | Arrens (Pyrénées) | 42.901 | 0.265 W | 10 | 5.5 | 0.581 (±0.236) | 0.824 (±0.087) | 0.294** | 0.116 | 0.214 | 0 | 0 | 0 | 1 | 0 |
| 44 | France | Ourey (Pyrénées) | 42.931 | 0.243 W | 10 | 5.3 | 0.599 (±0.272) | 0.793 (±0.134) | 0.236** | 0.049 | 0.049 | 0 | 0 | 0 | 1 | 0 |
| 45 | France | Sarrance (Pyrénées) | 43.041 | 0.602 W | 21 | 5.3 | 0.622 (±0.356) | 0.826 (±0.073) | 0.219** | 0.008 | 0 | 0 | 0 | 0 | 1 | 0 |
| 46 | France | Rouvières (Cévennes) | 44.337 | 3.814 E | 32 | 5.2 | 0.626 (±0.270) | 0.746 (±0.222) | 0.161** | 0.024 | 0.036 | 0 | 0 | 1 | 0 | 0 |
| 47 | France | Felgérolles (Cévennes) | 44.356 | 3.799 E | 20 | 5.2 | 0.670 (±0.248) | 0.732 (±0.236) | 0.086* | 0.016 | 0 | 0 | 0 | 1 | 0 | 0 |
| 48 | France | Runes (Cévennes) | 44.382 | 3.677 E | 29 | 5.4 | 0.678 (±0.238) | 0.766 (±0.228) | 0.117** | 0.050 | 0.038 | 0 | 0 | 1 | 0 | 0 |
| 49 | France | St Frézal (Cévennes) | 44.297 | 3.889 E | 19 | 5.2 | 0.657 (±0.232) | 0.746 (±0.220) | 0.117** | 0.041 | 0.120 | 0 | 0 | 1 | 0 | 0 |
| 50 | France | Le Cros (Cévennes) | 44.311 | 3.895 E | 13 | 4.7 | 0.659 (±0.294) | 0.677 (±0.293) | 0.028 | 0.016 | 0.116 | 0 | 0 | 1 | 0 | 0 |
| 51 | France | Murol (Auvergne) | 45.578 | 2.940 E | 27 | 4.7 | 0.757 (±0.168) | 0.753 (±0.138) | -0.006 | 0.013 | 0.014 | 0 | 0 | 1 | 0 | 0 |
| 52 | France | Morand (Auvergne) | 45.590 | 2.875 E | 6 | - | 0.731 (±0.293) | 0.697 (±0.147) | -0.054 | 0.019 | 0.147 | 0 | 0 | 1 | 0 | 0 |
| 53 | France | Pierrefort (Auvergne) | 44.952 | 2.940 E | 22 | 5.0 | 0.643 (±0.213) | 0.769 (±0.155) | 0.164** | 0.021 | 0.003 | 0 | 0 | 1 | 0 | 0 |
| 54 | France | Devèze (Auvergne) | 44.862 | 2.764 E | 20 | 4.5 | 0.612 (±0.244) | 0.721 (±0.163) | 0.144** | 0.045 | 0.047 | 0 | 0 | 1 | 0 | 0 |
| 55 | France | Ceillac (Queyras) | 44.631 | 6.818 E | 27 | 5.7 | 0.519 (±0.309) | 0.706 (±0.286) | 0.248** | 0.016 | 0.005 | 0 | 0 | 0 | 0 | 1 |
| 56 | France | Valpereyre (Queyras) | 44.814 | 6.975 E | 17 | 5.2 | 0.541 (±0.304) | 0.751 (±0.183) | 0.224** | 0.020 | 0.053 | 0 | 0 | 0 | 0 | 1 |
| 57 | France | Arvieux (Queyras) | 44.778 | 6.740 E | 19 | 5.4 | 0.527 (±0.334) | 0.732 (±0.244) | 0.232** | 0.014 | 0.080 | 0 | 0 | 0 | 0 | 1 |
| 58 | France | Aussois (Savoie) | 45.238 | 6.760 E | 32 | 5.1 | 0.478 (±0.247) | 0.734 (±0.185) | 0.341** | 0.216 | 0.190 | 0 | 0 | 0.241 | 0 | 0.759 |
| 59 | France | Mirebel (Jura) | 46.700 | 5.734 E | 20 | 4.1 | 0.435 (±0.197) | 0.684 (±0.153) | 0.351** | 0.153 | 0 | 0 | 0 | 0 | 0 | 1 |
| 60 | France | Serre forest (Jura) | 47.167 | 5.557 E | 19 | 4.2 | 0.511 (±0.216) | 0.693 (±0.149) | 0.262** | 0.096 | 0.012 | 0 | 0 | 1 | 0 | 0 |
| 61 | France | Les Granges (Jura) | 47.138 | 5.967 E | 24 | 4.7 | 0.527 (±0.242) | 0.730 (±0.179) | 0.260** | 0.033 | 0 | 0 | 0 | 0 | 0 | 1 |
| 62 | France | Savigny-les-Beaune (Bourgogne) | 47.093 | 4.806 E | 12 | 3.9 | 0.654 (±0.254) | 0.678 (±0.222) | 0.038 | 0.018 | 0 | 0 | 0 | 1 | 0 | 0 |
| 63 | France | Plombières-les-Dijon | 47.332 | 4.936 E | 12 | 5.0 | 0.623 (±0.292) | 0.674 (±0.252) | 0.077* | 0.024 | 0 | 1 | 0 | 0 | 0 | 0 |
| 64 | France | Corcelles-les-Monts (Bourgogne) | 47.311 | 4.928 E | 10 | 5.2 | 0.692 (±0.193) | 0.726 (±0.197) | 0.049* | 0.055 | 0.080 | 1 | 0 | 0 | 0 | 0 |
| 65 | France | Epagny (Bourgogne) | 47.448 | 5.043 E | 12 | 4.9 | 0.599 (±0.245) | 0.735 (±0.165) | 0.160** | 0.034 | 0.134 | 0.889 | 0.111 | 0 | 0 | 0 |
| 66 | France | Jaulny (Lorraine) | 48.966 | 5.894 E | 25 | 4.7 | 0.588 (±0.185) | 0.691 (±0.182) | 0.149** | 0.127 | 0.088 | 1 | 0 | 0 | 0 | 0 |
| 67 | Luxemburg | Dirbach (Oesling) | 49.921 | 6.038 E | 10 | 4.6 | 0.631 (±0.246) | 0.703 (±0.233) | 0.113** | 0.054 | 0.039 | 0.875 | 0.125 | 0 | 0 | 0 |

-: not applicable because of too small sample size.
Plastid diversity is provided as haplotype frequencies found in each population.
Morphological varieties in Great Britain are indicated as "UK – Sm" for *Silene nutans* var. *smitiana* and "UK – Sa" for var. *Salmoniana* and Belgium ecotypes as "Belgium – Ca" for the calcicolous ecotype and "Belgium – Si" for the silicicolous ecotype.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Population details and mean estimates of genetic variation for 111 populations of *Silene nutans* from western Europe. (continued)

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 68 | Luxemburg | Bavigne (Oesling) | 49.922 | 5.845 E | 11 | 4.7 | 0.649 (±0.197) | 0.697 (±0.193) | 0.076** | 0.042 | 0 | 1 | 0 | 0 | 0 | 0 |
| 69 | Luxemburg | Dudelange (Gutland) | 49.475 | 6.061 E | 7 | - | 0.615 (±0.337) | 0.509 (±0.249) | -0.231 | 0.012 | 0 | 1 | 0 | 0 | 0 | 0 |
| 70 | Luxemburg | Dudelange (Gutland) | 49.482 | 6.069 E | 12 | 3.6 | 0.506 (±0.248) | 0.544 (±0.231) | 0.070 | 0.022 | 0.116 | 1 | 0 | 0 | 0 | 0 |
| 71 | Luxemburg | Oberkorn (Gutland) | 49.547 | 5.878 E | 12 | 3.0 | 0.513 (±0.223) | 0.526 (±0.159) | 0.026 | 0.041 | 0.188 | 1 | 0 | 0 | 0 | 0 |
| 72 | Luxemburg | Brandenbourg (Oesling) | 49.911 | 6.140 E | 12 | 4.4 | 0.517 (±0.230) | 0.683 (±0.194) | 0.253** | 0.233 | 0.379 | 1 | 0 | 0 | 0 | 0 |
| 73 | Luxemburg | Schlinder (Oesling) | 49.936 | 6.068 E | 5 | - | 0.481 (±0.376) | 0.530 (±0.321) | 0.105** | 0.039 | 0 | 1 | 0 | 0 | 0 | 0 |
| 74 | Luxemburg | Marienthal (Gutland) | 49.711 | 6.060 E | 9 | 4.7 | 0.641 (±0.257) | 0.671 (±0.209) | 0.049 | 0.042 | 0.129 | 1 | 0 | 0 | 0 | 0 |
| 75 | Luxemburg | Ansembourg (Gutland) | 49.697 | 6.034 E | 12 | 4.8 | 0.633 (±0.224) | 0.705 (±0.198) | 0.106* | 0.061 | 0.087 | 1 | 0 | 0 | 0 | 0 |
| 76 | Luxemburg | Steinfort (Gutland) | 49.666 | 5.905 E | 12 | 4.1 | 0.629 (±0.239) | 0.633 (±0.227) | 0.003 | 0.018 | 0 | 0.917 | 0.083 | 0 | 0 | 0 |
| 77 | Belgium - Ca | Nismes (Namur) | 50.069 | 4.561 E | 19 | 3.8 | 0.590 (±0.181) | 0.639 (±0.139) | 0.076** | 0.025 | 0.063 | 1 | 0 | 0 | 0 | 0 |
| 78 | Belgium - Ca | Nismes (Namur) | 50.086 | 4.566 E | 28 | 4.7 | 0.550 (±0.200) | 0.698 (±0.210) | 0.215** | 0.150 | 0.134 | 1 | 0 | 0 | 0 | 0 |
| 79 | Belgium - Si | Houyet (Namur) | 50.186 | 5.008 E | 24 | 4.5 | 0.617 (±0.253) | 0.712 (±0.191) | 0.126** | 0.013 | 0.039 | 0 | 0 | 1 | 0 | 0 |
| 80 | Belgium - Ca | Dinant (Namur) | 50.271 | 4.918 E | 29 | 4.6 | 0.581 (±0.233) | 0.707 (±0.157) | 0.164** | 0.029 | 0 | 1 | 0 | 0 | 0 | 0 |
| 81 | Belgium - Si | Vodelée (Namur) | 50.171 | 4.726 E | 16 | 4.3 | 0.582 (±0.264) | 0.717 (±0.168) | 0.182** | 0.020 | 0.091 | 0 | 0 | 1 | 0 | 0 |
| 82 | Belgium - Si | Olloy-sur-Viroin (Namur) | 50.069 | 4.606 E | 96 | 4.6 | 0.598 (±0.166) | 0.739 (±0.106) | 0.188** | 0.023 | 0 | 0 | 0 | 1 | 0 | 0 |
| 83 | Belgium | Knokke (West-Vlaanderen) | 51.346 | 3.295 E | 15 | 4.3 | 0.490 (±0.172) | 0.700 (±0.118) | 0.308** | 0.152 | 0.031 | 1 | 0 | 0 | 0 | 0 |
| 84 | Belgium | Middelkerke (West-Vlaanderen) | 51.173 | 2.786 E | 12 | 3.6 | 0.539 (±0.253) | 0.619 (±0.175) | 0.120** | 0.068 | 0 | 0.909 | 0.091 | 0 | 0 | 0 |
| 85 | Belgium | De Panne (West-Vlaanderen) | 51.092 | 2.557 E | 2 | - | - | - | 0.514** | - | - | 1 | 0 | 0 | 0 | 0 |
| 86 | Belgium - Si | Hamoir (Liège) | 50.446 | 5.537 E | 23 | 3.9 | 0.499 (±0.212) | 0.645 (±0.188) | 0.222** | 0.031 | 0 | 0 | 0 | 1 | 0 | 0 |
| 87 | Belgium - Ca | Bomal-sur-Ourthe (Luxembourg) | 50.363 | 5.519 E | 28 | 4.7 | 0.626 (±0.253) | 0.671 (±0.251) | 0.069** | 0.033 | 0.073 | 1 | 0 | 0 | 0 | 0 |
| 88 | Belgium - Ca | Chokier (Liège) | 50.594 | 5.445 E | 18 | 3.9 | 0.560 (±0.231) | 0.645 (±0.164) | 0.132** | 0.035 | 0 | 1 | 0 | 0 | 0 | 0 |
| 89 | Belgium - Si | Nonceveux (Liège) | 50.472 | 5.733 E | 15 | 3.1 | 0.441 (±0.258) | 0.559 (±0.229) | 0.214** | 0.063 | 0.069 | 0 | 0 | 1 | 0 | 0 |
| 90 | The Netherlands | Katwijk aan Zee (Zuid-Holland) | 52.195 | 4.390 E | 11 | 4.5 | 0.480 (±0.315) | 0.660 (±0.245) | 0.268** | 0.057 | 0.207 | 1 | 0 | 0 | 0 | 0 |
| 91 | The Netherlands | Ijmuiden (Noord-Holland) | 52.457 | 4.569 E | 17 | 5.0 | 0.525 (±0.186) | 0.727 (±0.181) | 0.288** | 0.127 | 0.051 | 1 | 0 | 0 | 0 | 0 |
| 92 | The Netherlands | Egmond aan Zee (Noord-Holland) | 52.624 | 4.636 E | 11 | 4.8 | 0.587 (±0.181) | 0.693 (±0.205) | 0.155** | 0.041 | 0 | 1 | 0 | 0 | 0 | 0 |
| 93 | Germany | Klocksin (Müritz) | 53.647 | 12.522 E | 15 | 5.2 | 0.723 (±0.163) | 0.731 (±0.178) | 0.011 | 0.013 | 0 | 1 | 0 | 0 | 0 | 0 |
| 94 | Germany | Putbus (Rügen) | 54.341 | 13.521 E | 15 | 4.8 | 0.567 (±0.222) | 0.694 (±0.196) | 0.180** | 0.075 | 0.056 | 0.933 | 0.067 | 0 | 0 | 0 |
| 95 | Germany | Schöneberg (Uckermark) | 52.975 | 14.093 E | 15 | 4.1 | 0.651 (±0.263) | 0.634 (±0.216) | -0.028 | 0.015 | 0.017 | 1 | 0 | 0 | 0 | 0 |
| 96 | Germany | Neugernsdorf (Greiz) | 50.713 | 12.142 E | 13 | 4.0 | 0.641 (±0.296) | 0.638 (±0.239) | -0.010 | 0.009 | 0.009 | 1 | 0 | 0 | 0 | 0 |
| 97 | Germany | Hohenölsen (Greiz) | 50.732 | 12.060 E | 15 | 5.0 | 0.670 (±0.208) | 0.716 (±0.157) | 0.067* | 0.037 | 0.143 | 1 | 0 | 0 | 0 | 0 |
| 98 | Germany | Hellingen (Hildburghausen) | 50.268 | 10.702 E | 15 | 4.6 | 0.605 (±0.266) | 0.680 (±0.254) | 0.109** | 0.038 | 0.060 | 1 | 0 | 0 | 0 | 0 |
| 99 | Germany | Schwarzach (Schwandorf) | 49.415 | 12.228 E | 15 | 5.4 | 0.724 (±0.167) | 0.760 (±0.146) | 0.049* | 0.026 | 0.042 | 1 | 0 | 0 | 0 | 0 |
| 100 | Germany | Sinzing (Regensburg) | 48.971 | 12.018 E | 15 | 5.4 | 0.635 (±0.259) | 0.736 (±0.220) | 0.139** | 0.069 | 0.074 | 1 | 0 | 0 | 0 | 0 |
| 101 | Germany | Waidhofen (Neuburg- | 48.602 | 11.326 E | 14 | 5.4 | 0.736 (±0.150) | 0.759 (±0.155) | 0.031* | 0.036 | 0.045 | 1 | 0 | 0 | 0 | 0 |
| 102 | Germany | Mülheim (Tuttlingen) | 48.017 | 8.861 E | 15 | 4.8 | 0.496 (±0.274) | 0.727 (±0.255) | 0.312** | 0.063 | 0 | 0 | 0 | 0 | 0 | 1 |

-: not applicable because of too small sample size.
Plastid diversity is provided as haplotype frequencies found in each population.
Morphological varieties in Great Britain are indicated as "UK – Sm" for *Silene nutans* var. *smitiana* and "UK – Sa" for var. *Salmoniana* and Belgium ecotypes as "Belgium – Ca" for the calcicolous ecotype and "Belgium – Si" for the silicicolous ecotype.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



Population details and mean estimates of genetic variation for 111 populations of *Silene nutans* from western Europe. - Fin

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 103 | Germany | Hardt (Rottweil) | 48.184 | 8.374 E | 14 | 4.3 | 0.526 (±0.269) | 0.679 (±0.176) | 0.229** | 0.060 | 0 | 0 | 0 | 0 | 0 | 1 |
| 104 | Germany | Blieskastel (Saar-Pfalz) | 49.259 | 7.309 E | 15 | 4.3 | 0.567 (±0.216) | 0.633 (±0.242) | 0.111** | 0.075 | 0.014 | 1 | 0 | 0 | 0 | 0 |
| 105 | Germany | Saarbrücken (Saarbrücken) | 49.228 | 7.016 E | 15 | 5.3 | 0.662 (±0.300) | 0.738 (±0.217) | 0.106** | 0.018 | 0.013 | 1 | 0 | 0 | 0 | 0 |
| 106 | Germany | Schmelz (Saarlouis) | 49.464 | 6.851 E | 20 | 4.7 | 0.584 (±0.205) | 0.645 (±0.227) | 0.098 | 0.047 | 0.036 | 1 | 0 | 0 | 0 | 0 |
| 107 | Germany | Nonnweiler (Sankt Wendel) | 49.548 | 6.998 E | 5 | - | 0.677 (±0.277) | 0.721 (±0.216) | 0.069 | 0.023 | 0 | 1 | 0 | 0 | 0 | 0 |
| 108 | Germany | Loffenau (Rastatt) | 48.772 | 8.378 E | 11 | 3.7 | 0.488 (±0.228) | 0.596 (±0.233) | 0.185** | 0.116 | 0.094 | 1 | 0 | 0 | 0 | 0 |
| 109 | Germany | Zaberfeld (Heilbronn) | 49.019 | 8.937 E | 15 | 5.0 | 0.551 (±0.246) | 0.749 (±0.146) | 0.264** | 0.054 | 0.118 | 1 | 0 | 0 | 0 | 0 |
| 110 | Germany | Neckarzimmern (Neckar-Odenwald) | 49.323 | 9.129 E | 15 | 3.8 | 0.536 (±0.309) | 0.560 (±0.280) | 0.035* | 0.014 | 0 | 1 | 0 | 0 | 0 | 0 |
| 111 | Germany | Kerpen (Daun) | 50.313 | 6.760 E | 14 | 2.4 | 0.505 (±0.190) | 0.446 (±0.162) | -0.140 | 0.013 | 0.096 | 1 | 0 | 0 | 0 | 0 |

-: not applicable because of too small sample size.
Plastid diversity is provided as haplotype frequencies found in each population.
Morphological varieties in Great Britain are indicated as “UK – Sm” for *Silene nutans* var. *smitiana* and “UK – Sa” for var. *Salmoniana* and Belgium ecotypes as “Belgium – Ca” for the calcicolous ecotype and “Belgium – Si” for the silicicolous ecotype.


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

© 2016 Tous droits réservés.



**Table S2.2:** Sequences flanking the plastid SNPs that were used in *Silene nutans* to define KASP$^{tm}$ specific primers by LGC group. The SNPs are indicated within brackets.

| SNP | Flanking sequence |
|---|---|
| Cp 42 | TACATCCGCCCCCTTTTTCTAATTTCAACAACTTCTTAAAT[T/G]AAAGAAAAAGAAAAGACGAAYATTATACCCACCATGAACAGACGTATGAT |
| Cp 397 | ATGAGTATCATAATAATAATTGGAATAGTTTTTTTATTCCAAAAAACTCT[A/C]TTTCTTTTTTTTTTTTTTAGAAATCAAAGATTATTCGTGTTCTTATATAAT |
| Cp 540 | AACCAATCCTCTCATTTACGATCAACATCTTATAGAGCCCTTCTTGAACG[C/T]ACTTTTTTCTACGGAAAATTAGAATATTTAGTAAAACTTTTTACTAAGGA |
| Cp 656 | ATGGCTTTTCAAGGATCCTTCCCCGCATTCTGTTAGGTATAAGGGAAAAT[G/T]CATTCTGGCCTCAAAAGGGACATTCTTTTTGATGCATAAATGGAAATTTT |
| Cp 730 | TCTTTTTGATGCATAAATGGAAATTTTACCTTATTCATTTCTGGCAATGT[C/T]ATTTTTCTGTGTGGTCTCACCCAAGAAGAATCTATATTAATCGATTATCA |
| Cp 804 | AGAAGAATCTATATTAATCGATTATCAACCCATTGTCTTGACTTTATTGG[T/G]TTTCTTTCAAGTGTACAACTCACTTCTTCAGTGGTACGGAGTCAAATGTT |

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

# **Chapitre 2** : Reproductive isolation between *Silene nutans* lineages

**Hélène Martin, Pascal Touzet, Mathilde Dufaÿ, Cécile Godé, Eric Schmitt, Emna Lahiani, Lynda F Delph et Fabienne Van Rossum**

Ce chapitre est en cours d'évaluation dans la revue *Evolution*.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



Dans le chapitre précédent, nous avons décrit des lignées génétiques qui, malgré en contact géographique, ne semblent pas échanger de flux de gènes. En Belgique, les deux écotypes de *S. nutans* sont fortement isolés par des barrières à la reproduction postzygotique. Dans ce chapitre nous avons testé la présence ce type de barrières dans les autres régions de contact géographique. Pour ce faire, nous avons profité de données issues de croisement intra- et interlignées avec des populations provenant de l'Angleterre, de la France et de la Belgique. Nous avons observé une importante mortalité au stade plantule (probablement liée à la chlorose observée dès la germination) chez les descendants des croisements interlignées par rapport aux descendants des croisements intralignées, que l'on croise des parents provenant de la même région ou de régions différentes. Nous avons donc conclu que l'isolement reproducteur postzygotique observé en Belgique est associé à la divergence en allopatrie des lignées *Ouest* et *Est*.

Les données phénotypiques présentées dans ce chapitre sont issues de croisements contrôlés réalisées pas Fabienne Van Rossum en 2011. Lors de mon stage de M1 (2012), j'ai planté et suivi la descendance pendant 5 semaines et j'ai réalisé les premières analyses statistiques. Lors de cette thèse, j'ai ré-analysé les résultats, j'ai complété les données génétiques dans le cadre du stage de L3 de Thomas Lesaffre et j'ai écrit l'article en incorporant les corrections et remarques des co-auteurs.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



# Lineages of *Silene nutans* developed rapid, strong, asymmetric postzygotic reproductive isolation in allopatry

## 3.1 Introduction

Reproductive isolation is a key element in speciation: by reducing gene flow, it delimits species (Coyne and Orr, 2004). Nevertheless, reproductive isolation may be incomplete, allowing gene flow even between well-defined species (Jacquemyn et al., 2012). In plants, hybridization between groups is prevented by multiple barriers that are often classified as being prezygotic or postzygotic (see reviews by Rieseberg and Willis, 2007, Baack et al., 2015). Prezygotic barriers to reproduction can occur prior to mating, as a result of pollinator specialisation or non-overlapping flowering times. It can also occur post-pollination, such as when incompatibility between pollen and stigmas prevents fertilization. Postzygotic barriers are sometimes evident soon after fertilization (e.g., embryo/seed abortion), but can act much later as well, leading to the production of hybrid progeny that suffer from reduced viability and/or fertility.

Although prezygotic isolating barriers are generally thought to be stronger than postzygotic barriers in plants (Lowry et al., 2008a), this will depend whether lineages are found in allopatry or in sympatry (Coyne and Orr, 2004). Under the scenario of allopatric speciation (geographic isolation), postzygotic barriers are expected to occur as a result of the accumulation of genetic incompatibilities over time (Coyne and Orr, 2004). Secondary contact between differentiated lineages can then reinforce the reproductive isolation process through pre-mating isolating barriers (Noor, 1999, Hopkins, 2013).

Because intrinsic genetic incompatibilities accumulate via selective or neutral processes over time (Presgraves, 2010, Lynch and Hagner, 2015), postzygotic reproductive isolation is expected to increase with increasing genetic distance among taxa (Orr, 1995, Moyle et al., 2004, Scopece et al., 2008). Genetic incompatibilities are thought to involve epistasis between two or more loci following the Bateson-Dobzhansky-Muller (BDM) model (Fraïsse et al., 2014) or a pathway-based model (Lindtke and Buerkle, 2015). When genetic incompatibilities involve an interaction between the nuclear and uniparentally inherited cytoplasmic genomes (termed "cytonuclear"), the reproductive isolation is expected to be asymmetric between reciprocal crosses (Turelli and Moyle, 2007).

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



We investigated intrinsic postzygotic barriers between phylogeographic lineages of the perennial herb *Silene nutans*. Discrete genetic lineages of this species have previously been identified based on nuclear microsatellite and plastid markers, in relation to past climatic events and post-glacial migration history (Martin et al., 2016, Van Rossum et al., 2016, Van Rossum, unpublished results). Lineages can be divided into two main genetically distant western and eastern groups. The western lineage, mainly located in western Europe, consists of several genetically differentiated sub-lineages. The most widespread sub-lineage (W1) occurs in England, France, and Belgium, whereas other sub-lineages are restricted to Spain and southwestern France (W2) or to the Alps and Italy (W3). The most widespread eastern lineage (E1) primarily occurs in Eastern Europe, but has spread from the east, northward and westward up into western Europe (Figure 3.1a) (e.g., Great Britain, France, and Belgium). The W1 and E1 lineages both occur in southern Belgium, where they grow in parapatry as separate ecotypes, as a result of edaphic evolution to different soils (De Bilde, 1973, Van Rossum et al., 1999). Previous studies have shown that crosses between these two ecotypes from Belgium result in nearly complete postzygotic reproductive isolation (PRI), expressed as seedling chlorosis and reduced hybrid fitness (Van Rossum et al., 1996).

Given this PRI between ecotypes in Belgium, we investigated whether PRI between these lineages exists on a broader scale or is localized. While only one lineage exists in most areas of Europe, two or more lineages co-exist in southern England and eastern France and there is no evidence of ecotypic differentiation to soil types in these areas (Van Rossum, unpublished results). Hence, two possibilities exist. First, any PRI that occurs between the western and eastern lineages does so because the lineages diverged in allopatry prior to the spread of the eastern lineage. Second, the PRI occurs because of differentiation to local conditions following secondary contact, in a speciation-with-gene-flow process (Smadja and Butlin, 2011). Hence, we carried out a crossing experiment both within and between regions to investigate whether PRI results from allopatric vs. local processes. Under the first scenario, we predict that PRI between the lineages will occur regardless of their geographical provenance. If PRI is mainly a consequence of the accumulation of genetic incompatibilities since the separation of the western and eastern lineages, we predicted similar degrees of PRI between them regardless of whether the cross was within or between regions. In addition, we expected less PRI between western sub-lineages because of their more recent divergence. Under the second scenario, PRI is mainly a consequence of local conditions, and within-region crosses should exhibit stronger PRI than between region crosses. In addition, the PRI should vary between the western and eastern lineages depending on region.


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



PRI was estimated from hybrid progeny, ranging from the proportion of ovules fertilized to the mortality of five-week old seedlings, and the level of chlorosis and vegetative growth. In parallel, we investigated the occurrence of hybrids between lineages in regions where they co-occur using population genetic analysis of nuclear and plastid markers to estimate net reproductive isolation, i. e., the overall effect of prezygotic and postzygotic barriers, between lineages in nature.

## 3.2 Material and Methods

### 3.2.1 Study species and sampled populations

*Silene nutans* L. (Caryophyllaceae) is a long-lived perennial plant species, occurring in dry habitats, on rock outcrops, sand, or shingle, with a continental distribution, extending from western and northern Europe to central Siberia and the Southern Caucasus (Hepper, 1951, Fitter, 1978). Cytonuclear gynodioecy occurs in natural population leading to the coexistence of hermaphrodite and female (Dufay et al., 2010, Garraud et al., 2011); in this study only hermaphrodite plants were used for the crosses. Flowers open at dusk and are protandrous, with one whorl of anthers dehiscing on the first evening of opening, followed by a second whorl the following evening. Stigmas become receptive on the third evening, and remain receptive for about two days (Hepper, 1956). Several flowers are often simultaneously open on a plant, allowing geitonogamous self-fertilization that results in inbreeding depression (Hauser and Siegismund, 2000, Dufay et al., 2010). Flowering occurs from mid-April to mid-July, with considerable among-population variation in the timing (De Bilde, 1973, Hauser and Weidema, 2000). *Silene nutans* is insect-pollinated, mainly by nocturnal moths, and by long-tongued diurnal bees (Jürgens et al., 1996).

Based on nuclear and plastid sequences, *S. nutans* is split into genetically differentiated lineages in Europe, some of which co-occur in southern England, eastern France, and southern Belgium (Martin et al., 2016, Van Rossum et al., 2016, Van Rossum unpublised results). As described above, the western lineage is composed of three sub-lineages, including the W1 lineage (found in southern England, eastern France, and Belgium) and the W3 lineage (found in eastern France). Several lineages occur in Eastern Europe, but only the E1 lineage is found in the regions of focus (Martin et al., 2016, Van Rossum unpublised results). Living rosettes were collected from 176 individuals growing in 16 natural populations in four regions containing multiple lineages (Figure 3.1a and in appendices Table S3.1), and grown to flowering in a greenhouse for controlled

© 2016 Tous droits réservés.

 lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



crosses. Leaf material from 12 to 96 individuals per population was also harvested from the field, dried and kept in silica gel for further population genetic analyses (see Martin et al., 2016, for details).

### 3.2.2 Crossing experiment: design

In order to detect PRI between lineages, we conducted several types of reciprocal crosses within and between lineages within regions. In addition, when possible, crosses between populations located either in geographic vicinity (within regions) or in distinct regions (between regions) were compared to detect whether a local regional effect on the degree of PRI existed. The study mainly focused on the W1 and E1 lineages, with some additional crosses involving the W3 lineage, which is only present in eastern France (Jura) (Table S3.2 in appendices).

First, within three regions where western W1 and eastern E1 lineages co-occur (England, eastern France [Bourgogne], and southern Belgium), we performed four cross types – W1 x W1 and E1 x E1, W1 x E1 and E1 x W1 (with maternal lineage listed first for each cross type). These four cross types were also performed between regions (for instance, an E1 Belgian dam crossed with an E1 British sire, see Table S3.2 in appendices). As flowering periods in England and eastern France did not overlap, between-region crosses were only possible with individuals from Belgium. Depending on material availability, all cross combinations between regions were not possible.

Next, we examined the degree of PRI between the W3 and E1 lineages compared to the one observed between the W1 and E1 lineages. As the W3 and E1 lineages do not co-occur in the same region and the W3 lineage was only found in eastern France (Jura), crosses were performed between both the W1 and W3 lineages from Jura with the E1 lineage from Belgium. We performed four inter-lineage cross types: W3 x E1, E1 x W3, W1 x E1 and E1 x W1 (Table S3.2 in appendices).

Finally, we evaluated the degree of PRI between the two closely related western lineages, the W1 and W3 lineages, which co-occur in Eastern France (Jura). Within the Jura region, we performed four cross types: two intra-lineage crosses, W1 x W1 and W3 x W3, and two inter-lineage crosses, W1 x W3 and its reciprocal W3 x W1 (Table S3.2 in appendices).


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



### 3.2.3 Crossing experiment: procedures and fitness measurements

Pollinations for crosses were performed from mid-April to June 2011 on a total of 700 flowers. These flowers were emasculated just before opening (prior to dehiscence of any stamens) and bagged to avoid pollen contamination. Once stigmas were receptive (three days after emasculation), they were hand-pollinated early in the morning on two consecutive days using freshly dehiscent anthers from three pollen donors and re-bagged.

A total of 295 mature fruits were collected just before dehiscence, when seeds were ripe (about one month after pollination). Some pollinated flowers did not develop into a fruit. This can be caused by prezygotic reproductive barriers (Van Rossum et al., 1996), but can also be caused by ovary damage during emasculation. The number of unfertilized ovules, aborted seeds, and viable seeds were counted and summed to estimate the total number of ovules in each fruit. The proportion of ovules fertilized per fruit was calculated as the sum of aborted and viable seeds over the total number of ovules, while the proportion of seeds aborted per fruit was calculated as the number of aborted seeds over the sum of aborted and viable seeds.

In March 2012, up to 50 seeds per fruit were germinated in two Petri dishes on moist filter paper (Whatman) in controlled conditions (20°C, 16 hour day length) for a total of 243 fruits. The proportion of seeds that germinated after two weeks was calculated. The fitness of each seedling was characterized by its chlorosis level according to its color: yellow-white (chlorotic), light green (partially chlorotic), and green (healthy, non-chlorotic) (Figure S3.1 in appendices). Two to 14 (mostly 11) randomly selected seedlings per fruit were transplanted individually into pots containing a soil mixture (3/4 compost and 1/4 perlite) and their growth was followed for five weeks. Each week, the diameter of the rosette at the widest point was measured, and the number of leaves was counted for 2,587 individuals. The proportion of individuals that died within each family (i.e., from a single fruit) was recorded one week after transplantation and at the end of the five-week period.

### 3.2.4 Crossing experiment: statistical analyses

The proportion variables, i.e., ovules fertilized, aborted seeds, germination, and mortality were analysed using a logistic regression analysis (binomial distribution and log link function) using lme4 R package (Bates et al. 2015). For seedling chlorosis, a Chi-square test was performed. The number of leaves measured each week for five weeks was analysed using a multilevel mixed-effect model (Poisson distribution) using lme4 R package (Bates et al., 2015). Rosette diameter

© 2016 Tous droits réservés.



lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



measured each week for five weeks was analysed using a multilevel mixed-effect model (Gaussian distribution). Nevertheless, it should be noted that at each time step, the distribution of this variable did not follow a normal distribution (based on Shapiro-Wilks tests) and no transformation succeeded to meet normality. Hereafter, the results of the non-transformed variable are presented. Multiple comparisons of means using pairwise post-hoc Tukey contrast tests were performed when a significant effect was found.

To test for PRI between the W1 and E1 lineages with individuals originating from the same or different geographic regions, the models tested the effect of cross type (W1 x W1, W1 x E1, E1 x W1, and E1 x E1), regional level of the crosses (within vs. between regions), and their interaction as fixed effects. Maternal population nested within maternal region was tested as a random effect.

To compare the degree of PRI between W3 and E1 lineages with the degree of PRI between W1 and E1 lineages, the models tested cross type (E1 x W1, W1 x E1, E1 x W3, W3 x E1) as a fixed effect and maternal population as a random effect. The maternal region was not included in the analyses because there was only one maternal region in one level of cross type (Table S3.2 in appendices).

For crosses testing PRI between the two western lineages (W1 and W3) within the Jura region, models tested cross type (W3 x W3, W1 x W1, W3 x W1, W1 x W3) as a fixed effect and maternal population as a random effect. All statistical analyses were performed in the R environment (R Core Team Development, 2014).

Finally, following Lowry et al. (2008b,a) we determined the strength of the F1 survival barrier as RI = 1 - [(1 - mean inter-lineage progeny mortality)/(1 - mean intra-lineage progeny mortality)] for each reciprocal cross, using the proportion of seeds that had suffered mortality when measured five weeks after transplantation. The degree of asymmetry of PRI was calculated as the absolute value of the difference for the strength of RI between reciprocal crosses. The strength of the F1 survival barrier was estimated between W1 and E1 lineages and between the W1 and W3 lineages.

### 3.2.5 Molecular markers

DNA was extracted from 15-20 mg of dried leaf tissue using Macherey-Nagel (Düren, Germany) NucleoSpin® 96 Plant II kits following the standard protocol outlined in the manufacturer's handbook. An individual's assignment to one of three plastid groups was based on a combi-


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



nation of six plastid SNPs developed from the intergenic spacer sequences psbA-trnH and the matK gene fragment of *S. nutans* representative samples (Lahiani et al., 2013) and described in (Martin et al., 2016). To investigate the possibility of hybridization events in the regions of focus, we used a set of 24 microsatellites (Table S3.3 in appendices). Amplification procedures, multiplexing and genotyping were carried out following the standard protocols described in Godé et al. (2014). The populations of England, eastern France [Bourgogne and Jura] and four Belgian populations (HAM, LEFF, OLL and HER) were previously genotyped on the six plastid SNPs and 13 microsatellites (B09, E08, G01, H07, D10, Sil16, Sil19, Sil24, Sil31, Sil35, Sil36, Sil37 and Sil42) in Martin et al. (2016).

### 3.2.6 Population genetic variation and hybrid identification

The following measures of genetic variation were calculated within and over all populations: observed heterozygosity ($H_O$) and expected heterozygosity corrected for sample size ($H_E$) using GENECLASS2 (Piry et al., 2004), and the mean number of alleles (*An*) and allelic richness (*Ar*) based on a minimum sample size of eight individuals (El Mousadik and Petit, 1996) using FSTAT version 2.9.3.2 (Goudet, 1995). For each population, we estimated selfing rate (*s*) derived from the multilocus correlation structure of population samples, following a maximum likelihood based on the observed distribution of multilocus heterzygosity, described in David et al. (2007) and implemented in SPAGeDI version 1.5 (Hardy and Vekemans, 2002). Using GENEPOP version 4.3 (Rousset, 2008), deviations from Hardy–Weinberg equilibrium were also tested by estimating the intra-population fixation index ($F_{IS}$) following Weir and Cockerham (1984). Significance of $F_{IS}$ estimates was tested using exact probability tests across loci and populations. Markov chain method provided unbiased estimates of the Fisher's exact test probability using the following parameters: 100 000 dememorizations, 1 000 batches, and 100 000 iterations per batch. Over all populations, mean fixation indices ($F_{IS}$, $F_{IT}$, $F_{ST}$, following Weir and Cockerham, 1984) were estimated and their significance was tested with 20 000 permutations using SPAGeDI.

To infer population structure and hybrid detection within each region, a non-spatially explicit Bayesian clustering was carried out using STRUCTURE (Pritchard et al., 2000, Vähä and Primmer, 2006). The number of potential K clusters was assessed from 10 different runs of K ranging from 1 to 8. To ascertain adequate convergence of the Markov Chain Monte Carlo (MCMC) model, we allowed a burn-in of 100,000 iterations, followed by 2x106 MCMC replications without any prior geographic information on the putative affiliation of individuals. To identify the most likely number of K clusters, the ad hoc $\Delta K$ was calculated as described in

© 2016 Tous droits réservés.



lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Evanno et al. (2005). CLUMP version 1.1.2 (Jakobsson and Rosenberg, 2007) was used to find the optimal alignment of independent runs by averaging the top runs. The resulting most likely grouping of populations was plotted using DISTRUCT version 1.1 (Rosenberg, 2004). However, the assumption of panmixia implied in such Bayesian clustering does not necessarily hold in *S. nutans*, which can exhibit a mixed-mating system and female individuals (Lahiani et al., 2015). To avoid this potential bias, we performed a Principal Component Analysis (PCA) using adegenet R package (Jombart, 2008, R Core Team Development, 2014), which does not require any genetic assumptions, within each region.

## 3.3 Results

### 3.3.1 Can we detect PRI between the western and eastern lineages at any geographical scale?

Considering crosses between W1 and E1 lineages, a significantly higher proportion of seeds were aborted in the E1 x W1 crosses compared to the other cross types (Table 3.1a and see post-hoc Tukey tests in appendices Figure S3.2). The proportion of seeds that germinated also varied significantly among cross types, but post-hoc tests did not reveal a global significantly lower germination in hybrids compared to intra-lineage progeny (Table 3.1a, see post-hoc tests in appendices Figure S3.2). The effect of geographic distance (within vs. between regions) was significant for the proportion of ovules fertilized, with a significantly higher proportion in between-region crosses (Table 3.1a). A significant interaction effect between cross type and geographic distance was found for the proportion of ovules fertilized, seeds aborted, and seeds germinated (Table 3.1a). However, we did not found significantly higher hybrid fitness in between-region crosses compared to within-region crosses based on these traits (Figure S3.2in the appendices). In the first week of the survey, hybrid progeny suffered from high mortality: overall, an average 98% of E1 x W1 progeny and 42% of W1 x E1 progeny died, while most intra-lineage progeny survived, revealing a significant effect of cross type on mortality (Table 3.1a, Figure 3.1b). Although we found this same pattern within- and between-regions crosses (i.e., higher mortality when E1 was the maternal parent in a between-lineage cross), we found slightly higher mortality of W1 x E1 progeny in within-regions crosses compared to between-region crosses, leading to a significant interaction between cross type and geographical distance (Table 3.1a).

Mortality mirrored seedling chlorosis levels (Figure 3.1c). There was a significant association between the type of cross and the seedling chlorosis levels ($\chi^2$ (df = 6) = 4070.146, $P < 0.001$).


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

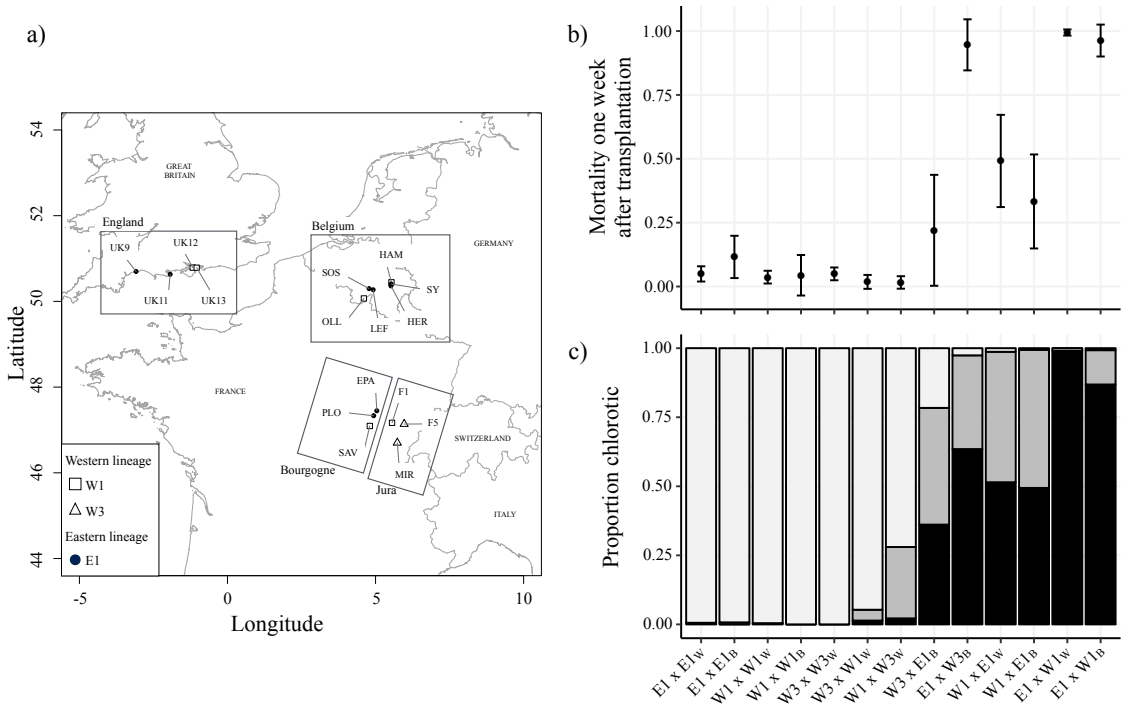



**Figure 3.1:** Geographic location of the 16 study populations in our crossing experiment from four regions – southern England, Bourgogne and Jura (eastern France), and Belgium (a). Mortality (in proportion) of seedlings one week after transplantation (b) and the proportion of one-week old seedlings that were chlorotic and the degree of their chlorosis (light bars - no chlorosis, gray bars - partial chlorosis, dark bars - complete chlorosis, Figure S3.1) (c) of progeny from intra- and inter-lineage crosses (dam x sire) are shown. Within (W) or between (B)-region crosses are indicated.

**Table 3.1:** Results of logistic regression analyses. Mortality refers to mortality one week after transplantation.

| A - Crosses between the W1 and E1 lineages within and between three regions (England, Bourgogne in eastern France, and Belgium) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cross type | | | Within- vs. between-region crosses | | | Cross type x within-/between-region crosses | | | Maternal population (nested in maternal region) | | Maternal region | |
| variable | df | $\chi^2$ | *P* | df | $\chi^2$ | *P* | df | $\chi^2$ | *P* | Variance | SD | Variance | SD |
| Ovules fertilized | 3 | 7.496 | 0.058 | 1 | 21.989 | < 0.001 | 3 | 273.907 | < 0.001 | 0.539 | 0.734 | < 0.001 | < 0.001 |
| Seeds aborted | 3 | 222.430 | < 0.001 | 1 | 0.004 | 0.952 | 3 | 86.320 | < 0.001 | 1.839 | 1.356 | < 0.001 | < 0.001 |
| Germination | 3 | 25.618 | < 0.001 | 1 | 2.724 | 0.100 | 3 | 25.874 | < 0.001 | 0.412 | 0.642 | 0.366 | 0.605 |
| Mortality | 3 | 182.283 | < 0.001 | 1 | 0.085 | 0.771 | 3 | 12.231 | 0.007 | 0.212 | 0.460 | 0.004 | 0.063 |

| B - Crosses between the W1 and W3 lineages from Jura (eastern France) with the E1 lineage from Belgium | | | | | |
|---|---|---|---|---|---|
| | Cross type | | | Maternal population | |
| variable | df | $\chi^2$ | *P* | Variance | SD |
| Ovules fertilized | 3 | 7.695 | 0.053 | 0.101 | 0.318 |
| Seeds aborted | 3 | 11.245 | 0.010 | 0.503 | 0.709 |
| Germination | 3 | 66.920 | < 0.001 | < 0.001 | < 0.001 |
| Mortality | 3 | 52.877 | < 0.001 | 0.024 | 0.154 |

| C - Crosses between the W1 and W3 lineages in Jura (eastern France) | | | | | |
|---|---|---|---|---|---|
| | Cross type | | | Maternal population | |
| variable | df | $\chi^2$ | *P* | Variance | SD |
| Ovules fertilized | 3 | 4.686 | 0.196 | 0.027 | 0.164 |
| Seeds aborted | 3 | 8.772 | 0.033 | 0.052 | 0.052 |
| Germination | 3 | 124.060 | < 0.001 | < 0.001 | < 0.001 |
| Mortality | 3 | 6.253 | 0.100 | < 0.001 | < 0.001 |

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



While intra-lineage progeny were mostly green and healthy (more than 99%), W1 x E1 progeny were half partially chlorotic (48%) and half fully chlorotic (51%), and most E1 x W1 progeny were fully chlorotic (98%). For both W1 x E1 and E1 x W1 cross type, there was no significant association between the seedling chlorosis levels and whether crosses were performed within- or between- regions ($\chi^2$ (df=2) = 0.993, P = 0.609 and $\chi^2$ (df=2) = 4.900, P = 0.087). Because most of the inter-lineage hybrid progeny died, we were not able to test for an effect of cross type on vegetative growth. However, we noted that the few surviving hybrids were markedly smaller than intra-lineage progeny regardless of whether the cross was within or between regions (data not shown).

At the end of five weeks, on average of 67.7% of W1 x E1 progeny and 99.6% of E1 x W1 progeny were dead, while most intra-lineage progeny survived (5.4% and 8.9% of the progeny died in W1 x W1 and E1 x E1 crosses respectively, Figure S3.2). Grouping within- and between-region crosses, the strength of the F1 survival barrier was $RI_{E1xW1}$ = 0.996 and $RI_{W1xE1}$ = 0.645, resulting in a degree of asymmetry of 0.351.

The pattern of PRI between the E1 lineage in Belgium and both the Jura W1 and W3 lineages revealed significant differences between cross types in the proportion of seeds aborted, germinated, and mortality (Table 3.1b, Figure S3.3in the appendices). Seed abortion was significantly higher for E1 x W1 progeny than E1 x W3 progeny (Figure S3.3in the appendices). While the proportion of seeds germinated was similar when the maternal lineage was from both W1 and W3 lineages, we found a stronger reduction in germination for E1 x W1 crosses than E1 x W3 crosses (Figure S3.3in the appendices). Mortality one week following transplantation was significantly higher when the maternal lineage was from the E1 lineage than when it was from either the W1 or W3 lineages (Table 3.1b, Figure 3.1b). The level of seedling chlorosis mirrored the pattern of mortality, with more fully chlorotic seedlings occurring when the E1 lineage was the maternal lineage (Figure 3.1b). Finally, at the end of the five weeks of survey, most of the survivors were from the W3 x E1 cross type (Table S3.2 and Figure S3.3 in the appendices).

### 3.3.2 Can we detect PRI between the W1 and W3 lineages?

For the crosses within Jura, both the proportion of ovules fertilized and seeds aborted were similar between cross types (despite a significant effect of cross type on the proportion of seeds aborted, Table 3.1c, post hoc tests did not reveal any differences, Figure S3.4 in the appendices). The significant effect of cross type on germination was caused by lower germination for W1 x W3 seeds and higher germination for seeds from the intra-lineage W3 x W3 cross compared


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



to seeds from the other cross types (Table S3.2 and Figure S3.4 in the appendices). W1 x W3 progeny had the highest chlorosis level—26.7% of seedlings were partially chlorotic (Figure 3.1c). Fully chlorotic seedlings were very rare and no significant differences in the proportion of seeds that died were found among the cross types (Figure 3.1b, Table 3.1c). Finally, we did not see a reduction in growth for inter-lineage progeny based on the number of leaves ($\chi^2$(df = 12) = 11.708, $P$ = 0.469). We detected a significant reduction in rosette diameter in inter-lineage progeny compared to intra-lineage progeny ($\chi^2$(df = 12) = 304.305, $P$ < 0.001) (Figure S3.4in the appendices), with inter-lineage progeny being on average 10% smaller than intra-lineage progeny at the end of the survey (mean value: 10.9 cm vs. 12.2 cm, respectively). Five weeks after transplantation, the proportion of seedling that had died was less than 0.100. The strength of the F1 survival barriers was $RI_{W1xW3}$ = 0.016 and $RI_{W3xW1}$ = 0.018, with a degree of asymmetry = 0.002.

### 3.3.3 Can we detect inter-lineage hybrids in natural populations?

At the population level, $F_{IS}$ values ranged from 0.01 to 0.17 and mirrored low estimates of selfing ($s$ = 0.00 to 0.13; Table S3.1). The combination of six plastid SNPs assigned individuals to the three lineages and no intra-population polymorphism was found. Within each region, the highest likelihood found using the $\Delta K$ statistic revealed clustering of individuals into two groups corresponding to the two lineages, with no hybrids identified (Figure 3.2). The first axis of the PCA clearly differentiated the lineages without intermediate individuals and explained from 7.8% of the variance in Belgium to 16.3% in England (Figure 3.2). The second axis split E1 populations or W3 populations, while the two W1 populations in England and Belgium stayed grouped together and explained from 3% of the variance in Belgium to 7.7% in England (Figure 3.2).

## 3.4 Discussion

Western and eastern lineages of *S. nutans* have come into secondary contact following post-glaciation spread (Martin et al., 2016). We took advantage of the presence of regions of contact to test whether reproductive isolation between lineages was a consequence of the accumulation of incompatibilities that arose when the lineages were allopatric, or whether this isolation was a result of local divergence in sympatry following the spread. We found strong PRI between both of the western lineages (W1 and W3) with the eastern E1 lineage across all study regions regardless of whether the crosses were within or between regions. The PRI was expressed mainly in the form of seedling chlorosis that resulted in mortality. Mortality among cross types was asymmet-

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



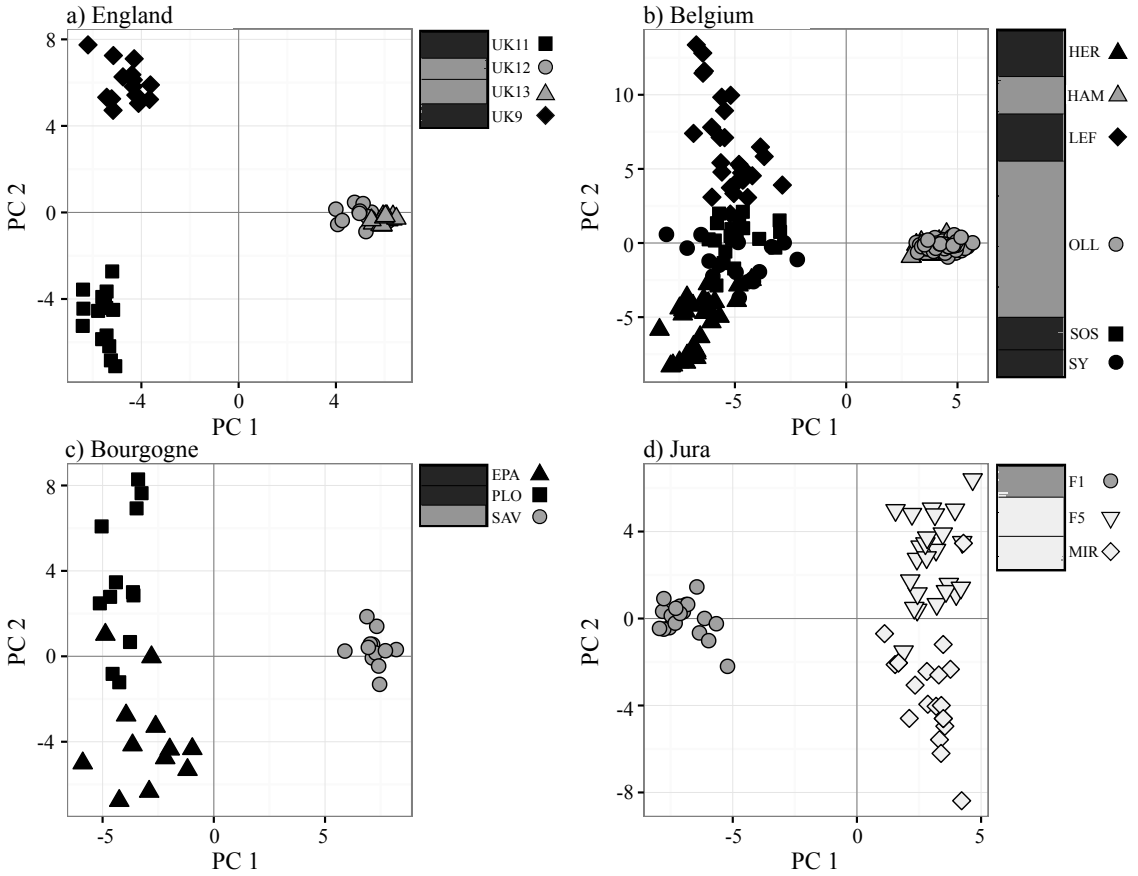



**Figure 3.2:** Genetic assignment of individuals sampled from natural populations from which parents of the crosses originated. For each region a) England, b) Belgium c) Bourgogne and d) Jura, the plot of the first and second axes of the Principal Component Analysis (PCA) and bar-plot of assignment probabilities of an individual's membership into $K$=2 clusters based on their genotype at 25 microsatellite loci are given. Each individual is represented by a thin line partitioned into two colored segments displaying the individual's estimated membership fraction in the two clusters. Label colors in the bar-plot refer to plastid SNP combinations.

ric and dependent on which lineage acted as the maternal parent, consistent with cytonuclear incompatibility. Furthermore, no hybrids were found in nature. Together, these findings reveal strong barriers between the western and eastern lineages that arose prior to secondary contact, and indicate that these lineages should be considered separate species. PRI was not as strong between the two western lineages under study (W1 and W3) in the one region of overlap, as expected given that they diverged more recently than either did from the E1 lineage. We suggest that the relatively rapid, strong PRI that developed between western and eastern lineages is consistent with the hypothesis that speciation can occur quickly when cytonuclear incompatibilities are involved.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



### 3.4.1 Strong PRI between western and eastern lineages reveals more than one species

It was previously found that the E1 lineage, which originated in eastern Europe, was more genetically divergent to the two lineages that originated in western Europe, the W1 and W3 lineages, than the latter two were to each other (Martin et al., 2016, Van Rossum et al., 2016). In addition, the W1 and E1 lineages were found to be differentiated to different soil types in southern Belgium (e.g. Van Rossum et al., 1999), where strong PRI was found (Van Rossum et al., 1996). This raised the possibility that the high PRI seen between these lineages in Belgium was a result of local ecotypic differentiation. However, by performing a crossing experiment using plants from multiple regions of overlap, we found strong PRI between both the western lineages under study (W1 and W3) and the E1 lineage that was widespread across Europe. In accordance with inter-lineage hybrids always suffering from high seedling mortality, our population genetic analysis did not detect any inter-lineage hybrids.

The high seedling mortality between W1 and E1 observed both in within- and between-region crosses led to estimates of barriers to F1 survival that are relatively high ($RI_{E1xW1}$ = 0.996 and $RI_{W1xE1}$ = 0.645) compared to those reported in a review by Lowry et al. (2008a). Using nineteen plant systems, they compiled the strength of prezygotic and postzygotic reproductive isolating mechanisms. In five plant systems, they report hybrid vigor leading to negative values of the F1 viability barriers, and when positive, these values did not exceed 0.190.

Taken together, our results suggest that reproductive isolation between the western and eastern lineages is complete and occurs as the result of genetic incompatibilities that accumulated in allopatry prior to the spread of the eastern lineage to the west. In other words, subsequent post-glaciation secondary contact and local ecotypic differentiation do not appear to have led to an increase in the extent of PRI.

### 3.4.2 The degree of PRI varies with genetic relatedness

While chlorosis leading to high mortality was seen for hybrids between more genetically distant lineages, this form of PRI was not as strong between the two more closely related lineages, W1 and W3. In addition, their juvenile growth was only mildly affected following inter-lineage hybridization. Nevertheless, our population genetic analysis indicates that hybrids between the two study lineages that originated in the west, W1 and W3, do not occur in natural populations. In this study, as well as in a previous work, we showed gene flow between populations

© 2016 Tous droits réservés.

 lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



of the same lineage at similar geographic distances (Figure 3.2 and Martin et al., 2016); therefore, geographic distance does not seem to explain the absence of hybrids between lineages in natural populations. However, it should be noted that our survey lasted only five weeks after germination, so we cannot exclude PRI expressed at later developmental stages, e. g., adult mortality or sterility. In addition, the involvement of prezygotic reproductive barriers between lineages occurring in close geographic vicinity cannot be excluded, such as pollen competition favoring intra-lineage crosses. Several pollinator groups visit *S. nutans* flowers: Noctuidae (especially *Hadena* sp.) and one taxa of *Sphingidae* are the main nocturnal pollinators. There are also diurnal visitors, in particular several taxa of *Apidae* and *Syrphidae* (Hepper, 1956, Jürgens et al., 1996). Further studies are needed to determine whether the observed variation in flower morphology (Van Rossum, unpublished results) and possible variation in floral scent among *S. nutans* lineages impacts pollinator preferences and results in prezygotic reproductive isolation, as observed in other Silene species (Waelti et al., 2008, Montgomery et al., 2010).

### 3.4.3 Tempo of PRI

Our results raise questions regarding the tempo of the accumulation of genetic incompatibilities that led to the complete reproductive isolation of these phylogenetic lineages, which until now have been assumed to constitute a single species. This single-species designation was based on the lack of any discernable geographic pattern of variation in morphological tendencies, mainly based on vegetative traits (Jeanmonod and Bocquet, 1983). We now know that the geographical distribution of the genetically differentiated lineages exhibits a clear pattern that can be assigned to Quaternary climate oscillations, i.e., fewer than 1 My ago (Martin et al., 2016). When compared to other species pairs, the rise of reproductive isolating barriers seems to be fast. In the same genus, the dioecious species *S. latifolia* and *S. dioica*, which diverged more than 1.5 My ago, have been described as fully cross fertile (Karrenberg and Favre, 2008), even though postzygotic isolation that follows Haldane's rule, lower fitness of the hybrids of the heterogametic sex, has been documented (Brothers and Delph, 2010). *Helianthus annus* and *H. petiolaris*, which diverged approximately 1 My ago, hybridize naturally (Strasburg and Rieseberg, 2008), even though postzygotic isolation is expressed in the form of hybrid sterility (Lowry et al., 2008a). *Arabidopsis arenosa* and *A. lyrata*, which diverged 1-2 My ago, also naturally hybridize (Muir et al., 2015). Even animal taxa, e.g., *Mytilus edulis* and *M. galloprovincialis* (diverged 2.5 My ago), exhibit strong but incomplete reproductive isolation, such that hybridization occurs in nature after secondary contact (Bierne et al., 2003, Roux et al., 2014).


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



The relatively rapid rate of complete reproductive isolation between lineages of *S. nutans* may be a result of cytonuclear incompatibility. *Silene nutans* is gynodioecious, with nuclear-cytoplasmic control of sex; this breeding system gives rise to genomic conflict between cytoplasmic male sterility factors and nuclear restorers of fertility (Garraud et al., 2011). The "arms-race" between a selfish element and a counteracting restorer can drive the evolution of genetic incompatibilities (Phadnis and Orr, 2009, Case et al., 2016). In particular, when a conflict occurs between genomes and these genomes have evolved independently across populations, cytonuclear incompatibility can arise rapidly (Chou and Leu, 2010). Interestingly, inter-lineage seedling chlorosis and hybrid mortality were asymmetric in both types of crosses involving either the W1 or W3 lineages with the E1 lineage, with higher mortality when the maternal parent was from the E1 lineage. Asymmetry in reproductive isolation is often seen when reciprocal crosses are conducted between species (Orr, 1995). Several genetic causes have been proposed to explain this asymmetry, including cytonuclear incompatibilities (Levin, 2003, Turelli and Moyle, 2007). This mechanism can also lead to hybrid chlorosis (or bleaching Greiner et al., 2011). The involvement of such incompatibilities is gaining support as an important component of the speciation process (Greiner et al., 2011, Gagnaire et al., 2012, Simon et al., 2016, Case et al., 2016), and as seen here, may be involved in rapid development of reproductive isolation.

## Acknowledgements

We thank S. Le Cadre for her contribution to the plant material sampling; the BSBI (D.A. Pearman, R.E.N. Smith, MJ Hawksford, S. Wild, W. McCarthy, A. Willmot, M.W. Rand, A. Knapp, J. Knight, Q. Groom) for providing location of populations; local managers (e.g. G. Bryant - Natural England, M. Bartlett - Browndown Training Camp) for access to natural sites; T. Lesaffre for molecular analyses; C. Glorieux, A. Bourceaux, N. Faure, F. Joly, and L. Frachon, for seedling transplantation and plant maintenance in the greenhouse. This experiment was supported by grants from the National Fund for Scientific Research (FNRS, Belgium) to FVR, the French National research Agency (ANR-11-BSV7-013-03, TRANS) to PT, French Research Ministry (PhD) and Collège doctoral Lille Nord de France to HM.

## 3.5 Bibliography

Baack, E., Melo, M. C., Rieseberg, L. H., and Ortiz-barrientos, D. (2015). The origins of reproductive isolation in plants. *New Phytologist*, 207:968–984.



© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67:1–48.

Bierne, N., Borsa, P., Daguin, C., Jollivet, D., Viard, F., Bonhomme, F., and David, P. (2003). Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. *Molecular Ecology*, 12:447–461.

Brothers, A. N. and Delph, L. F. (2010). Haldane's rule is extended to plants with sex chromosomes. *Evolution*, 64:3643–3648.

Case, A. L., Finseth, F. R., Barr, C. M., and Fishman, L. (2016). Selfish evolution of cytonuclear hybrid incompatibility in *Mimulus*. *Proceedings of the Royal Society B*, 283:20161493.

Chou, J.-Y. and Leu, J.-Y. (2010). Speciation through cytonuclear incompatibility: insights from yeast and implications for higher eukaryotes. *Bioessays*, 32:401–411.

Coyne, J. A. and Orr, H. A. (2004). *Speciation*. Sinauer associates, Sunderland, Massachusetts, USA, 1st edition.

David, P., Pujol, B., Viard, F., Castella, V., and Goudet, J. (2007). Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology*, 16:2474–2487.

De Bilde, J. (1973). Etude génécologique du *Silene nutans* L. en Belgique: populations du *Silene nutans* L. sur substrats siliceux et calcaires. *Revue Générale de Botanique*, 80:161–176.

Dufay, M., Lahiani, E., and Brachi, B. (2010). Gender variation and inbreeding depression in gynodioecious-gynomonoecious *Silene nutans* (Caryophyllaceae). *International Journal of Plant Sciences*, 171:53–62.

El Mousadik, A. and Petit, R. J. (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinos* (L.) Skeels] endemic to Morocco. *Theoretical and Applied Genetics*, 92:832–839.

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

Fitter, A. (1978). *An atlas of the wild flowers of Britain and Northern Europe*. Collins, London, UK.

Fraïsse, C., Elderfield, J. A. D., and Welch, J. J. (2014). The genetics of speciation: are complex incompatibilities easier to evolve? *Journal of Evolutionary Biology*, 27:688–699.

Gagnaire, P.-A., Normandeau, E., and Bernatchez, L. (2012). Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American Eels. *Molecular Biology and Evolution*, 29:2909–2919.

Garraud, C., Brachi, B., Dufay, M., Touzet, P., and Shykoff, J. A. (2011). Genetic determination of male sterility in gynodioecious *Silene nutans*. *Heredity*, 106:757–764.

Godé, C., Touzet, P., Martin, H., Lahiani, E., Delph, L. F., and Arnaud, J.-F. (2014). Characterization of 24 polymorphic microsatellite markers for *Silene nutans*, a gynodioecious-gynomonoecious species, and cross-species amplification in other *Silene* species. *Conservation Genetics Resources*, 6:915–918.

Goudet, J. (1995). FSTAT (Version 1.2): a computer program to calculate *F*-statistics. *The Journal of Heredity*, 86:485–486.

Greiner, S., Rauwolf, U., Meurer, J., and Herrmann, R. G. (2011). The role of plastids in plant speciation. *Molecular Ecology*, 20:671–691.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



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

Hauser, T. P. and Siegismund, H. R. (2000). Inbreeding and outbreeding effects on pollen fitness and zygote survival in *Silene nutans* (Caryophyllaceae). *Journal of Evolutionary Biology*, 13:446–454.

Hauser, T. P. and Weidema, I. R. (2000). Extreme variation in flowering time between populations of *Silene nutans*. *Hereditas*, 132:95–101.

Hepper, F. N. (1951). The variations of *Silene nutans* L. in Great Britain. *Watsonia*, 2:80–90.

Hepper, F. N. (1956). Biological flora of the British Isles: *Silene nutans* L. *Journal of Ecology*, 44:693–700.

Hopkins, R. (2013). Reinforcement in plants. *New Phytologist*, 197:1095–1103.

Jacquemyn, H., Brys, R., Honnay, O., and Roldan-Ruiz, I. (2012). Asymmetric gene introgression in two closely related *Orchis* species: evidence from morphometric and genetic analyses. *BMC Evolutionary Biology*, 12:178–189.

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

Jeanmonod, D. and Bocquet, G. (1983). Propositions pour un traitement biosystématique du *Silene nutans* L. Caryophyllaceae. *Candollea*, 38:267–295.

Jombart, T. (2008). *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24:1403–1405.

Jürgens, A., Witt, T., and Gottsberger, G. (1996). Reproduction and pollination in central European populations of *Silene* and *Saponaria* species. *Botanica Acta*, 109:316–324.

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

Lahiani, E., Dufaÿ, M., Castric, V., Le Cadre, S., Charlesworth, D., Van Rossum, F., and Touzet, P. (2013). Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity*, 111:157–164.

Lahiani, E., Touzet, P., Billard, E., and Dufay, M. (2015). When is it worth being a self-compatible hermaphrodite? Context-dependent effects of self-pollination on female advantage in gynodioecious *Silene nutans*. *Ecology and Evolution*, 5:1854–1862.

Levin, D. A. (2003). The cytoplasmic factor in plant speciation. *Systematic Botany*, 28:5–11.

Lindtke, D. and Buerkle, C. A. (2015). The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution*, 69:1987–2004.

Lowry, D. B., Modliszewski, J. L., Wright, K. M., Wu, C. A., and Willis, J. H. (2008a). The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363:3009–3021.

Lowry, D. B., Rockwood, R. C., and Willis, J. H. (2008b). Ecological reproductive isolation of coast and inland races of *Mimulus guttatus*. *Evolution*, 62:2196–2214.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Lynch, M. and Hagner, K. (2015). Evolutionary meandering of intermolecular interactions along the drift barrier. *Proceedings of the National Academy of Sciences*, 112:30–38.

Martin, H., Touzet, P., Van Rossum, F., Delalande, D., and Arnaud, J.-F. (2016). Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity*, 116:286–294.

Montgomery, B. R., Soper, D. M., and Delph, L. F. (2010). Asymmetrical conspecific seed-siring advantage between *Silene latifolia* and *S. dioica*. *Annals of Botany*, 105:595–605.

Moyle, L. C., Olson, M. S., and Tiffin, P. (2004). Patterns of reproductive isolation in three angiosperm genera. *Evolution*, 58:1195–1208.

Muir, G., Ruiz-Duarte, P., Hohmann, N., Mable, B. K., Novikova, P., Schmickl, R., Guggisberg, A., and Koch, M. A. (2015). Exogenous selection rather than cytonuclear incompatibilities shapes asymmetrical fitness of reciprocal *Arabidopsis* hybrids. *Ecology and Evolution*, 5:1734–1745.

Noor, M. A. F. (1999). Reinforcement and other consequences of sympatry. *Heredity*, 83:503–508.

Orr, H. A. (1995). Population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*, 139:1805–1813.

Phadnis, N. and Orr, H. A. (2009). A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. *Science*, 323:376–379.

Piry, S., Alapetite, A., Cornuet, J.-M., Paetkau, D., Baudouin, L., and Estoup, A. (2004). GENECLASS2: a software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, 95:536–539.

Presgraves, D. C. (2010). The molecular evolutionary basis of species formation. *Nature Reviews Genetics*, 11:175–180.

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

R Core Team Development (2014). R: a language and environment for statistical computing. *R Foundation for Statistical Computing Vienna, Austria*, https://www.r–project.org/.

Rieseberg, L. H. and Willis, J. H. (2007). Plant speciation. *Science*, 317:910–914.

Rosenberg, N. A. (2004). DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4:137–138.

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

Roux, C., Fraïsse, C., Castric, V., Vekemans, X., Pogson, G. H., and Bierne, N. (2014). Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *Journal of Evolutionary Biology*, 27:1662–1675.

Scopece, G., Widmer, A., and Cozzolino, S. (2008). Evolution of postzygotic reproductive isolation in a guild of deceptive orchids. *The American Naturalist*, 171:315–326.

Simon, M., Durand, S., Pluta, N., Gobron, N., Botran, L., Ricou, A., Camilleri, C., and Budar, F. (2016). Genomic conflicts that cause pollen mortality and raise reproductive barriers in *Arabidopsis thaliana*. *Genetics*, 203:1353–1367.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Smadja, C. M. and Butlin, R. K. (2011). A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, 20:5123–5140.

Strasburg, J. L. and Rieseberg, L. H. (2008). Molecular demographic history of the annual sunflowers *Helianthus annuus* and *H. petiolaris* - large effective population sizes and rates of long-term gene flow. *Evolution*, 62:1936–1950.

Turelli, M. and Moyle, L. C. (2007). Asymmetric postmating isolation: Darwin's corollary to Haldane's rule. *Genetics*, 176:1059–1088.

Vähä, J. P. and Primmer, C. R. (2006). Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, 15:63–72.

Van Rossum, F., De Bilde, J., and Lefèbvre, C. (1996). Barriers to hybridization in calcicolous and silicolous populations of *Silene nutans* from Belgium. *Belgian Journal of Botany*, 129:13–18.

Van Rossum, F., Meerts, P., Gratia, E., and Tanghe, M. (1999). Ecological amplitude in *Silene nutans* in relation to allozyme variation at the western margin of its distribution. *Journal of Vegetation Science*, 10:253–260.

Van Rossum, F., Weidema, I. R., Martin, H., Le Cadre, S., Touzet, P., Prentice, H. C., and Philipp, M. (2016). The structure of allozyme variation in *Silene nutans* (Caryophyllaceae) in Denmark and in north-western Europe. *Plant Systematics and Evolution*, 302:23–40.

Waelti, M. O., Muhlemann, J. K., Widmer, A., and Schiestl, F. P. (2008). Floral odour and reproductive isolation in two species of *Silene*. *Journal of Evolutionary Biology*, 21:111–121.

Weir, B. S. and Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, 38:1358–1370.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



## 3.6 Appendices


**Figure S3.1:** Categories of the degree of chlorosis seen in the progeny.

**Table S3.1:** Populations details and within-population genetic-variation estimates from collections of individual plants in the wild: geographic coordinates, sample size (*n*), lineages (eastern, E1 and western, W1 and W3), allelic richness (*An*), observed heterozygosity ($H_O$), expected heterozygosity ($H_E$), intra-population fixation index ($F_{IS}$) and selfing rate (s). Populations were assigned to lineages using plastid SNP combinations.

| Region | Population | Latitude N | Longitude | *n* | Lineage | *Ar* | $H_O$ | $H_E$ | $F_{IS}$ | *s* |
|---|---|---|---|---|---|---|---|---|---|---|
| Southern | UK9 | 50.698 | 3.087 W | 14 | E1 | 4.4 | 0.47 | 0.56 | 0.10 | 0.11 |
| England | UK11 | 50.632 | 1.936 W | 16 | E1 | 4.4 | 0.51 | 0.59 | 0.01 | 0.00 |
| | UK12 | 50.789 | 1.175 W | 13 | W1 | 3.2 | 0.41 | 0.49 | 0.02 | 0.00 |
| | UK13 | 50.786 | 1.041 W | 15 | W1 | 2.6 | 0.36 | 0.38 | 0.03 | 0.14 |
| Belgium | HAM | 50.446 | 5.537 E | 23 | W1 | 4.6 | 0.41 | 0.51 | 0.02 | 0.07 |
| | OLL | 50.069 | 4.606 E | 96 | W1 | 6.3 | 0.49 | 0.60 | 0.02 | 0.03 |
| | SOS | 50.298 | 4.780 E | 20 | E1 | 6.9 | 0.51 | 0.67 | 0.06 | 0.02 |
| | SY | 50.404 | 5.526 E | 16 | E1 | 4.6 | 0.50 | 0.60 | 0.05 | 0.03 |
| | LEF | 50.271 | 4.918 E | 29 | E1 | 6.8 | 0.51 | 0.62 | 0.02 | 0.00 |
| | HER | 50.385 | 5.514 E | 28 | E1 | 6.5 | 0.56 | 0.62 | 0.01 | 0.02 |
| Bourgogne, | EPA | 47.448 | 5.043 E | 12 | E1 | 5.6 | 0.53 | 0.66 | 0.04 | 0.04 |
| eastern France | PLO | 47.332 | 4.936 E | 12 | E1 | 5.5 | 0.52 | 0.62 | 0.02 | 0.00 |
| | SAV | 47.093 | 4.806 E | 12 | W1 | 3.7 | 0.48 | 0.52 | 0.02 | 0.00 |
| Jura, | F1 | 47.167 | 5.557 E | 19 | W1 | 4.4 | 0.39 | 0.50 | 0.08 | 0.05 |
| eastern France | F5 | 47.138 | 5.967 E | 24 | W3 | 6.4 | 0.47 | 0.61 | 0.03 | 0.00 |
| | MIR | 46.700 | 5.734 E | 20 | W3 | 5.0 | 0.37 | 0.57 | 0.17 | 0.13 |

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



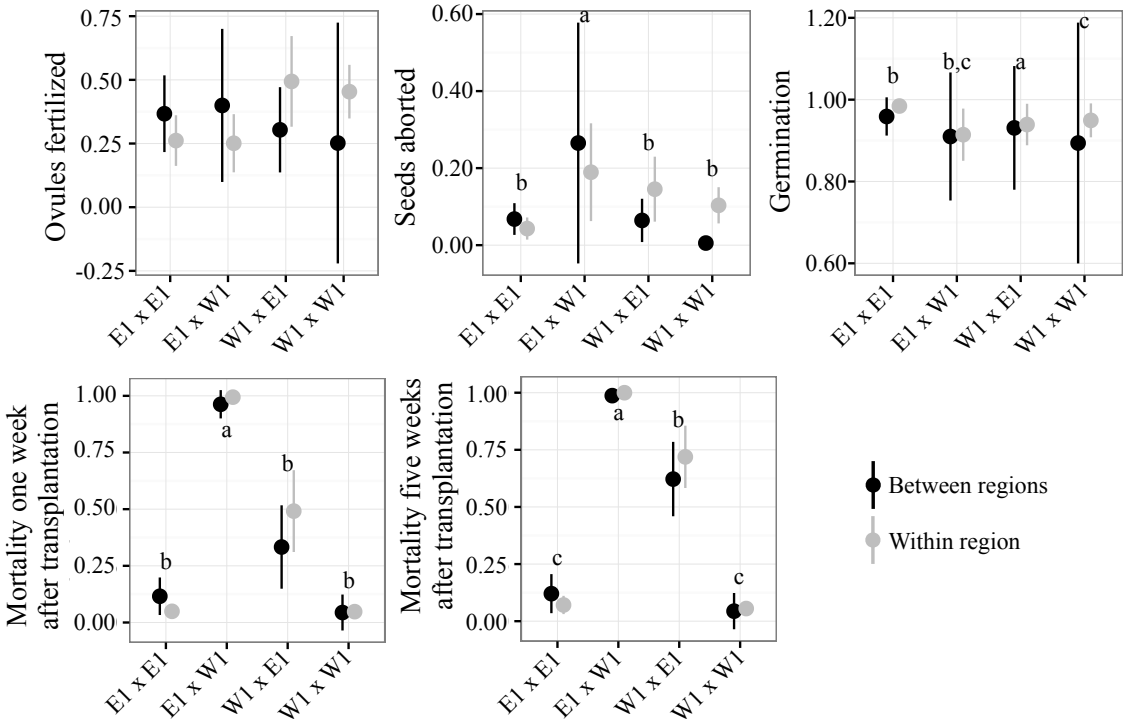



**Figure S3.2:** Fitness measurements of the progeny from crosses between the W1 and E1 lineages within and between three regions (England, Bourgogne in eastern France, and Belgium). For each cross type, the mean and 95% CI of the proportion of ovules fertilized, proportion of seeds aborted, germination (in proportion), and mortality (in proportion) one week and five weeks following transplantation of within- and between- crosses are plotted. Small letters a, b and c show the results of post-hoc Tukey tests on pairwise comparison between cross types (when significant), within and between region crosses together. Different letters indicate a significant difference ($P < 0.05$).

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



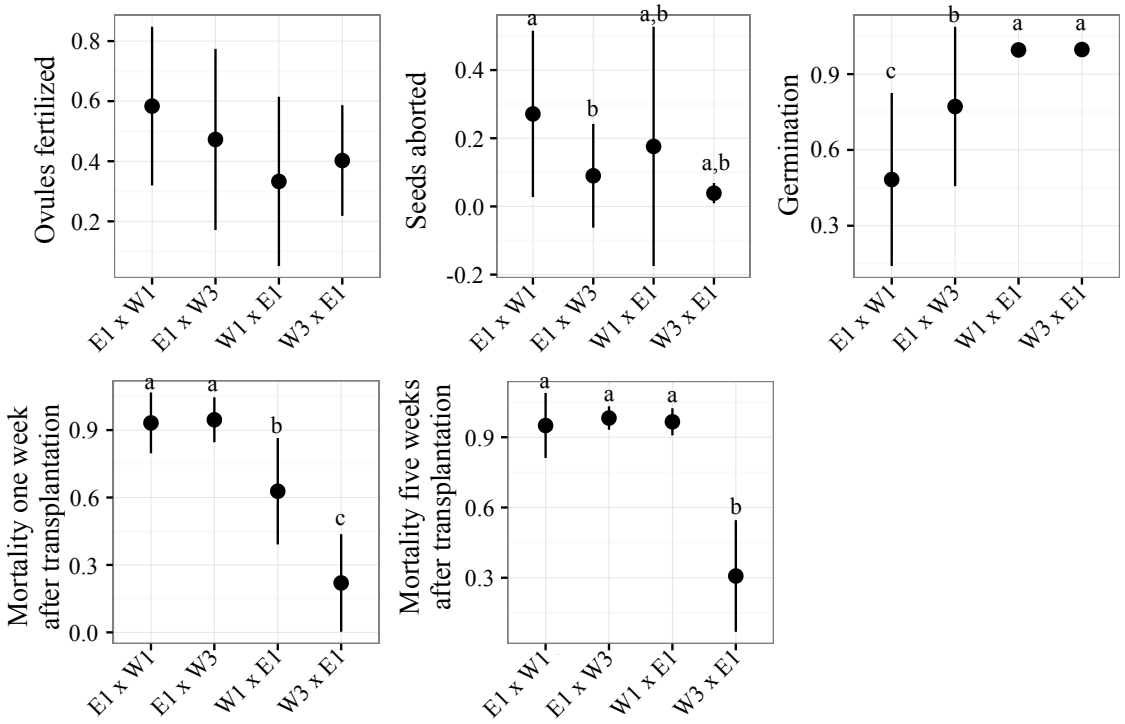



**Figure S3.3:** Fitness measurements of the progeny from crosses between the W1 and W3 lineages from Jura (eastern France) with the E1 lineage from Belgium. For each cross type, we plot the mean and 95% CI of the proportion of ovules fertilized, proportion of seeds aborted, germination (in proportion), and mortality (in proportion) one week and five weeks following transplantation. All crosses were performed between regions. Small letters a, b and c show the results of post-hoc Tukey tests on pairwise comparison between cross types (when significant). Different letters indicate a significant difference ($P < 0.05$).

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



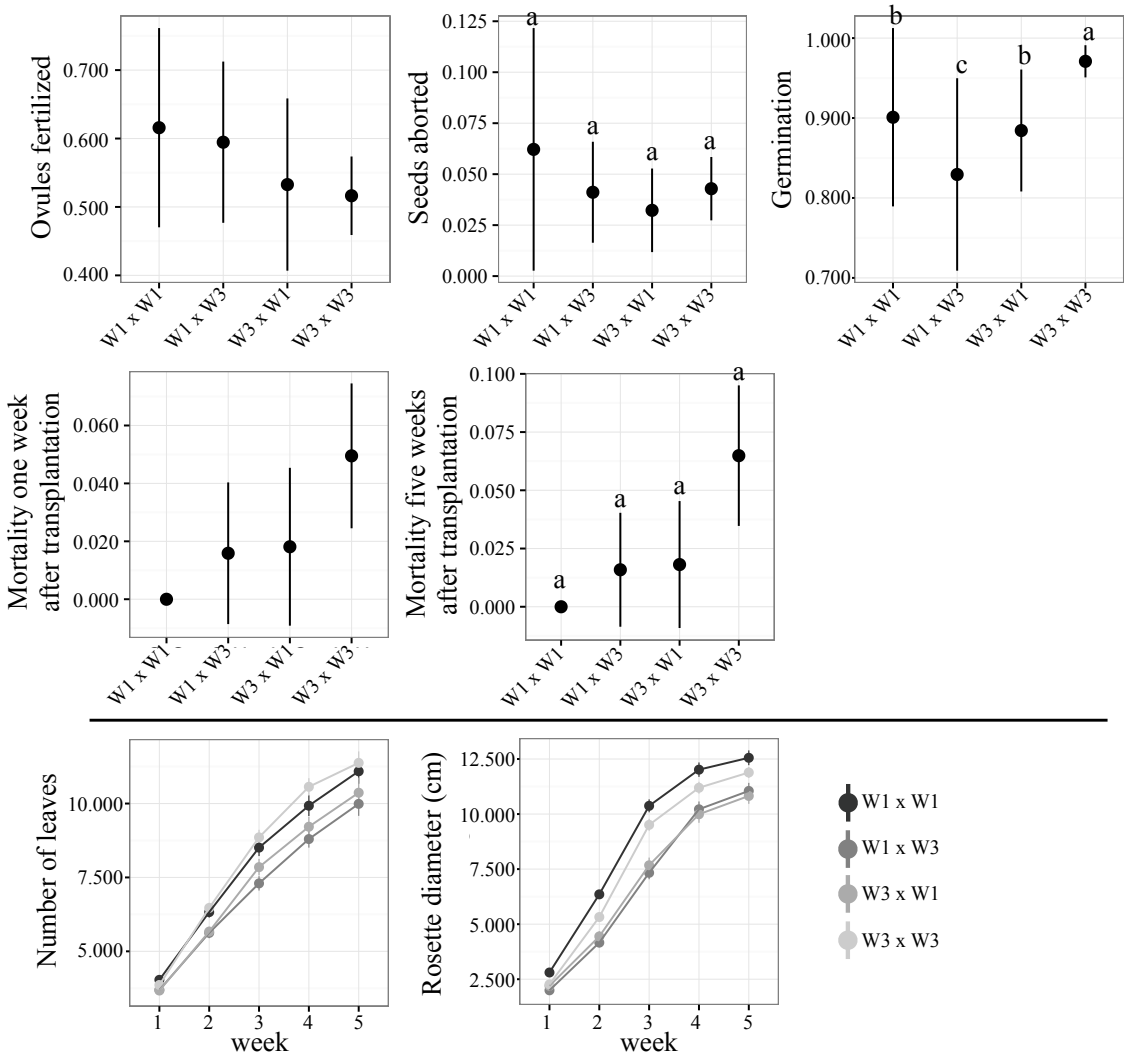



**Figure S3.4:** Fitness measurements of the progeny from crosses between the W1 and W3 lineages in Jura (eastern France). For each cross type, we plot the mean and 95% CI of the proportion of ovules fertilized, proportion of seeds aborted, germination (in proportion), and mortality (in proportion) one week and five weeks following transplantation. Growth is represented based on the number of leaves and the rosette diameter. All crosses were within Jura. Small letters a, b and c show the results of post-hoc Tukey tests on pairwise comparison between cross types (when significant). Different letters indicate a significant difference ($P < 0.05$).

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



**Table S3.2:** Summary of the crosses from the crossing experiment: cross type, regions used for the crosses (UK = southern England, Bo = Bourgogne, Ju = Jura, and Be = Belgium), number of crosses made, mean ±SD of the number of seeds per cross (fruit), the proportion of ovules fertilized, the proportion of seeds that were aborted and germinated, the number of seedlings transplanted per cross, and the mortality (in proportion) following the first and fifth week of transplantation. W = within region, B = between region.

| Cross type | Region | No. crosses | Mean number of seeds per cross | Ovules fertilized | Seeds aborted | Germination | Mean no. seedlings per cross | Mortality after one week | Mortality after five weeks |
|---|---|---|---|---|---|---|---|---|---|
| A - Crosses between the W1 and E1 lineages within and between three regions (England, Bourgogne in eastern France, and Belgium) | | | | | | | | | |
| W1 x $W1_W$ | Be x Be | 16 | 32.6 ±20.04 | 0.481 ±0.347 | 0.124 ±0.118 | 0.961 ±0.090 | 10.3 ±2.15 | 0.046 ±0.098 | 0.058 ±0.102 |
| W1 x $W1_W$ | Bo x Bo | 2 | 27.5 ±31.82 | 0.431 ±0.465 | 0.313 ±0.428 | 0.800 ±0.283 | 6.5 ±6.36 | 0.045 ±0.064 | 0.045 ±0.064 |
| W1 x $W1_W$ | UK x UK | 15 | 35.8 ±15.88 | 0.428 ±0.235 | 0.053 ±0.052 | 0.957 ±0.116 | 11.4 ±0.91 | 0.050 ±0.089 | 0.055 ±0.088 |
| W1 x $W1_B$ | Be x UK | 4 | 14.7 ±23.67 | 0.242 ±0.436 | 0.007 ±0.014 | 1.000 ±0.000 | 5.2 ±5.31 | 0.055 ±0.068 | 0.055 ±0.068 |
| W1 x $W1_B$ | Bo x Be | 1 | 36.0 (NA) | 0.290 (NA) | 0.000 (NA) | 0.470 (NA) | 11.0 (NA) | 0.000 (NA) | 0.000 (NA) |
| E1 x $E1_W$ | Be x Be | 21 | 18.4 ±17.27 | 0.207 ±0.235 | 0.032 ±0.069 | 0.990 ±0.036 | 7.6 ±3.67 | 0.056 ±0.094 | 0.083 ±0.124 |
| E1 x $E1_W$ | Bo x Bo | 3 | 12.0 ±11.79 | 0.159 ±0.175 | 0.078 ±0.069 | 0.950 ±0.056 | 7.3 ±4.73 | 0.030 ±0.052 | 0.030 ±0.052 |
| E1 x $E1_W$ | UK x UK | 7 | 33.9 ±19.50 | 0.471 ±0.326 | 0.062 ±0.108 | 0.984 ±0.026 | 10.0 ±3.60 | 0.039 ±0.048 | 0.051 ±0.071 |
| E1 x $E1_B$ | Be x Bo | 8 | 28.0 ±18.82 | 0.340 ±0.319 | 0.041 ±0.049 | 0.941 ±0.125 | 9.8 ±3.01 | 0.109 ±0.101 | 0.109 ±0.101 |
| E1 x $E1_B$ | Bo x Be | 2 | 6.0 ±4.24 | 0.127 ±0.128 | 0.050 ±0.071 | 1.000 ±0.000 | 6.0 ±4.24 | 0.000 ±0.000 | 0.000 ±0.000 |
| E1 x $E1_B$ | Be x UK | 8 | 34.9 ±19.12 | 0.498 ±0.305 | 0.107 ±0.111 | 0.961 ±0.087 | 10.0 ±2.88 | 0.166 ±0.239 | 0.178 ±0.246 |
| E1 x $E1_B$ | UK x Be | 1 | 2.0 (NA) | 0.015 (NA) | 0.000 (NA) | 1.000 (NA) | 2.0 (NA) | 0.000 (NA) | 0.000 (NA) |
| W1 x $E1_W$ | Be x Be | 5 | 38.2 ±18.42 | 0.420 ±0.252 | 0.094 ±0.140 | 0.996 ±0.009 | 10.4 ±1.34 | 0.478 ±0.382 | 0.796 ±0.191 |
| W1 x $E1_W$ | Bo x Bo | 5 | 38.0 ±19.56 | 0.585 ±0.374 | 0.275 ±0.042 | 0.862 ±0.091 | 9.6 ±3.13 | 0.434 ±0.319 | 0.618 ±0.292 |
| W1 x $E1_W$ | UK x UK | 3 | 38.7 ±19.63 | 0.466 ±0.280 | 0.016 ±0.027 | 0.973 ±0.030 | 11.0 ±0.00 | 0.610 ±0.104 | 0.760 ±0.137 |
| W1 x $E1_B$ | Be x Bo | 1 | 3.0 (NA) | 0.026 (NA) | 0.000 (NA) | 1.000 (NA) | 3.0 (NA) | 0.670 (NA) | 0.670 (NA) |
| W1 x $E1_B$ | Bo x Be | 4 | 40.5 ±18.34 | 0.409 ±0.241 | 0.016 ±0.022 | 1.000 ±0.000 | 11.0 ±0.00 | 0.090 ±0.000 | 0.500 ±0.058 |
| W1 x $E1_B$ | Be x UK | 5 | 29.2 ±20.46 | 0.276 ±0.220 | 0.115 ±0.083 | 0.862 ±0.298 | 9.4 ±3.58 | 0.460 ±0.203 | 0.710 ±0.298 |
| E1 x $W1_W$ | Be x Be | 13 | 15.1 ±7.48 | 0.255 ±0.213 | 0.214 ±0.253 | 0.950 ±0.061 | 9.3 ±3.17 | 1.000 ±0.000 | 1.000 ±0.000 |
| E1 x $W1_W$ | Bo x Bo | 2 | 9.0 ±5.66 | 0.082 ±0.057 | 0.000 ±0.000 | 0.670 ±0.184 | 6.5 ±3.54 | 1.000 ±0.000 | 1.000 ±0.000 |
| E1 x $W1_W$ | UK x UK | 1 | 34.0 (NA) | 0.536 (NA) | 0.244 (NA) | 0.940 (NA) | 11.0 (NA) | 0.910 (NA) | 1.000 (NA) |
| E1 x $W1_B$ | Be x UK | 5 | 22.2 ±18.66 | 0.471 ±0.369 | 0.362 ±0.361 | 0.874 ±0.193 | 8.4 ±3.71 | 0.966 ±0.076 | 1.000 ±0.000 |
| E1 x $W1_B$ | UK x Be | 2 | 21.5 ±2.12 | 0.221 ±0.003 | 0.024 ±0.034 | 1.000 ±0.000 | 13.0 ±2.83 | 0.955 ±0.064 | 0.955 ±0.064 |
| B - Crosses between the W1 and W3 lineages from Jura (eastern France) with the E1 lineage from Belgium | | | | | | | | | |
| W3 x $E1_B$ | Ju x Be | 8 | 33.4 ±14.26 | 0.402 ±0.220 | 0.039 ±0.036 | 0.998 ±0.007 | 11.5 ±1.10 | 0.220 ±0.260 | 0.308 ±0.285 |
| E1 x $W3_B$ | Be x Ju | 5 | 39.0 ±15.39 | 0.473 ±0.243 | 0.090 ±0.123 | 0.772 ±0.254 | 11.0 ±0.00 | 0.946 ±0.080 | 0.982 ±0.040 |
| W1 x $E1_B$ | Ju x Be | 5 | 28.6 ±20.34 | 0.333 ±0.227 | 0.176 ±0.283 | 0.996 ±0.009 | 10.0 ±2.83 | 0.628 ±0.191 | 0.966 ±0.047 |
| E1 x $W1_B$ | Be x Ju | 7 | 33.3 ±17.73 | 0.584 ±0.286 | 0.271 ±0.264 | 0.483 ±0.370 | 8.8 ±3.19 | 0.932 ±0.109 | 0.950 ±0.112 |
| C - Crosses between the W1 and W3 lineages in Jura (eastern France) | | | | | | | | | |
| W1 x $W1_W$ | Ju x Ju | 10 | 46.9 ±7.03 | 0.616 ±0.204 | 0.062 ±0.083 | 0.901 ±0.156 | 11.0 ±0.00 | 0.000 ±0.000 | 0.000 ±0.000 |
| W3 x $W3_W$ | Ju x Ju | 41 | 43.9 ±9.56 | 0.516 ±0.182 | 0.043 ±0.049 | 0.971 ±0.064 | 11.0 ±0.63 | 0.050 ±0.079 | 0.065 ±0.096 |
| W1 x $W3_W$ | Ju x Ju | 17 | 46.2 ±7.20 | 0.595 ±0.229 | 0.041 ±0.048 | 0.829 ±0.234 | 11.0 ±0.00 | 0.016 ±0.048 | 0.016 ±0.048 |
| W3 x $W1_W$ | Ju x Ju | 16 | 44.9 ±9.65 | 0.533 ±0.236 | 0.032 ±0.038 | 0.884 ±0.143 | 10.9 ±0.57 | 0.018 ±0.051 | 0.018 ±0.051 |

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



**Table S3.3:** Estimates of nuclear genetic variation for 24 microsatellite loci over 16 populations of Silene nutans for each locus and over all loci: the mean fixation indices ($F_{IT}$, $F_{IS}$, $F_{ST}$), total number of alleles (An), observed heterozygosity ($H_O$), and expected heterozygosity ($H_E$).

| Locus | $F_{IT}$ | $F_{IS}$ | $F_{ST}$ | *An* | $H_O$ | $H_E$ |
|---|---|---|---|---|---|---|
| B09 | 0.422*** | 0.11*** | 0.351*** | 14 | 0.48 | 0.56 |
| H07 | 0.386*** | 0.26*** | 0.169*** | 25 | 0.57 | 0.75 |
| G01 | 0.185*** | 0.11*** | 0.083*** | 53 | 0.76 | 0.88 |
| E08 | 0.553*** | 0.29*** | 0.368*** | 21 | 0.39 | 0.51 |
| D10 | 0.446*** | 0.20*** | 0.303*** | 20 | 0.43 | 0.57 |
| Sil19 | 0.339*** | $0.06^{NS}$ | 0.299*** | 14 | 0.59 | 0.60 |
| Sil24 | 0.184*** | $-0.03^{NS}$ | 0.212*** | 54 | 0.75 | 0.72 |
| Sil36 | 0.419*** | 0.24*** | 0.237*** | 19 | 0.49 | 0.65 |
| Sil42 | 0.445*** | 0.32*** | 0.184*** | 25 | 0.49 | 0.70 |
| Sil16 | 0.485*** | 0.32*** | 0.243*** | 16 | 0.49 | 0.68 |
| Sil31 | 0.483*** | 0.24*** | 0.317*** | 25 | 0.35 | 0.49 |
| Sil35 | 0.335*** | 0.16*** | 0.208*** | 35 | 0.65 | 0.73 |
| Sil37 | 0.189*** | $0.05^{NS}$ | 0.149*** | 28 | 0.72 | 0.76 |
| Sil03 | 0.320*** | $0.03^{NS}$ | 0.300*** | 11 | 0.42 | 0.44 |
| Sil08 | 0.420*** | $0.05^{NS}$ | 0.387*** | 9 | 0.33 | 0.35 |
| Sil15 | 0.758*** | 0.31*** | 0.651*** | 9 | 0.23 | 0.35 |
| Sil01 | 0.378*** | $0.08^{NS}$ | 0.325*** | 9 | 0.47 | 0.50 |
| Sil05 | 0.315*** | 0.12*** | 0.221*** | 7 | 0.47 | 0.54 |
| Sil18 | 0.685*** | 0.56*** | 0.279*** | 9 | 0.20 | 0.42 |
| Sil29 | 0.712*** | 0.60*** | 0.288*** | 8 | 0.20 | 0.54 |
| Sil41 | 0.280*** | 0.10*** | 0.204*** | 10 | 0.55 | 0.59 |
| Sil26 | 0.497*** | 0.34*** | 0.241*** | 24 | 0.33 | 0.55 |
| Sil27 | 0.311*** | 0.11*** | 0.221*** | 31 | 0.58 | 0.67 |
| Sil30 | 0.282*** | $0.07^{NS}$ | 0.227*** | 10 | 0.23 | 0.26 |
| Multilocus | 0.387*** | 0.17*** | 0.260*** | 20 | 0.47 | 0.56 |

*** $P < 0.001$, NS not significant.

© 2016 Tous droits réservés.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

# **Chapitre 3** : Demographic scenario of speciation of *Silene nutans* lineages

**Hélène Martin, Xavier Vekemans, Camille Roux, Sophie Gallina et Pascal Touzet**

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



Pour tenter de retrouver l'architecture génétique de l'isolement reproducteur mis en évidence entre les lignées *Ouest* et *Est* dans le chapitre précédent, nous avons réalisé un scan génomique. Afin de l'interpréter correctement nous avons testé différents scénarios de divergence en tenant compte de l'hétérogénéité des flux de gène le long du génome mais également l'hétérogénéité de la sélection. L'approche ABC que nous avons utilisé nous as permis de montrer que l'hétérogénéité de la sélection le long du génome a façonné le patron de différenciation génétique entre les deux lignées. De ce fait nous n'avons pas été capable de retrouver l'architecture génétique de l'isolement reproducteur.

Une partie des données RNA-seq utilisée dans ce chapitre a été généré avant cette thèse. Lors de cette thèse, j'ai complété l'échantillonnage et traité les données brutes avec l'aide de Sophie Gallina. L'analyse des données nettoyées et l'approche ABC ont été réalisé en étroite collaboration avec Camille Roux. J'ai ensuite écrit le chapitre en incorporant les corrections et remarques des co-auteurs.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



# Linked selection drove the heterogeneous level of differentiation between *Silene nutans* lineages

## 4.1 Introduction

The genetic basis of reproductive isolation (RI) and the process driving their rise are key elements in the study of speciation. Postzygotic RI are thought to be due to negative epistatic interactions that could lead to the sterility or death of the hybrids (Coyne and Orr, 2004). Under the Bateson-Dobzhansky-Müller model of genetic incompatibilities, two isolated populations have diverged from each other and fixed mutations that arose independently in each population at different loci. If both populations hybridize, they bring independent substitutions that might negatively interact (Orr and Presgraves, 2000). If gene flow occurs between these populations, such loci will not introgress. While the rest of the genome is homogenized by gene flow, they differentiate and form outliers of differentiation (Wu, 2001, Bierne et al., 2011). Genomic variation in introgression rate has been successfully tested using simulation methods (Roux et al., 2013, 2014, Tine et al., 2014) and genomic outliers of differentiation have been used to infer the genetic architecture of postzygotic isolation (Harr, 2006, Gagnaire et al., 2013).

However, linked selection and recombination rate can create similar patterns of differentiation (Noor and Bennett, 2009, Cruickshank and Hahn, 2014). By reducing diversity at linked neutral sites, natural selection mechanically increases genetic differentiation and low recombination rates amplify its effect (Charlesworth, 1998). In this scenario, rather than being the result of homogenization by gene flow, regions of low level of differentiation reflect shared ancestral polymorphism, and outliers of differentiation are the result of linked selection that quickly sort out lineages and thus are not necessarily barriers to introgression involved in RI (Cruickshank and Hahn, 2014). Therefore, genome-wide heterogeneity of differentiation could be shaped without gene flow and outliers of differentiation do not necessarily co-occur with loci involved in RI. Contrasting patterns of differentiation with divergence, within-population diversity, and recombination rate, recent studies suggested that linked selection played a major role in shaping genomic variation of differentiation (Renaut et al., 2013, Cruickshank and Hahn, 2014, Burri et al., 2015, Elyashiv et al., 2016, Wang et al., 2016). Therefore, there are two alternative scenarios to explain the genomic landscape of differentiation between two species: either linked selection, that will concomitantly generate genome wide rates of diversity, or heterogeneous gene flow.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



In this study, we focus on two lineages (eastern and western) of the herbaceous *Silene nutans* that co-occur in parapatry in Western Europe, after postglacial recolonization from distinct refugia (Martin et al., 2016, Van Rossum et al., 2016). Although they are found in close parapatry in some regions, no evidence of gene flow was recorded, but a strong postzygotic RI was reported (Van Rossum et al., 1996, and Chapter 2). Using transcriptome data, we tested for genome-wide heterogeneity of diversity *vs* gene flow in thirteen scenarios of speciation that take into account the history of gene flow in relation with glacial cycles. Finally, taking advantage of the synteny among *Silene* species genomes, we contrast patterns of diversity and divergence in *Silene otites* and *Silene paradoxa* to infer the impact of low recombination regions, assumed to be shared among species, on the pattern of differentiation between *Silene nutans* lineages.

## 4.2 Material and methods

### 4.2.1 Sampling and sequencing

A total of 11 individuals of the Eastern lineage of *S. nutans* were collected in 6 natural populations (Germany, Finland, Grand Duchy of Luxembourg, Belgium, and United-Kingdom, Figure 4.1). For the Western lineage, we sampled a total of 11 individuals in 6 French populations: 7 individuals belonging to W1 group, 2 individuals belonging to the W2 group, and 2 individuals belonging to the W3 group (Figure 4.1). Rosettes collected in natural populations grew in controlled greenhouse conditions (20°C, 16-hour day length) and flower buds were collected to total to 20 to 50 mg of tissue per individual library. RNA was extracted following the Sigma Spectrum protocol. The libraries were sequenced on a Hiseq 2000 (Illumina, Inc.) following an Illumina paired-end protocol (fragment lengths 150-250 bp, 100 bp sequenced from each end). RNA extraction, libraries preparation and sequencing were performed by MGX platform (Montpellier, France).

### 4.2.2 Assembly, mapping, and genotyping

Adaptors and low quality reads were removed using Cutadapt (Martin, 2011) and PRINSEQ (Schmieder and Edwards, 2011). *De novo* transcriptome assembly was performed with TRINITY (Haas et al., 2013) followed by a run of Cap3 (Huang and Madan, 1999) on combined individuals of the most represented population in the cleaned library. Open-Reading frames were predicted using TransDecoder (Haas et al., 2013) on the assembly. Individual reads were



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



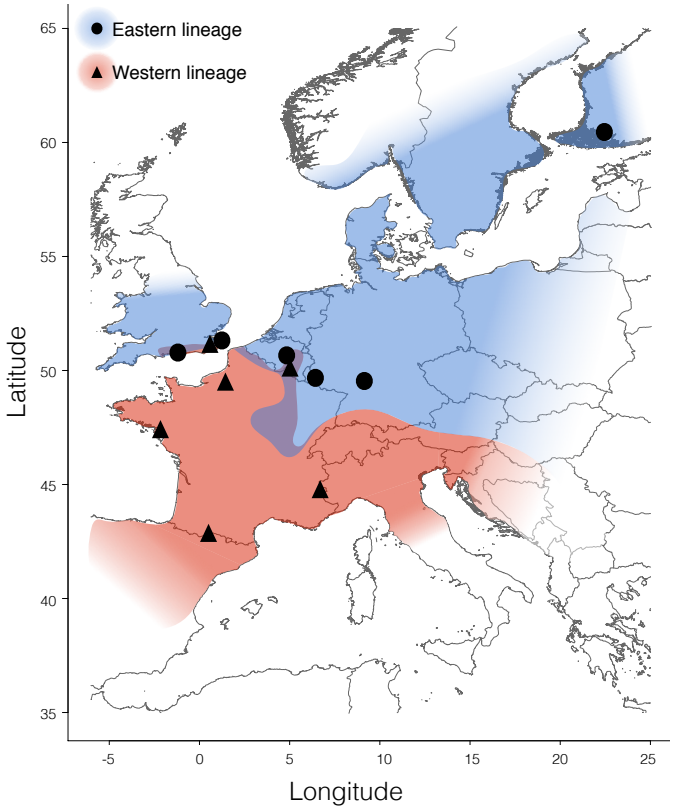



**Figure 4.1:** Geographic distribution of *S. nutans* lineages: 11 samples from the Eastern lineage and 11 samples from the Western lineage.

then mapped on predicted coding regions with bowtie2 (Langmead and Salzberg, 2012). Contigs covered at 2.5X or less across all individuals were discarded. At each position in individual alignments, diploid genotypes were called using the reads2snps program, which implements the method described by Tsagkogeorga et al. (2012) and improved by Gayral et al. (2013). This method estimates the error rate in the maximum likelihood framework, then calculates the posterior probability of each genotype and calls missing data for genotypes supported with probability below than 0.95. Following Burgarella et al. (2015), we required a minimum coverage of 10X per position and per individual to call a genotype. We filtered SNPs for possible hidden paralogs using a likelihood ratio test based on explicit modeling of paralogy (`paraclean` option of reads2snps).

### 4.2.3 Global population genetic statistics

We restricted our analysis to contigs for which all individuals in both lineages were successfully genotyped and with a size ranking from 100 to 5 000 pb after excluding codons containing indels or more than two segregating alleles. For each locus, we computed an array of statistics related to polymorphism and divergence at synonymous sites: genetic diversity measured by both the

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



nucleotide diversity $\pi$ and Watterson's $\theta_w$ estimator, relative measures of differentiation using both the $F_{ST}$ estimated as $1 - \pi_S/\pi_T$, where $\pi_S$ is the mean pairwise diversity for both species and $\pi_T$ is the pairwise diversity calculated from the total alignment), and the "net" interspecific divergence ($d_a$), the Tajima's D (Tajima 1989), the absolute measure of interspecific divergence ($d_{XY}$), the number of biallelic positions exclusively polymorphic in the Western lineage (SxW), in the Eastern lineage (SxE), or shared by both species (Ss), and the number of fixed differences between the two lineages (Sf). This set of summary statistics does not rely on genotype phasing. The average and standard error of each of these statistics were used as summary statistics in the following ABC.

### 4.2.4 Models of speciation simulated

To investigate the demographic history between the two main lineages of *Silene nutans*, we compared six speciation models (Figure 4.2) using an ABC framework (Tavaré et al., 1997, Beaumont et al., 2002). All models considered an instantaneous split at $T_{split}$ of the ancestral population of effective size $N_{anc}$ into two daughter populations of independent constant sizes $N_{east}$ and $N_{west}$. We simulated the standard SI model as well as five other scenarios assuming migration events at different timescales (Figure 4.2). In the Isolation with Migration (IM) model, introgression occurred continuously over time from the ancestral split to the present date. The Ancient Migration (AM) model simulated a period of introgression right after $T_{split}$, followed to a strict isolation, while the Secondary Contact (SC) model assumed a period of isolation right after $T_{split}$, followed to a period of introgression to present. The Periodic Ancient Migration (PAM) model and the Periodic Secondary Contact (PSC) model were similar to the AM and SC model, respectively, with an additional period of introgression.

### 4.2.5 Homogeneous and heterogeneous versions of speciation models

Distinct versions of the six models were developed to test for genome-wide heterogeneity in introgression rates (for the five models involving periods of migration) and in effective size (for the SI model). Migration was simulated using the scaled effective migration rates $M = 4Nm$ ($M_{east}$ is the effective migration rate from Western lineage to Eastern lineage, and $M_{west}$ is the effective migration rate from Eastern lineage to Western lineage), where m is the fraction of a population which is made up of effective migrants from the other population at each generation. For each model involving gene flow, we compared two alternative versions: an homogeneous


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



version ($homo_M$) where m is shared by all loci, and an heterogeneous version ($hetero_M$) where m varied among loci.

Similarly, we tested for heterogeneity in effective size comparing two alternative scenarios: a homogeneous version ($homo_{N_e}$) where all loci shared the same effective size, and an heterogeneous version ($hetero_{N_e}$) where the effective size varied among loci independently among populations (a locus could have a reduced effective size in the Western lineage but not in the Eastern lineage). Finally, if a locus occurs in a low recombination region shared among lineages, its effective size is expected to be lower in the two lineages as well as in the ancestral population. Therefore, we simulated a second heterogeneous version, $hetero_{N_e\_recomb}$ where effective size varied among loci in a similar way as among populations (a locus had a reduced effective size in both lineages and the ancestral population).

### 4.2.6 Simulations and prior distributions

For each version of each model, we simulated $2 \times 10^6$ multilocus data sets, each composed of 900 loci fitting the length and sampling size of 900 observed loci randomly chosen among 4 067 loci. We used large uniform prior distributions for all parameters common to all models.

Population sizes were simulated with respect to the mutation parameter of reference $\theta_{ref} = 4N_{ref}\mu$, with $N_{ref}$ the number of effective individuals of the reference population, arbitrarily set at 100 000, and $\mu$ the mutation rate of $2.76 \times 10^{-8}$ bp/generation. $\theta_{east}/\theta_{ref}$ and $\theta_{west}/\theta_{ref}$ were sampled in a (0 – 10) bounded uniform distribution. Therefore, $N_{east}$ and $N_{west}$ could take values over (0 – 1 000 000). We also integrated intralocus recombination at a rate of $2.763 \times 10^{-8}$

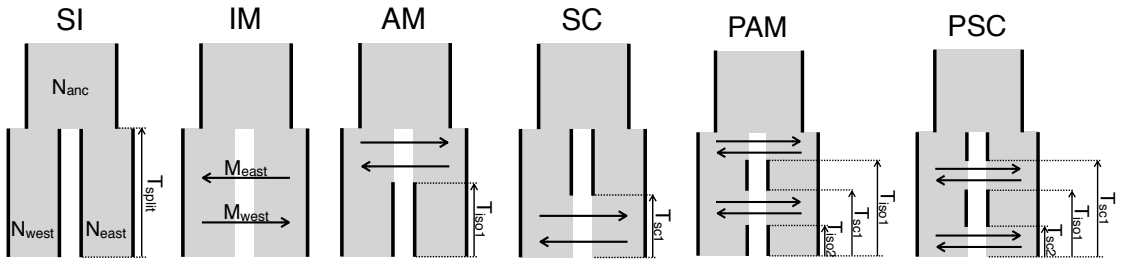



**Figure 4.2:** Alternative scenarios of speciation for Eastern and Western lineage. Six classes of models with different temporal patterns of migration have been compared: SI, IM, AM, SC, PAM and PSC. Four parameters are shared by all models: $T_{split}$ is the number of generations since the speciation time; $N_{anc}$, $N_{east}$, and $N_{west}$ are the number of effective individuals in the ancestral population, Eastern lineage and Western lineage respectively. $T_{iso1}$ and $T_{iso2}$ are the number of generations since the two nascent species have stopped to exchange migrants for the first and second time respectively. $T_{sc1}$ and $T_{sc2}$ are the number of generations since the two daughter species experiment a SC after a period of isolation for the first and second time, respectively. The effective migration rates $M_{east}$, and $M_{west}$ are expressed in 4Nm units, where m is the proportion of a population made up of migrants from the other population per generation.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



bp/generation. The ancestral population size was simulated as $\theta_{anc}/\theta_{ref}$ and has a (0–40) bounded uniform distribution; therefore $N_{anc}$ could take values over (0 – 4 000 000).

The time of speciation was simulated using the $T_{split}/(4N_{ref})$ ratio sampled in a (0 – 3) bounded uniform distribution. Therefore, $T_{split}$ could take values over (0 – 1 200 000) generations; (0 – 3 600 000) years, assuming there is one generation of *S. nutans* in three years in natural populations. This $T_{split}/(4N_{ref})$ parameter conditioned the other time parameters uniformly chosen within: $T_{iso1}<T_{split}$ for the AM model, $T_{iso2}<T_{sc1}<T_{iso1}<T_{split}$ for the PAM model, $T_{sc1}<T_{split}$ for the SC model and $T_{sc2}<T_{iso1}<T_{sc1}<T_{split}$ for the PSC model.

To simulate the $homo_M$ version, two single effective migration rates $M_{east}$ and $M_{west}$ were independently sampled from a (0–10) bounded uniform distribution, and the values were shared by all loci along the transcriptome. For the alternative $hetero_M$ version, 900 – p loci (with p randomly chosen in the (0 – 900) bounded uniform distribution) had their effective migration rates $M_{east}$ and $M_{west}$ reduced by the factor $f_M$. We generated a beta distribution, shaped by the two parameters $\alpha$ and $\beta$ randomly chosen in ($10^{-5}$ – 50), in which we sampled 900 – p times the value of $f_M$. For each multilocus simulation, two beta distributions were generated, one for each lineage.

Similarly, to simulate the $hetero_{N_e}$ version of the SI model, 900 – p loci (with p randomly chosen in the (0 – 900) bounded uniform distribution) had an effective size $\theta/\theta_{ref}$, reduced by the factors $f_{N_e}$. The reduction factor was sampled 900 – p times in a beta distribution shaped by two parameters, $\alpha$ and $\beta$ randomly chosen in ($10^{-5}$ – 50). The parameter p and the beta distribution were independent among the three populations. In this case, the same locus could have a reduced effective compared to the rest of the transcriptome in the Eastern population (linked selected), but not to the rest of the transcriptome in the Western or the ancestral population (neutral, not linked selected).

To simulate the $hetero_{N_e\_recomb}$ version of the SI model, the parameter p and the beta distribution were shared among the three populations. In this case, one locus had an effective size reduced by the same value of the factor $f_{N_e}$ in the three populations.

Prior distributions were computed using a modified version of Priorgen software (Ross-Ibarra et al., 2008, Roux et al., 2013, 2014), and coalescent simulations were run using Msnsam (Ross-Ibarra et al., 2008, Roux et al., 2013, 2014), a modified version of the Ms program (Hudson, 2002), allowing variations in sample size among loci under an infinite site mutation model.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



### 4.2.7 Model selection

We statistically evaluated alternative speciation models using a hierarchical procedure (Roux et al., 2013). First, for each of the six speciation models, we evaluated the posterior probabilities of the two alternative homogeneous ($homo_M$ and $homo_{N_e}$) and heterogeneous ($hetero_M$ and $hetero_{N_e}$) versions. Second, we separately compared the best version of each of the five migration models with the best $homo_{N_e}$ *vs* $hetero_{N_e}$ SI model. Finally, to attest for the impact of recombination in the SI model, we compared the $hetero_{N_e}$ version with the $hetero_{N_e\_recomb}$ version. Posterior probabilities for each candidate model were estimated using a feed-forward neural network implementing nonlinear multivariate regression by considering the model itself as an additional parameter to be inferred under the ABC framework using R package abc (Csilléry et al., 2012). The 0.025% replicate simulations nearest to the observed values for the summary statistics were selected, and these were weighted by an Epanechnikov kernel that peaks when $S_{obs}$=$S_{sim}$. Computations were performed using 50 trained neural networks and 10 hidden networks in the regression.

Models were checked through an analysis of 500 pseudo-observed data sets simulated for each compared model, using random parameter values from the same prior distributions as the analysis performed on the observed data set. For each pairwise model comparison, we applied the above-described model choice procedure to compute the posterior probabilities of the pseudo-observed data set given a model. The relative probability distributions over the 500 replicates were then used to compute the probabilities that the best supported models were correct given the posterior probabilities obtained for the observed data set.

### 4.2.8 Parameter estimation

Posterior distributions of the parameters describing the best model were estimated using a nonlinear regression procedure. Parameters were first transformed according to a log-tangent transformation. Only the 1 000 replicate simulations providing the smallest associated Euclidean distance $\delta = ||S_{obs}-S_{sim}||$ were considered. Then, the joint posterior parameter distribution was obtained by means of weighted nonlinear multivariate regressions of the parameters on the summary statistics. For each regression, 50 feed-forward neural networks and 20 hidden networks were trained using R package abc (Csilléry et al., 2012).

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



### 4.2.9 Inference of the impact of low recombination regions

Loci above the last 5% of the empirical distribution of $F_{ST}$ ere identified as putative outliers of differentiation. We performed Mann-Whitney tests to compare the level of synonymous nucleotide diversity $\pi$, and the absolute interspecific divergence ($d_{XY}$) between outliers and non-outliers in *S. nutans*. Estimation of recombination can be tricky with RNAseq data, as the distance between SNPs within a locus is probably underestimated due to contigs lacking introns. To overcome this problem, we assumed a good synteny among *Silene* species. Therefore, regions of low recombination were assumed to be shared among close species. In this case, if outliers of differentiation are found in regions of low recombination, they should exhibit lower level of polymorphism in *S. nutans* lineages as their orthologs in the two close species *Silene otites* and *Silene paradoxa*.

For eight individuals of *S. otites* and twelve individuals of *S. paradoxa*, we used the above described methodology for transcriptome assembly, read mapping, and genotyping. Orthologous pairs of contigs between *S. nutans* and the two other species were identified using reciprocal best hits on BLASTn results, a hit being considered as valid when alignment length was above 130 bp, sequence similarity above 80%, and a e value below $e^{-50}$ (Gayral et al., 2013, Burgarella et al., 2015). *S. otites* and *S. paradoxa* sequences were added separately to *S. nutans* alignments using a profile-alignment version of MACSE (Ranwez et al., 2011), a program dedicated to the alignment of coding sequences and the detection of frameshift. Contigs were only retained if no frameshift was identified by MACSE. We estimated the $d_{XY}$ between *S. nutans* (grouping individuals from both lineages) and *S. otites*/*S. paradoxa* and $\pi$ in *S. otites*/*S. paradoxa* using synonymous sites. A Mann-Witney test was performed to test for differences between outlier othologs and non-outlier orthologs in both *S. otites* and *S. paradoxa*. Because GC content can be positively correlated to recombination (Serres-Giardi et al., 2012), we also estimated the level of GC3 in outliers of differentiation.

Finally, the genetic map and reference sequences of *Silene latifolia* from Papadopulos et al. (2015) were used to map both the non-outliers and outliers contigs using reciprocal best hits on BLASTn results between the *S. nutans* reference transcriptome and scaffolds of *S. latifolia*. A hit was considered as valid when alignment length was above 130 bp, sequence similarity above 80%, and e value below $e^{-50}$.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



## 4.3 Results

### 4.3.1 Global population genetic statistics

To investigate transcriptome-wide heterogeneity, we obtained a polymorphism data set from 22 individual transcriptomes of *S. nutans* (11 from both lineages). We analyzed 4 067 loci for which polymorphism data were available in both lineages. Levels of synonymous diversity, measured by $\pi$ (Figure 4.3a) or Watterson's $\theta_w$ were significantly different between both lineages (Wilcoxon signed-rank test, $V = 2679670$, $p < 0.001$ for $\pi$, and $V = 1447882$, $p < 0.001$ for $\theta_w$, with the Western lineage being more polymorphic than the Eastern lineage. Accordingly, we found a higher proportion of private polymorphic sites in the Western lineage (49.14% on average) than in the Eastern lineage (29.75% on average). The average absolute divergence between the two lineages was 0.0184, with values ranking from 0.0003 to 0.1242 (Figure 4.3b). We found much heterogeneity in genetic differentiation among contigs with values ranking from 0 to 0.9184 (average= 0.1927, SD= 0.1897, Figure 4.3c). The net interspecific divergence, $d_a$, was ranking from 0 to 0.0967 (average= 0.0059, SD= 0.0082). Such low levels of genetic differentiation were in agreement with the low proportion of fixed polymorphism (1.12% of polymorphic sites on average) compared to the proportion of shared polymorphic sites (19.98% on average). Tajima's D was slightly higher in the Western lineage (mean= 0.5089, SC= 0.8540) compared to the Eastern lineage (mean= 0.2568, SD= 0.9015).

### 4.3.2 Model selection

The shared polymorphism could be explained by incomplete lineage sorting, secondary introgression, or both, with introgression being either homogenous or heterogeneous across loci. We used a model selection approach to identify a relevant speciation model in an ABC framework. We performed coalescent simulations under fourteen models to compute the expected distributions of summary statistics related to polymorphism and divergence patterns. Then, we compared the simulated statistics with the observed ones calculated from the data.

We considered five models involving migration: isolation with migration (IM), ancient migration (AM), periodic ancient migration (PAM), secondary contact (SC), and periodic secondary contact (PSC). Within each model, genome wide heterogeneity (GWH) was tested by comparing two alternative introgression versions that assume either a single introgression rate shared by all loci ($homo_M$) or independently sampled introgression rates across loci ($hetero_M$). In four cases

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



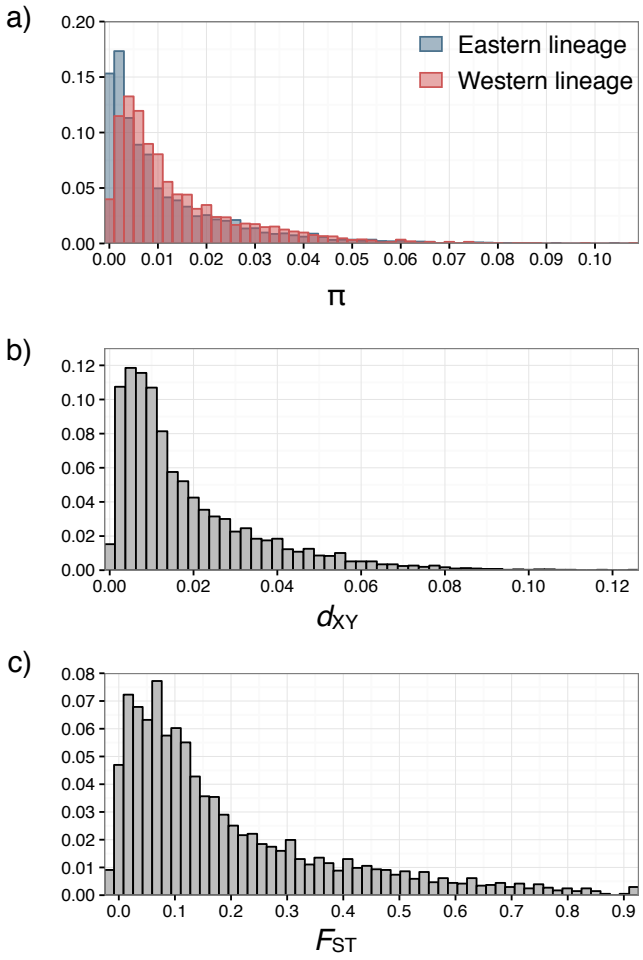



**Figure 4.3:** Patterns of polymorphism and divergence between the Eastern and the Western lineage. Histograms show the empirical distributions of a) the $\pi$ diversity estimator b) the absolute inter-specific divergence $d_{XY}$, and c) the between-species differentiation measured by $F_{ST}$ Each statistic was measured at synonymous positions.

(IM, AM, SC, PSC scenarios), the hetero$_M$ version fit our data best but had a low relative probability to be the correct one given its observed posterior probability (robustness ranking from 0.508 to 0.73, Table 4.1). This approach enabled us to clearly discriminate between the versions of models involving migration. We considered a sixth model without migration (strict isolation scenario, SI) and tested GWH comparing two alternative effective size versions assuming either a single effective size shared by all loci and distinct among lineages and the ancestral population (homo$_{N_e}$) or independently sampled effective size across loci and populations (hetero$_{N_e}$). In this case, the hetero$_{N_e}$ version clearly fit our data set better than the homo$_{N_e}$ version, as supported by the high relative probability that hetero$_{N_e}$ was the correct version given the observed posterior probability (Table 4.1).

We then tested the various hypotheses regarding the history of gene flow between the two lineages using independent comparisons of the best supported version of each scenario involving gene flow against the hetero$_{N_e}$ version of the SI model (Table 4.1). In all cases, the SI


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



model clearly fit our data set better than the models involving gene flow, as supported by the high relative probability that SI model was the correct scenario given the observed posterior probability (Table 4.1).

To take into account the effect of low recombination on the heterogeneity of effective size, we considered a third version of the SI model where a locus experienced a reduction of its effective size shared by the ancestral population as well as the Western and the Eastern lineages ($hetero_{N_e\_recomb}$). When tested against the $hetero_{N_e}$ version of the SI model, the $hetero_{N_e\_recomb}$ fit our data best, as supported by its high relative probability to be the correct scenario given the observed posterior probability (Table 4.1).

### 4.3.3 Parameter estimation

We examined parameter estimates under the best-fitting SI $hetero_{N_e\_recomb}$ model of *S. nutans* lineage divergence. The posterior distributions of parameters built from 1 000 accepted simulations were highly differentiated from their prior distributions, suggesting that the data offered adequate information, and that we appropriately explored the parameter space (Figure S4.1). In this analysis, mutation rate was assumed to be constant over time and across loci, with a value of $2.76 \times 10^{-8}$ bp/generation. Estimates of the effective population sizes suggest an approximately 2.8-fold larger population in the Western lineage (699 415, 95% confidence interval [CI]: 339 975 – 976 419) than in the Eastern lineage (247 896, 95% CI: 84 834 – 763 776). Our analysis

**Table 4.1:** Classification of models of speciation

| | model choice | | Robustness |
|---|---|---|---|
| (1) vs (2) | P(1) | P(2) | |
| IM $hetero_M$ vs IM $homo_M$ | **0.6302** | 0.3698 | 0.541 |
| SC $hetero_M$ vs SC $homo_M$ | **0.9054** | 0.0946 | 0.737 |
| PSC $hetero_M$ vs PSC $homo_M$ | **0.8027** | 0.1973 | 0.653 |
| AM $hetero_M$ vs AM $homo_M$ | 0.4510 | **0.5490** | 0.541 |
| PAM $hetero_M$ vs PAM $homo_M$ | **0.6084** | 0.3916 | 0.508 |
| SI $hetero_{Ne}$ vs SI $homo_{Ne}$ | **0.9996** | 0.0004 | |
| | | | |
| IM $hetero_M$ vs SI $hetero_{Ne}$ | 0.0157 | **0.9843** | 0.998 |
| SC $hetero_M$ vs SI $hetero_{Ne}$ | 0.0039 | **0.9961** | 1 |
| PSC $hetero_M$ vs SI $hetero_{Ne}$ | 0.0106 | **0.9894** | 0.994 |
| AM $homo_M$ vs SI $hetero_{Ne}$ | 0.0015 | **0.9985** | 0.975 |
| PAM $hetero_M$ vs SI $hetero_{Ne}$ | 0.0034 | **0.9966** | 0.997 |
| | | | |
| SI $hetero_{Ne}$ vs SI $hetero_{Ne\_recomb}$ | 0.0420 | **0.9580** | 0.971 |

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



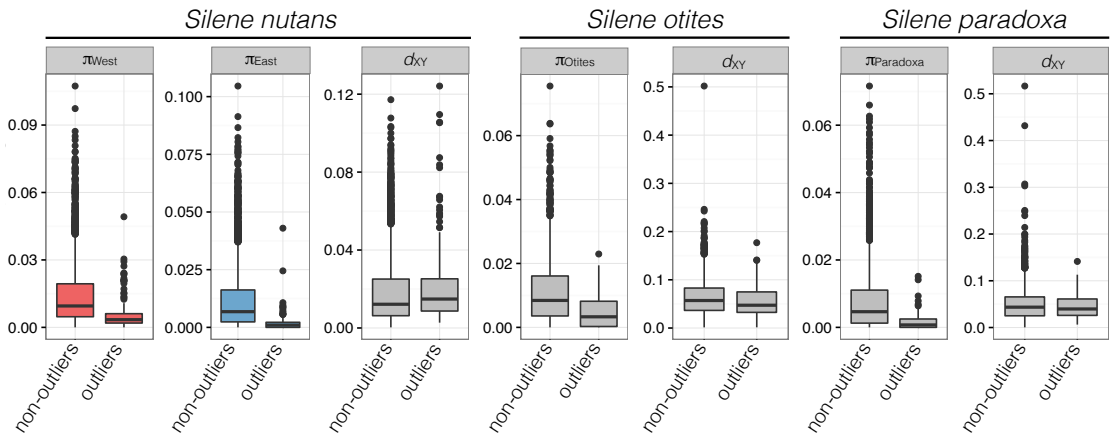



**Figure 4.4:** Levels of within population diversity in *S. nutans* Western lineage ($\pi_{West}$), *S. nutans* Eastern lineage ($\pi_{East}$), *S. otites* ($\pi_{Otites}$), and *S. paradoxa* ($\pi_{Paradoxa}$), and level of divergence $d_{XY}$ between the two lineages of *S. nutans*, between *S. nutans* and *S. otites*, and *S. nutans* and *S. paradoxa*. For each parameter, a comparison was made between non-outliers and outliers of differentiation among *S. nutans* lineages (the last 5% of the $F_{ST}$ istribution).

also suggests that the ancestral population was smaller than the Western lineage (286 466, 95% CI: 222 967 – 396 711). Regarding timing, our estimations indicate that the split in the ancestral *S. nutans* population probably occurred between 62 888 and 122 794 generations (point estimate: 90 509 generations) thus between 188 664 and 368 382 years assuming a 3-year generation.

We then investigated the predicted distribution of the $N_e$ reduction factor c, shared by the two lineages and the ancestral population, that reduced the estimated effective size of a locus. Its distribution was skewed to the left with a mode at 0.165 (95% CI: 0.158 – 0.172), suggesting that linked-selected loci had an effective size six times lower than neutral loci. Among 900 simulated loci, 77.5% exhibited a reduction of $N_e$ (698, 95% CI: 696 – 704).

We then compared observed and simulated summary statistics under a goodness-of-fit procedure from the joint-posterior distribution, and found that the estimated SI hetero$_{N_e_recomb}$ model fit well with the data (Figure S4.2).

### 4.3.4 Inference of the impact of low recombination regions

The five higher percent of the distribution of $F_{ST}$ t formed 203 outliers of differentiation (called hereafter *outliers*, while the rest of the analysed transcriptome will be called *non-outliers*), with levels of differentiation ranking from 0.6139 to 0.91841 (average= 0.7269, SD= 0.0849). We found a reduction of diversity in the outliers in both lineages (mean ranks of the non-outliers and outliers were 0.0117 and 0.0018, respectively; $U = 649\,403$, $Z = 15.7797$, $p < 0.001$, r = 0.2474 for the Eastern lineage; and mean ranks of the non-outliers and outliers were 0.0142 and 0.0054, respectively; $U = 598\,383$, $Z = 12.6443$, $p < 0.001$, r = 0.1983 for the Western lineage, Figure 4.4). We did not find


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



strong support for a difference in levels of divergence between both kinds of loci (mean ranks of the non-outliers and outliers were 0.0182 and 0.0219, respectively; $U = 332\,687.5$, $Z = -3.6493$, $p < 0.001$, r = 0.0572, Figure 4.4).

Half of the contigs of *S. nutans* had an ortholog in *S. otites* (2 184 and 118 contigs were orthologous to the non-outliers and the outliers, respectively) and *S. paradoxa* (2 353 and 122 contigs were orthologous to the non-outliers and the outliers, respectively). We found a reduction of diversity at orthologs of outliers in *S. otites* (mean ranks of the non-outliers and outliers were 0.0111 and 0.0052, respectively; $U = 178\,499.5$, $Z = 7.0618$, $p < 0.001$, r = 0.1472, Figure 4.4) and in *S. paradoxa* (mean ranks of the non-outliers and outliers were 0.0081 and 0.0017, respectively; $U = 219\,306$, $Z = 9.8637$, $p < 0.001$, r = 0.1983, Figure 4.4).

The level of GC3 was similar between outliers and non-outliers in both lineages (mean ranks of the non-outliers and outliers of differentiation were 0.4113 and 0.4178, respectively; $U = 367\,621.5$, $Z = -1.507$, $p = 0.132$, r = 0.0335 for the Eastern lineage; mean ranks of the non-outliers and outliers of differentiation were 0.4110 and 0.4181, respectively; $U = 367\,559.5$, $Z = -1.5108$, $p = 0.131$, r = 0.0237 for the Western lineage).

Finally, although few of our data set were mapped on the *S. latifolia* genetic map (among the 203 outliers, 193 found an ortholog on a *S. latifolia* scaffold, and 30 of them were mapped; among the 3 864 non-outliers, 3 674 had an ortholog on a *S. latifolia* scaffolds, and 720 of them were mapped), non outliers as well as outliers were spread over all linkage groups of *S. latifolia* (Figure S4.3).

## 4.4 Discussion

In order to find out which evolutionary forces led to the speciation process of two *Silene nutans* lineages that exhibit near-complete post-zygotic isolation, and in which time scale, we investigated their evolutionary demographic history.

Using an ABC approach based on transcriptomic data, we showed that the best scenario was that both lineages of *Silene nutans* had not exchanged any gene flow since their split, around 300 000 years ago. This estimated date coincides with the early stage of the Riss glacial period. It is in agreement with the geographical distribution of the lineages that reflect post-glacial recolonization pathways in Europe (Martin et al., 2016). In addition, it confirms the hypothesis of allopatric speciation deduced from results of inter-lineage crosses from different geographical

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



regions (see chapter 2). Despite this recent history of speciation, *S. nutans* lineages show strong reproductive isolation (Van Rossum et al., 1996; Chapter 2). This raises the question of the nature of genetic incompatibilities that rapidly accumulated and could result from cytonuclear genetic conflict (Chapter 2). This question could be approached by analyzing organelle gene diversity in order to search for co-evolution between organelle and nuclear encoded proteins within the lineages (de Juan et al., 2013, Burton et al., 2013). Interestingly, both lineages of *S. nutans* globally look alike (leading to the single-species designation by Jeanmonod and Bocquet, 1983), although the Western lineage morphologically differs from the Eastern lineage on their capsule size and leave shape (Hepper, 1951, Van Rossum et al., 1996). The strong reproductive isolation and the mild morphological differentiation are in line with the theoretical expectations of Orr (1995): since reproductive isolation increases faster than linearly over time due to epistasis (the so-called snowball effect), it could precede the evolution of morphological or physiological differences (Moyle and Nakazato, 2010).

Although we found, on average, a low genetic differentiation between the Western and the Eastern lineages (average $F_{ST} = 0.124$), the range of $F_{ST}$ values was wide (min= 0, max= 0.918, Figure 4.3c). We found that this heterogeneity of genetic differentiation was more likely driven by heterogeneity of effective size along the transcriptome than by heterogeneity of gene flow (Table 4.1). Our results suggest that each lineage, strictly isolated since the split of the ancestral population, was independently subject to natural selection that lowered levels of synonymous diversity at linked sites. Reducing levels of diversity mechanically increases levels of genetic differentiation, as any differences between populations are nearly fixed (Charlesworth, 1998). On the contrary, the $d_{XY}$ divergence is not influenced by within population diversity, but reflects both the diversity of the ancestral population and the differences accumulated since the ancestral population split (Nei and Li, 1979). As the whole transcriptome started to diverge at the same time, since the ancestral population split, and without migration, $d_{XY}$ should be similar along the genome, and reflect the diversity in the ancestral populations (see Cruickshank and Hahn, 2014). In agreement with this scenario, we found a reduced synonymous diversity in the outliers of differentiation compared to the rest of the transcriptome, while divergence was barely affected (Figure 4.4).

The reduction of effective size is the result of the action of linked selection along the genome. In a recent study, Elyashiv et al. (2016) predicted that linked selection reduced diversity levels by 77% – 89% along the genome of *Drosophila melanogaster*. Such a drastic reduction in diversity meets our estimates in *S. nutans*, with a predicted reduction of effective size by ~83.5% (Figure S4.1f). Both positive (selective sweep) and negative (background selection) se-

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



lection are possibly involved in linked selection (Maynard-Smith and Haigh, 1974, Charlesworth et al., 1993, Cutter and Payseur, 2013). Quantifying the relative parts of selective sweep and background selection is still a challenge, even though most evidence points out a predominant role of background selection (Josephs and Wright, 2016). Therefore, when the lineages were isolated in distinct glacial refugees, the adaptation of each lineage to different environments may not be the main evolutionary force that shaped the landscape of differentiation.

Our best-fitting scenario modeled a reduction of effective size of the same loci in both the lineages and the ancestral populations. Such a scenario is expected if loci are found in low recombination regions conserved among populations. Burri et al. (2015) sequenced the genome of five related species of *Ficedula* flycatchers and showed a good conservation of the recombination landscape among species. They showed that in regions of low recombination, all species had reduced diversity. Following the same idea, we analysed the diversity of the orthologous loci of *S. nutans* outliers in two other close *Silene* species, *S. otites* and *S. paradoxa*, under the hypothesis that they shared the same regions of low recombination. We indeed found a reduction of diversity of these outlier orthologs in *S. otites* and *S. paradoxa*. The conserved number of chromosomes (2n = 24, Široký et al., 2001, Yildiz et al., 2008, Minareci et al., 2009) and small-scale synteny between the pseudo-autosomal region of *S. latifolia* and its autosomal ortholog in *S. vulgaris* (Blavet et al., 2012) back up our hypotheses of rough synteny among *Silene* species (see Chapter 4 of this thesis), although local inversion in *S. latifolia* (Filatov, 2005) and reciprocal translocation in *S. diclinis* (Howell et al., 2009) occurred during the evolution of their sex chromosomes. Taking advantage of this rough synteny, we used the genetic map of *S. latifolia* (Papadopulos et al., 2015) to map our analyzed contigs. Contrary to what is expected under an IM scenario (Turner et al., 2005), outliers were not clustered into few large discrete regions, but spread along the genome, with the present limitation that a large portion of them could not be mapped.

The impact of linked selection and low recombination regions in shaping islands of differentiation has received an increasing interest in the past few years (Noor and Bennett, 2009, Turner and Hahn, 2010, Nachman and Payseur, 2012, Cruickshank and Hahn, 2014), with recent genomic data fitting well with this scenario (Burri et al., 2015, Wang et al., 2016). Nevertheless, contrary to the alternative model involving heterogeneous levels of migration along the genome, islands of differentiation do not necessarily coincide with regions involved in reproductive isolation. Therefore, the genetic architecture of reproductive isolation is difficult to identify. Nevertheless, quick lineage sorting in low recombining regions could increase the potential for negative epistatic interactions to arise, leading to Bateson-Dobzansky-Müller genetic incompat-

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



ibilities and post-zygotic reproductive isolation. In this scenario, reproductive isolation would be a by-product, and would not be selected for, contrarily to the classic view of speciation with gene flow initiated by differential adaptation (Wu, 2001).

## Acknowledgements

Numerical results presented in this paper were carried out using the HPC service of the Centre de Ressources Informatiques (CRI) of Lille 1 University. We thank the technical staff for providing the technical support and infrastructure. This work was supported by the Agence Nationale de la Recherche (ANR-11-BSV7-013-03) to PT and XV and a PhD fellowship from the French Research Ministry to HM.

## 4.5 Bibliography

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002). Approximate Bayesian Computation in population genetics. *Genetics*, 162:2025–2035.

Bierne, N., Welch, J., Loire, E., Bonhomme, F., and David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, 20:2044–2072.

Blavet, N., Blavet, H., Čegan, R., Zemp, N., Zdanska, J., Janoušek, B., Hobza, R., and Widmer, A. (2012). Comparative analysis of a plant pseudoautosomal region (PAR) in *Silene latifolia* with the corresponding *S. vulgaris* autosome. *BMC Genomics*, 13:226–235.

Burgarella, C., Gayral, P., Ballenghien, M., Bernard, A., David, P., Jarne, P., Correa, A., Hurtrez-Boussès, S., Escobar, J., Galtier, N., and Glémin, S. (2015). Molecular evolution of freshwater snails with contrasting mating systems. *Molecular Biology and Evolution*, 32:2403–2416.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružic, M., Sæther, S. A., Sætre, G. P., Török, J., and Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research*, 25:1656–1665.

Burton, R. S., Pereira, R. J., and Barreto, F. S. (2013). Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematics*, 44:281–302.

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, 15:538–543.

Charlesworth, B., Morgan, M. T., and Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134:1289–1303.

Coyne, J. A. and Orr, H. A. (2004). *Speciation*. Sinauer associates, Sunderland, Massachusetts, USA, 1st edition.

Cruickshank, T. E. and Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23:3133–3157.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Csilléry, K., François, O., and Blum, M. G. B. (2012). abc: an R package for Approximate Bayesian Computation (ABC). *Methods in Ecology and Evolution*, 3:475–479.

Cutter, A. D. and Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*, 14:262–274.

de Juan, D., Pazos, F., and Valencia, A. (2013). Emerging methods in protein co-evolution. *Nature Reviews Genetics*, 14:249–261.

Elyashiv, E., Sattath, S., Hu, T. T., Strustovsky, A., McVicker, G., Andolfatto, P., Coop, G., and Sella, G. (2016). A genomic map of the effects of linked selection in *Drosophila*. *PLoS Genetics*, 12:e1006130.

Filatov, D. A. (2005). Evolutionary history of *Silene latifolia* sex chromosomes revealed by genetic mapping of four genes. *Genetics*, 170:975–979.

Gagnaire, P.-A., Pavey, S. A., Normandeau, E., and Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by rad sequencing. *Evolution*, 67:2483–2497.

Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., Lourenco, J. M., Alves, P. C., Ballenghien, M., Faivre, N., Belkhir, K., Cahais, V., Loire, E., Bernard, A., and Galtier, N. (2013). Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genetics*, 9:e1003457.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., and Regev, A. (2013). De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocols*, 8:1494–1512.

Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research*, 16:730–737.

Hepper, F. N. (1951). The variations of *Silene nutans* L. in Great Britain. *Watsonia*, 2:80–90.

Howell, E. C., Armstrong, S. J., and Filatov, D. A. (2009). Evolution of neo-sex chromosomes in *Silene diclinis*. *Genetics*, 182:1109–1115.

Huang, X. and Madan, A. (1999). CAP 3: a DNA sequence assembly program. *Genome Research*, 9:868–877.

Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics Applications Note*, 18:337–338.

Jeanmonod, D. and Bocquet, G. (1983). Propositions pour un traitement biosystématique du *Silene nutans* L. Caryophyllaceae. *Candollea*, 38:267–295.

Josephs, E. B. and Wright, S. I. (2016). On the trail of linked selection. *PLOS Genetics*, 12:e1006240.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9:357–359.

Martin, H., Touzet, P., Van Rossum, F., Delalande, D., and Arnaud, J.-F. (2016). Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity*, 116:286–294.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17:10–12.

Maynard-Smith, J. and Haigh, J. (1974). The hitch-hiking effect of a favorable gene. *Genetics Research*, 23:23–35.

Minareci, E., Yildiz, K., and Çirpici, A. (2009). Karyological studies in some species of the genus *Silene* (Caryophyllaceae). *Cytologia*, 74:245–252.

Moyle, L. C. and Nakazato, T. (2010). Hybrid incompatibility "snowballs" between *Solanum* species. *Science*, 329:1521–1523.

Nachman, M. W. and Payseur, B. A. (2012). Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367:409–421.

Nei, M. and Li, W.-H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76:5269–5273.

Noor, M. A. F. and Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, 103:439–444.

Orr, H. A. (1995). Population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*, 139:1805–1813.

Orr, H. A. and Presgraves, D. C. (2000). Speciation by postzygotic isolation: forces, genes and molecules. *BioEssays*, 22:1085–1094.

Papadopulos, A. S. T., Chester, M., Ridout, K., and Filatov, D. A. (2015). Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proceedings of the National Academy of Sciences*, 112:13021–13026.

Ranwez, V., Harispe, S., Delsuc, F., and Douzery, E. J. P. (2011). MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One*, 6:e22594.

Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., Bowers, J. E., Burke, J. M., and Rieseberg, L. H. (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature communications*, 4:1827.

Ross-Ibarra, J., Wright, S. I., Foxe, J. P., Kawabe, A., DeRose-Wilson, L., Gos, G., Charlesworth, D., and Gaut, B. S. (2008). Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS One*, 3:e2411.

Roux, C., Fraïsse, C., Castric, V., Vekemans, X., Pogson, G. H., and Bierne, N. (2014). Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *Journal of Evolutionary Biology*, 27:1662–1675.

Roux, C., Tsagkogeorga, G., Bierne, N., and Galtier, N. (2013). Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, 30:1574–1587.

Schmieder, R. and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27:863–864.

Serres-Giardi, L., Belkhir, K., David, J., and Glémin, S. (2012). Patterns and evolution of nucleotide landscapes in seed plants. *Plant Cell*, 24:1379–1397.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Široký, J., Lysák, M. A., Doležel, J., Kejnovský, E., and Vyskot, B. (2001). Heterogeneity of rDNA distribution and genome size in *Silene* spp. *Chromosome Research*, 9:387–393.

Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, 145:505–518.

Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R. S. T., Hecht, J., Knaust, F., Belkhir, K., Klages, S., Dieterich, R., Stueber, K., Piferrer, F., Guinand, B., Bierne, N., Volckaert, F. A. M., Bargelloni, L., Power, D. M., Bonhomme, F., Canario, A. V. M., and Reinhardt, R. (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications*, 5:5770.

Tsagkogeorga, G., Cahais, V., and Galtier, N. (2012). The population genomics of a fast evolver: high levels of diversity, functional constraints, and molecular adaptation in the Tunicate *Ciona intestinalis*. *Genome Biology and Evolution*, 4:852–861.

Turner, T. L. and Hahn, M. W. (2010). Genomic islands of speciation or genomic islands and speciation? *Molecular Ecology*, 19:848–850.

Turner, T. L., Hahn, M. W., and Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, 3:1572–1578.

Van Rossum, F., De Bilde, J., and Lefèbvre, C. (1996). Barriers to hybridization in calcicolous and silicolous populations of *Silene nutans* from Belgium. *Belgian Journal of Botany*, 129:13–18.

Van Rossum, F., Weidema, I. R., Martin, H., Le Cadre, S., Touzet, P., Prentice, H. C., and Philipp, M. (2016). The structure of allozyme variation in *Silene nutans* (Caryophyllaceae) in Denmark and in north-western Europe. *Plant Systematics and Evolution*, 302:23–40.

Wang, J., Street, N. R., Scofield, D. G., and Ingvarsson, P. K. (2016). Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American Aspens. *Molecular Biology and Evolution*, 33:1754–1767.

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14:851–865.

Yildiz, K., Minareci, E., Çirpici, A., and Dadandi, M. Y. (2008). A karyotypic study on *Silene*, section *Siphonomorpha* species of Turkey. *Nordic Journal of Botany*, 26:368–374.

## 4.6 Appendix

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



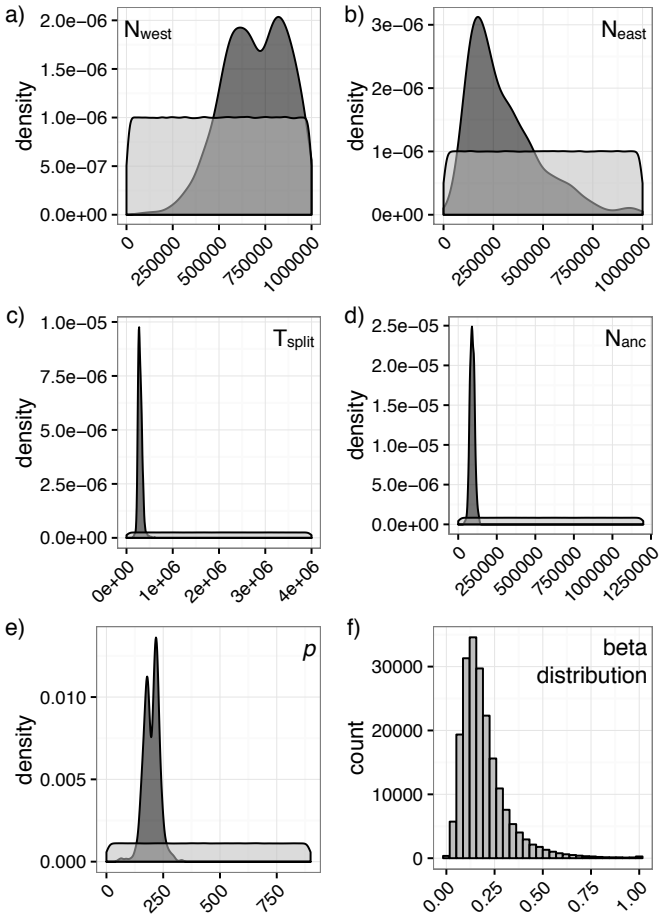



**Figure S4.1:** Prior (light grey) and posterior (dark grey) distribution of the parameters describing the $hetero_{N_e\_recomb}$ version of the SI model: a) $N_{west}$ the western population size, b) $N_{east}$ the eastern population size, c) $N_{anc}$ the ancestral population size, d) $T_{split}$ the time of the ancestral population split, in number of generation, e) p the number of loci that were not concerned by the reduction of effective size. d) posterior distribution of the reducing factor $f_{N_e}$.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

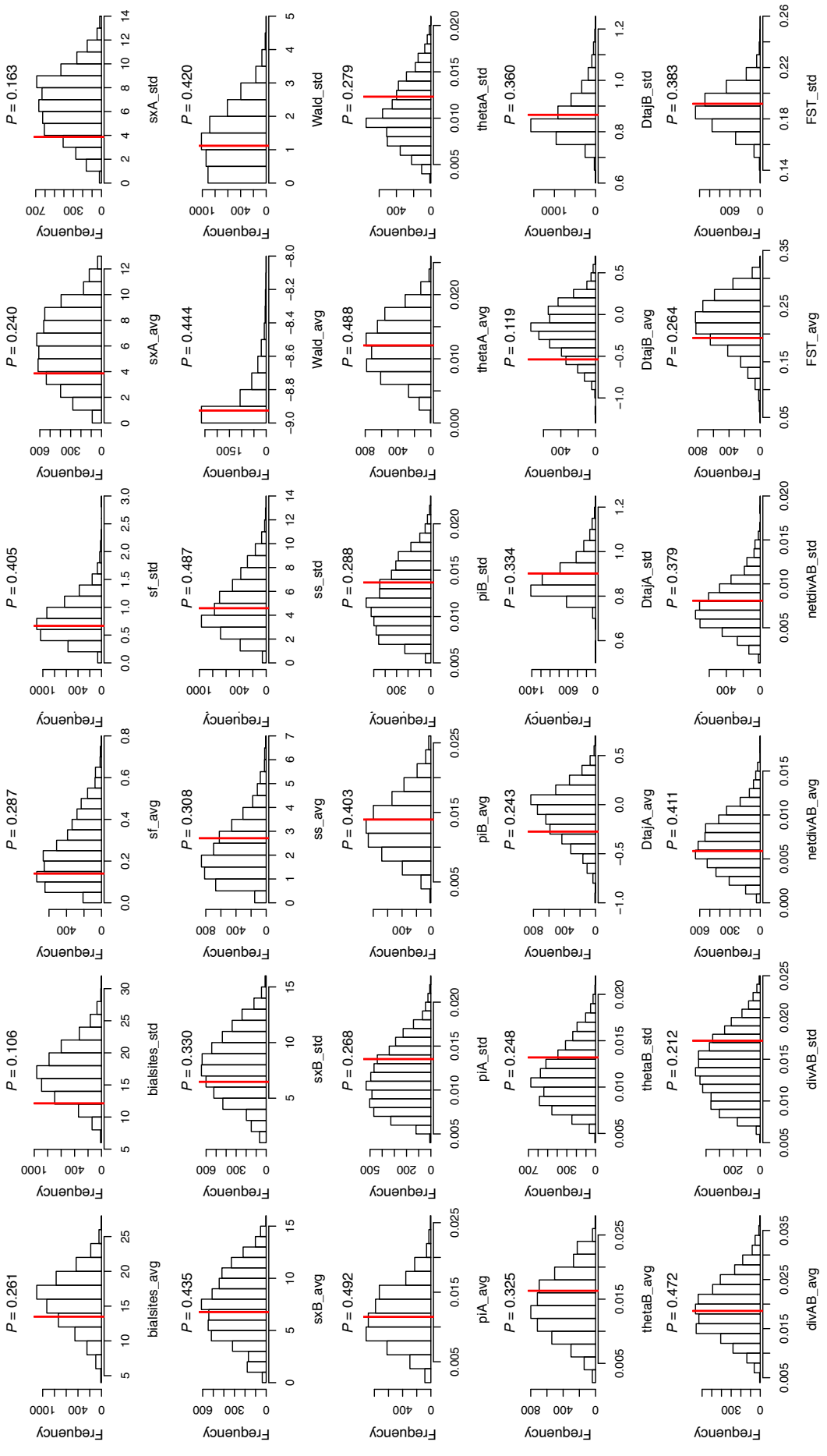



**Figure S4.2:** Goodness of fit of the SI hetero$_{N_e_recomb}$ scenario of speciation based on 5 000 simulations from the joint posterior distribution. The red line represent the observed value of the statistic.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



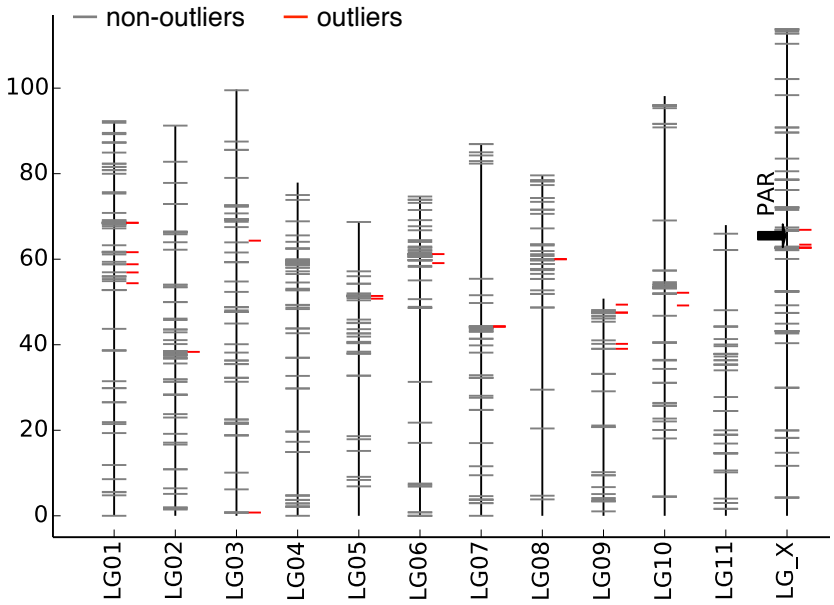



**Figure S4.3:** Localization of non-outliers (in grey) and outliers (in red) of differentiation between the two lineages of *S. nutans* on the genetic map of *Silene latifolia* (Papadopulos et al., 2015). PAR = Pseudo-Autosomal Region.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

# **Chapitre 4** : Evolution of sex chromosomes in the *Otites* section

**Hélène Martin, Fantin Carpentier, Sophie Gallina, Aline Muyle, Gabriel Marais et Pascal Touzet**

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



Sansome (1938) propose que l'architecture génétique de l'isolement reproducteur (biais de sex-ratio) entre *S. pseudotites* et *S. otites* est lié à différents déterminisme du sexe entre les deux espèces. Dans ce chapitre nous avons testé son hypothèse en utilisant une récente méthode probabilistique implémenté dans le programme SEX-Detector. Nous avons identifié, pour *S. pseudotites*, un déterminisme du sexe de type XY. Pour *S. otites*, nous avons identifié deux paires de chromosomes l'une présentant une ségrégation de type XY et l'autre de type ZW.

Une partie des données présente dans ce chapitre a été généré avant cette thèse. Lors de cette thèse, j'ai réalisé les croisements, traité les données brutes et lancé les analyses Sex-Detector. Fantin Carpentier, dont j'ai participé à l'encadrement du stage de M2, a réalisé les premières analyses. J'ai ensuite repris les données, continué les analyses et écrit ce chapitre en incorporant les corrections et remarques des co-auteurs.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



# Evolution of sex chromosomes in the *Otites* section

## 5.1 Introduction

Sex determination is controlled by diverse mechanisms that may evolve rapidly in many taxa (Bachtrog et al., 2014). In genotypic sex determination mechanisms, male (XY) and female (ZW) heterogamety or polygenic systems are described (Charlesworth and Mank, 2010, Moore and Roberts, 2013). Transitions from a system to another, as well as sex chromosome turnovers, have been reported in several groups (e.g., Ross et al., 2009) and are particularly likely when the ancestral sex chromosome exhibits little genetic differentiation (Bachtrog et al., 2014). With the increased availability of genetic markers in non-model organisms, hypotheses of sex chromosome evolution and lability can be tested, and broader patterns can be inferred (Charlesworth and Mank, 2010).

Sex chromosomes evolved from a pair of autosomes carrying sex-determining loci that stopped recombining with each other (Bergero and Charlesworth, 2009). Consequently, sex chromosomes differentiated from each other, and the sex-specific chromosome (Y or W) started to degenerate because of the strong reduction of its effective population size, $N_e$, as well as Hill-Robertson effects (Charlesworth and Charlesworth, 2000, Bachtrog, 2006). Under a neutral model, the $N_e$ of the Y (and W) chromosomes is one quarter of that of the autosomes, while the $N_e$ of the X and Z) chromosomes represents three quarters of that of the autosomes. Nevertheless, sexual selection and mating systems can change these ratios and differentially impact the polymorphism of the X, Z, Y, and W chromosomes (Ellegren, 2011). In the case of polygyny, the $N_e$ of the X chromosome tends to increase while the $N_e$ of the Z tends to decrease (Ellegren, 2009, Bachtrog et al., 2011).

In extreme cases, the degeneration of sex-specific chromosomes leads to male monosomy of the X chromosome, or to female monosomy of the Z. Therefore, the heterogametic sex is expected to upregulate the X or Z chromosome to compensate its monosomy and reach the ancestral diploid expression level. This dosage compensation can occur rapidly during sex chromosome evolution (Muyle et al., 2012) and can be partial or complete (Mank, 2013). In particular, sexual selection and strong reproductive variance in males can slow down the evolution of dosage compensation in a female heterogametic system (Mullon et al., 2015).

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



In plants, sex chromosomes are present in 48 species across 20 families and are found in various forms: homomorphic, heteromorphic, heterogamety, and multiple sex chromosome systems (Ming et al., 2011). Sex chromosome turnover is also found between sister genera (Hou et al., 2015), as well as within genus (Geraldes et al., 2015). In the genus *Silene*, sex chromosomes evolved independently in at least two sections from a non-dioecious ancestor (Desfeux et al., 1996, Marais et al., 2011). The *Melandrium* section includes, among others, the well-studied species *Silene latifolia* and its heteromorphic XY chromosome (Bernasconi et al., 2009), but also *Silene diclinis*, with its multiple sex chromosomes X, $Y_1$, and $Y_2$ that result from a reciprocal translocation between the ancestral Y and an autosome (Howell et al., 2009). In the *Otites* subsection, a different linkage group (LG) than the one in *S. latifolia* seems to have been recruited to form the homomorphic XY chromosome of *Silene colpophylla* (Mrackova et al., 2008). In the same subsection, Sansome (1938) performed interspecies crosses and assigned an XY sex determinism to *S. pseudotites* and a ZW sex determinism to *S. otites*. A few years later, Warmke (1942) questioned the sex determinism of *S. otites*, as crosses involving autotetraploid *S. otites* led him to conclude a XY sex determinism. Recently, a study based on the genetic mapping of AFLP markers in *S. otites* concluded in favor of Sansome; i.e., female heterogamy (Slancarova et al., 2013). Nevertheless, the few AFLP markers and the important lability of sex chromosomes in the genus call for a confirmation.

In this study, we used RNA-seq data from controlled crosses of *S. otites* and *S. pseudotites* to identify sex-linked genes using a recent probabilistic approach implemented in SEXDETector (Muyle et al., 2016). We used the genetic map and sequences of *S. latifolia* (Papadopulos et al., 2015) to infer the autosomal origin of sex-linked genes. Finally, using samples from natural populations and the non-dioecious *S. nutans* as an outgroup, we tested the expected consequences of sex chromosome evolution by estimating the divergence, polymorphism, and expression level of sex-linked genes in both species.

## 5.2 Material and Methods

### 5.2.1 Plant material and sequencing

To identify sex-linked genes, we generated RNA-seq data from a controlled cross per species. Parents came from seeds sampled in natural populations in France. They were sown and grown in a greenhouse with standard conditions (20°C, 16-hour day length). Total RNA was extracted from flower buds of parents, five males, and five females of each progeny, using the Spectrum



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Plant Total RNA kit (Sigma Inc., USA), and following the manufacturer's protocol. The total RNA was also treated with a DNAse. Libraries were prepared with the TruSeq RNA Strand Specific Preparation kit (Illumina Inc., USA). Each cDNA library was sequenced using a paired-end protocol on a HiSeq2500 sequencer, producing 100 bp reads (6 libraries pooled in equi-proportion per lane). Demultiplexing was performed using CASAVA 1.8.1 (Illumina) to produce paired sequence files containing reads for each sample in the Illumina FASTQ format. RNA extraction and sequencing were done by the sequencing platform in the AGAP laboratory (Montpellier, France). To estimate polymorphism and divergence, we used RNA-seq data from flower buds of individuals sampled in natural populations (different populations from the parents) and grown in the greenhouse in the same standard conditions. Two males and two females of *S. otites*, as well as one *S. pseudo otites* female, were added. One hermaphrodite *S. nutans* individual was used as an outgroup. To test for dosage compensation, we added RNA-seq data from the flower buds of five more hermaphrodite *S. nutans* individuals sampled in natural populations and grown in the greenhouse in the same standard conditions, totaling six hermaphrodites.

### 5.2.2 RNA-seq cleaning, assembly, and genotyping

We filtered the raw data from the *S. otites* and *S. pseudotites* families using Cutadapt (Martin, 2011) to remove sequencing adapters, as well as PRINSEQ (Schmieder and Edwards, 2011) to remove low-quality reads and poly-A tails. We assembled *de novo* reference transcriptomes by mixing all members of a family using Trinity (Haas et al., 2013) and a subsequent CAP3 run (Huang and Madan, 1999) with the default settings. Finally, we used TransDecoder (Haas et al., 2013) to predict coding sequences. We mapped reads from each member of a family onto their reference transcriptome using Bowtie2 (Langmead and Salzberg, 2012). SAM files were compressed and sorted using SAMTOOLS (Li et al., 2009). At each position in individual alignments, diploid genotypes were called using the reads2snps program, which implements the method described by Tsagkogeorga et al. (2012) and improved by Gayral et al. (2013). Following Burgarella et al. (2015), we required a minimum coverage of 10X per position and per individual to call a genotype. We did not clean data for paralogous SNPs, as X/Y SNPs tend to be filtered out by the *paraclean* option of reads2snps that uses a likelihood ratio test based on explicit modeling of paralogy.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



### 5.2.3 Identification of sex-linked genes and autosomal origin

To test for the sex determination system (i.e., no sex chromosomes, Z/W system, or X/Y system), and to infer a segregation type for contigs, we used the recent probabilistic methods implemented in SEX-DETector (Muyle et al., 2016). These methods are based on the genotypes of two parents and ten of their progeny (five males and five females), from which the segregation type of each contig is inferred. Each possible sex determination system yields a model, with the assumption that SNPs can be transmitted to the progeny by four segregation modes:

(i) autosomal;

(ii) sex-linked with both X and Y (or Z and W) alleles present;

(iii) X (or Z) hemizygous; i.e., sex-linked with only the X (or Z) allele present (the Y or W allel being inactivated, lost, too weakly expressed, or in a different contig due to X/Y or Z/W divergence);

(iv) undefined; e.g., non-polymorphic contigs.

A model is selected using the Bayesian Information Criterion (BIC), and its parameters are then estimated using an EM algorithm. Contig segregation type is then inferred by assessing the *posterior* probability of each segregation type for each SNP given the observed genotype data, and by averaging SNP posteriors. We filtered posterior segregation type probabilities to be higher than 0.9. We used the genetic map and reference sequences of *Silene latifolia* from Papadopulos et al. (2015) to map ORFs of both speciaes using reciprocal best hits on BLASTn results between ORFs of both species and scaffolds of *S. latifolia*. A hit was considered valid when alignment length was above 130 bp, sequence similarity above 80%, and e value below $e^{-50}$.

### 5.2.4 X-Y and Z-W divergence

We used the putative X-, Y-, Z- and W-linked sequences inferred by SEX-DETector to test whether the Y- or W-linked copies had incorporated more changes than the X- or Z-linked alleles, as expected if the Y or W is degenerated. We used *S. nutans* sequences from a previous study as outgroups. We identified orthologs of focal species and *S. nutans* sequences from reciprocal best hits on BLASTn results, and we aligned them using MACSE (Ranwez et al., 2011). We calculated the number of synonymous ($d_S$) and nonsynonymous ($d_N$) substitutions per site using the



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



PopPhyl tool dNdSpiNpiS (Gayral et al., 2013). While only one sequence of Y (from the father) or W (from the mother) can be reconstructed, up to three alleles can be identified for X- or Z-linked contigs. To avoid bias due to the number of sequences used to calculate these statistics, we randomly chose only one sequence of the putative X or Z. Finally, we estimated synonymous divergence between the X and Y haplotype contigs ($dS_{XY}$) using dNdSpiNpiS.

### 5.2.5 Sex-linked *vs.* autosomal polymorphism divergence and selection efficacy

To test whether sex-linked contigs had less polymorphism than autosomal contigs due to their smaller effective size ($N_e$), we focused on samples from natural populations obtained in a previous study (two female *S. pseudotites* individuals and eight samples—four males and four females—for *S. otites*). For natural population samples, reads2snps randomly attributed alleles to sequences (sequences do not represent haplotypes). To avoid misattribution of alleles of sex-linked contigs in the heterogametic sex (X allele attributed to the Y allele and *vice versa* in males under the XY system, and Z allele attributed to the W allele and *vice versa* under the ZW system), we only focused on the homogametic sex (female XX and male ZZ). Orthologous sequences from natural population samples and *S. nutans* (as outgroup) were identified from reciprocal best hits on BLASTn results and added to the family alignment using MACSE. For each gene, we calculated per-site synonymous ($\pi_S$) and nonsynonymous ($\pi_N$) diversity in *S. pseudotites* and *S. otites* using dNdSpiNpiS (Gayral et al., 2013), on a minimum of two individuals per focal species.

The global strength of purifying selection was evaluated through the $\pi_N/\pi_S$ ratio. Because expression level is usually found to affect the intensity of purifying selection (e.g., Burgarella et al., 2015), we then controlled for the effect of expression level on $\pi_N/\pi_S$. To have sufficient polymorphism, we grouped contigs into classes of expression containing a total of 100 SNPs and we computed the mean $\pi_N/\pi_S$ for each class. We used the average expression coverage per gene as a proxy for the expression level. For each individual, the sum of the length of matching reads was divided by the contig length and the library size of the individuals before multiplying the result by $10^9$ to obtain a normalized expression level in RPKM. Average coverage per gene was then calculated by averaging over individuals. To test for segregation type (autosomal and X-linked for *S. pseudotites*; autosomal, X-linked, and Z-linked for *S. otites*) differences in $\pi_N/\pi_S$ while controlling for expression level, we performed an ANCOVA including segregation type, expression level, and their interaction.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



### 5.2.6 Dosage compensation

We studied dosage compensation in *S. pseudotites* and *S. otites* by comparing the expression level of sex-unbiased sex-linked contigs of the focal species with the expression level of the orthologs in *S. nutans*. If dosage compensation occurs, the expression in males should be equivalent to the expression in females, while the allele expression in the heterogametic sex should be different, the Y or the W being lower than the X or the Z. To be able to separately study the X, Y, Z, and W allele expression, expression was examined at the SNP level.

To identify sex-unbiased contigs, we first computed the average expression over SNPs for each contig by dividing the sum of the number of reads at each SNP within a contig by the number of SNPs within a contig. The average expression was then normalized using the TMM normalization method implemented in edgeR (Robinson et al., 2010). A QL F-test was done to test for differential expression between sexes, also using edgeR. A contig was considered biased if its expression is significantly at least doubled in one sex compared to the other one (fold-changes significantly greater than two).

We focused on the average expression over SNPs of contigs of the 6 females of a family (the mother and its 5 daughters), the 6 males of a family (the father and its 5 sons) and 6 hermaphroditic individuals of *S. nutans*. For each contig, we computed the following $\log_2$ ratios: (Xm+Ym):XXf, which should be close to 0 as we worked on sex-unbiased contigs; Xm:XXf, which should be greater than -1 if dosage compensation occurs; and Ym:Xm as an estimate of Y degeneration. To estimate Y degeneration and possible subsequent dosage compensation in males more accurately, we also used *S. nutans* expression of sex-linked orthologs as a proxy of ancestral autosomal expression (Mank 2013), and thus computed the following $\log_2$ ratios: XXf:AAnut, (Xm+Ym):AAnut, Xm:AAnut, and Ym:AAnut. In the same way, for Z/W contigs, we computed the following $\log_2$ ratios: (Zf+Wf):ZZm, Zf:ZZm, Wf:ZZm, Wf:Zf, ZZm:AAnut, (Zf+Wf):AAnut, Zf:AAnut and Wf:AAnut.

## 5.3 Results

### 5.3.1 Sex determination in *S. pseudotites* and *S. otites*

The probabilistic methods implemented in SEX-DETector allowed us to test for the presence of sex chromosomes in the data (either an X/Y or a Z/W system) using the BIC. For both species,

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



the BIC was higher for the model without sex chromosomes, suggesting the presence of sex chromosomes (Table 5.1). Hereafter, we considered that *S. pseudotites* had an XY system. Over the 174 contigs inferred as sex-linked, 167 contigs had an ortholog on *S. latifolia* scaffolds of Papadopulos et al. (2015), and 44 of them were mapped on their genetic map. In the same way, over 7,233 contigs inferred as autosomes in *S. pseudotites*, 6,668 contigs had an ortholog on *S. latifolia* scaffolds, and 1,154 of them were mapped. While inferred autosomal contigs of *S. pseudotites* were spead over all linkage groups of *S. latifolia* (even on the X chromosome of *S. latifolia*), inferred sex-linked contigs of *S. pseudotites* were grouped on the LG6 of *S. latifolia* (Figure 5.1a).

The sex determinism of *Silene otites* was less clear. While the XY model had a slightly lower BIC than the ZW model, it also inferred a hundred less sex-linked contigs (Table 5.1). We therefore considered as autosomes the 2,160 contigs that were inferred as autosomes under both XY and ZW models, as XY sex-linked contigs the 155 contigs that were inferred as sex-linked under the XY model and undefined under the ZW model, and as ZW sex-linked contigs the 188 contigs that were inferred as sex-linked under the ZW model and undefined under the XY model. Over the 2,160 autosomal contigs, 2,008 contigs had an ortholog on a *S. latifolia* scaffold, and 326 of them were mapped on the genetic map of *S. latifolia*. In the same fashion, over the 155 XY sex-linked contigs, 143 had an ortholog on a *S. latifolia* scaffold, and 37 of them were mapped. Finally, over the 188 ZW sex-linked contigs, 75 had an ortholog on a *S. latifolia* scaffold, and 44 of them were mapped.

**Table 5.1:** Model comparison using SEX-DETector on *Silene pseudotites* and *Silene otites*

| species | Sex determination system | Chromosomal category | Number of contigs | BIC |
|---|---|---|---|---|
| *Silene pseudotites* | XY | Sex-linked | 174 | 3,422,710.589 |
| | | Autosomal | 7,233 | |
| | | Undefined | 37,351 | |
| | ZW | Sex-linked | 0 | 3,430,155.856 |
| | | Autosomal | 7,608 | |
| | | Undefined | 37,150 | |
| | No sex chromosome | Sex-linked | 0 | 3,438,114.512 |
| | | Autosomal | 8,638 | |
| | | Undefined | 36,120 | |
| *Silene otites* | XY | Sex-linked | 222 | 2,844,016.544 |
| | | Autosomal | 4,827 | |
| | | Undefined | 50,005 | |
| | ZW | Sex-linked | 329 | 2,844,920.705 |
| | | Autosomal | 5,232 | |
| | | Undefined | 49,493 | |
| | No sex chromosome | Sex-linked | 0 | 2,862,178.838 |
| | | Autosomal | 7,008 | |
| | | Undefined | 48,046 | |

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



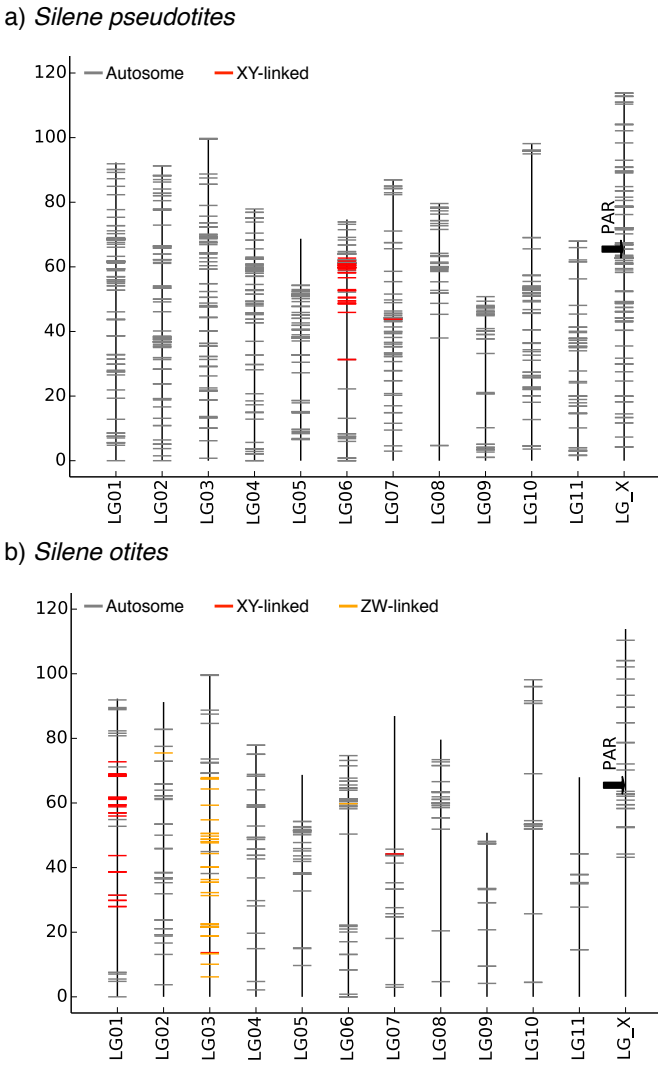



**Figure 5.1:** Mapping of contigs inferred as autosomal (in grey), XY-linked (in red) and ZW-linked (in orange) of (a) *Silene pseudotites* and (b) *Silene otites*, on the genetic map of *Silene latifolia* (Papadopulos et al., 2015). Chromosomal category was inferred using a probabilistic method based on segregation types, implemented in SEX-DETector.

While inferred autosomal contigs of *S. otites* were spread over all linkage groups of *S. latifolia* (even on the X chromosome of *S. latifolia*), inferred XY sex-linked contigs were grouped on LG1 of *S. latifolia* and inferred XY sex-linked contigs were grouped on LG3 (Figure 5.1b).

### 5.3.2 X-Y and Z-W divergence

In *S. pseudotites*, both the number of synonymous and nonsynonymous substitutions were slightly but significantly higher in the Y haplotype than in the X haplotype (G test, $G = 4.491$, $p = 0.034$ and $G = 9.692$, $p = 0.002$, respectively, Table 5.2 ). Therefore, $d_S$ and $d_N$ were slightly higher in the Y haplotype (Table 5.2). On the contrary, in *S. otites*, we only found a small significant excess of nonsynonymous substitutions in the W haplotype (G test, $G = 3.952$, $p = 0.046$),

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



**Table 5.2:** X-Y and Z-W divergence. Synonymous $d_S$ and nonsynonymous $d_N$ substitution per site using *S. nutans* as outgroup and the synonymous divergence between the X (or Z) and Y haplotype ($dS_{XY}$)

| | *S. pseudotites* | | *S. otites* | | | |
|---|---|---|---|---|---|---|
| | X-linked | Y-linked | X-linked | Y-linked | Z-linked | W-linked |
| Total number of contigs | 125 | | 109 | | 127 | |
| Number of synonymous substitution | 1821 | 1951 | 1444 | 1510 | 1798 | 1860 |
| $d_S$ | 0.060 [0.0548; 0.0667] | 0.0647 [0.0586; 0.0715] | 0.0589 [0.0506; 0.0698] | 0.0616 [0.0533; 0.0726] | 0.0525 [0.0447; 0.0621] | 0.0543 [0.0462; 0.0640] |
| Number of nonsynonymous substitution | 782 | 910 | 1062 | 1102 | 980 | 1070 |
| $d_N$ | 0.0087 [0.0075; 0.0100] | 0.0101 [0.0088; 0.0115] | 0.0146 [0.0106; 0.0197] | 0.0151 [0.0112; 0.0203] | 0.0096 [0.0077; 0.0122] | 0.0105 [0.0086; 0.0131] |
| $d_N/d_S$ | 0.1439 [0.1242; 0.1660] | 0.1563 [0.1355; 0.1798] | 0.2478 [0.1922; 0.3042] | 0.2459 [0.1935; 0.3000] | 0.1839 [0.1558; 0.2124] | 0.1941 [0.1664; 0.2219] |
| $dS_{XY}$ | 0.0192 [0.0166; 0.0222] | | 0.0122 [0.010; 0.0147] | | 0.0111 [0.0093; 0.0131] | |

otherwise differences were not significant (Table 5.2).

The divergence between the X and Y haplotypes of *S. pseudotites* was higher than both the divergences between the X and Y haplotypes and between the Z and W haplotypes of *S. otites* (Table 5.2).

### 5.3.3 Sex-linked *vs.* autosomal polymorphism and selection efficacy

Synonymous and nonsynonymous polymorphism were similar between autosomes and homogametic genotypes in both species (Table 5.3). Although $\pi_N/\pi_S$ decreases with expression levels in both species ($F_{(1,253)} = 27.831$, $p < 0.001$ for *S. pseudotites* and $F_{(1,126)} = 5.575$, $p = 0.020$ for *S. otites*), no difference was observed between segregation type ($F_{(1,253)} = 0.156$, $p = 0.693$ for *S. pseudotites* and $F_{(2,126)} = 0.772$, $p = 0.464$ for *S. otites*) and their interaction ($F_{(1,253)} = 0.497$, $p = 0.482$ for *S. pseudotites* and $F_{(2,126)} = 0.599$, $p = 0.551$ for *S. otites*).

### 5.3.4 Degeneration and dosage compensation of the sex-limited chromosome

The expression levels of sex-linked contigs in both species were highly correlated between male and females (Figure S5.1 in the appendix section). In *S. pseudotites*, 151 out of 174 XY-linked contigs were sex-unbiased (after removing five contigs with missing data), while in *S. otites*, 135 out of 155 XY-linked contigs were sex-unbiased (after removing eleven contigs with missing data) and 160 out of 188 ZW-linked contigs were sex-unbiased (after removing ten contigs with missing data). The fact that most sex-linked genes were sex-unbiased could suggest that the Y

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



**Table 5.3:** Characteristics of the data sets on which polymorphism was calculated for *S. pseudotites* and *S. otites*. Statistics on autosomes and XX-linked were calculated on females and on ZZ-linked were calculated on males.

| Species | *S. pseudotites* | | *S. otites* | | |
|---|---|---|---|---|---|
| Segregation type | Autosomes | XX-linked | Autosomes | XX-linked | ZZ-linked |
| No. of individuals | 2 | 2 | 4 | 4 | 4 |
| No. cleaned contigs | 3072 | 85 | 972 | 56 | 93 |
| No. of SNPs | 24939 | 614 | 11403 | 571 | 971 |
| $\pi_S$ | 0.0135 [0.0128; 0.0141] | 0.0126 [0.0105; 0.0149] | 0.0117 [0.0112; 0.0124] | 0.0102 [0.0061; 0.0175] | 0.0091 [0.0076; 0.0108] |
| $\pi_N$ | 0.0025 [0.0023; 0.0027] | 0.0020 [0.0016; 0.0025] | 0.0020 [0.0018; 0.0021] | 0.0017 [0.0013; 0.0021] | 0.0019 [0.0015; 0.0024] |
| $\pi_N/\pi_S$ | 0.1883 [0.1771; 0.2002] | 0.1607 [0.1244; 0.2030] | 0.1696 [0.1579; 0.1824] | 0.1643 [0.1100; 0.2625] | 0.2111 [0.1689; 0.2584] |

or W chromosomes have a limited degeneration or that dosage compensation is effective on the X or Z chromosomes in the heterogametic sex.

Focusing on sex-unbiased expressed genes that were heterozygous in the heterogametic sex, we first compared the levels of expression of the sex chromosomes in the heterogametic sex with the homogametic sex: we expected the Y or W chromosome to be less expressed than X or Z in the homogametic sex in case of degeneration, and a higher expression of X or Z in the heterogametic sex than in the homogametic sex in case of dosage compensation. In *S. pseudotites*, both Xm:XXf and Ym:XXf had a distribution centered close to −1 (median = −0.9176 and −1.0936, respectively, Figure5.2a), but Xm:XXf was significantly higher than Ym:XXf (Wilcoxon signed-rank paired test, $V = 6\,500$, $p < 10^{-5}$). Consequently, Ym:Xm has a distribution centered slightly below 0 (median = −0.1906, Figure 5.2a). Similarly, in *S. otites*, the Y and the W alleles had a slightly reduced expression than the X and the Z alleles of the heterogametic sex when compared to the genotypic expression of the homogametic sex (median of Xm:XXf = −1.0509, and median of Ym:XXf = −1.2259, Wilcoxon signed-rank paired test, $V = 5\,406$, $p < 10^{-5}$; median of Zf:ZZm = −0.8317, and median of Wf:ZZm = −0.9842, Wilcoxon signed-rank paired test, $V = 8\,228$, $p < 10^{-4}$, Figure 5.2b).

Second, as suggested by Mank (2013), we measured the expression of sex-linked genes in reference to the autosomal and ancestral expression of non-dioecious *S. nutans*, a sister species that belongs to the same sub-genus. Once again, in case of sex-specific chromosome degeneration we expected its expression to be reduced, and the other chromosome to be overexpressed in the heterogametic sex in case of dosage compensation. Overall, expression of *S. pseudotites* was slightly higher than *S.nutans* (Figure 5.2a). Xm:AAnut was higher than Ym:AAnut (median = −0.8229 and −1.0244 respectively, Wilcoxon signed-rank paired test, $V = 3\,482$, $p < 0.001$, Fig-


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



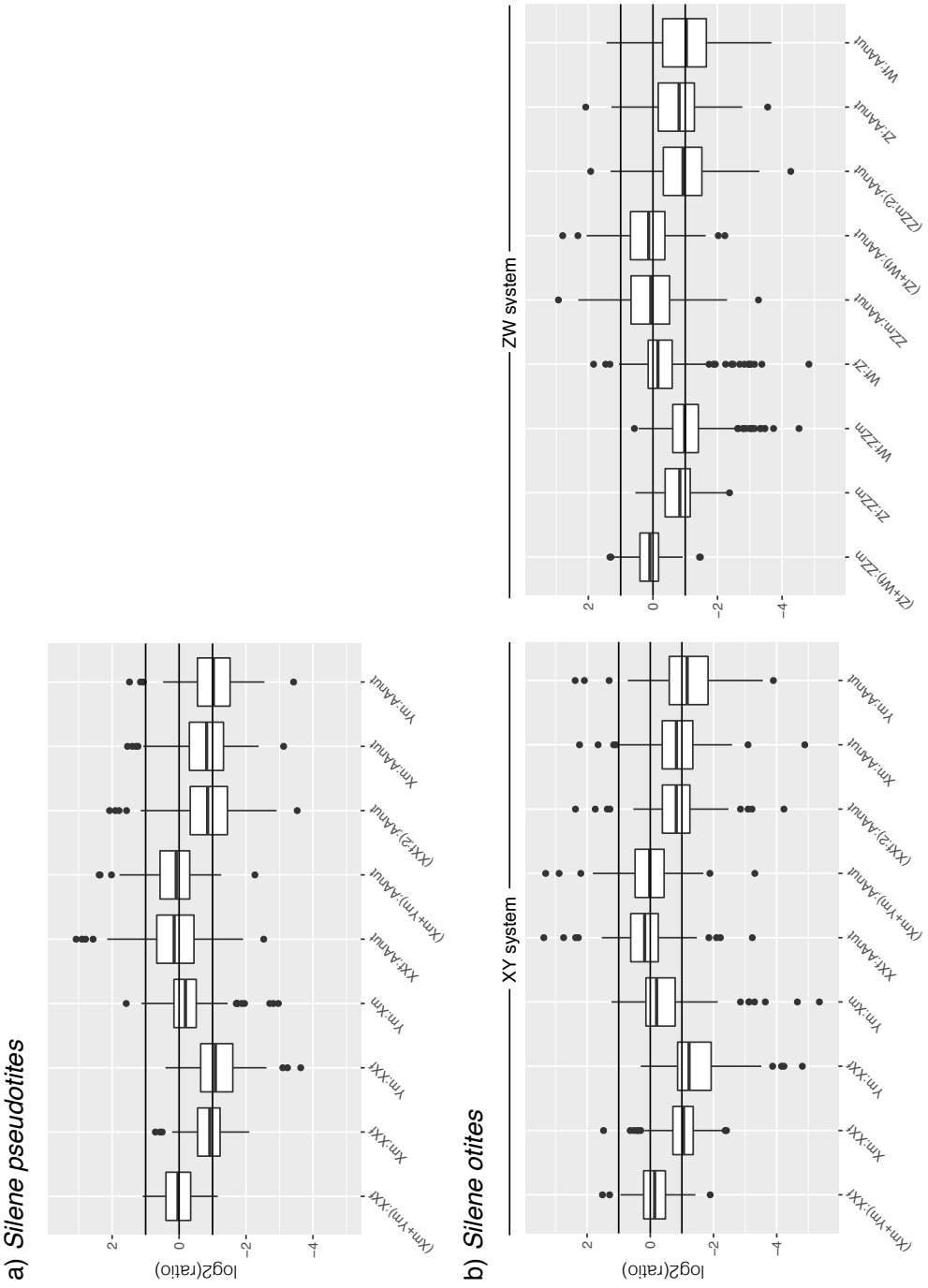



**Figure 5.2:** Gene expression ratio in *S. pseudotites* and *S. otites*. Xm = male X allele expression, Ym = male Y allele expression, XXf = female genotype expression, Zf = female Z allele expression, Wf = female W allele expression, ZZm = male genotype expression and AAnut = genotype expression of orthologs in the non dioecious *S. nutans* that do not have sex chromosomes (therefore, orthologs are on autosomes). Horizontal lines across the plot show the values of double, half and no change in expression (1, −1 and 0, respectively). In *S. pseudotites*, n = 133 for the first four boxplots and n = 98 otherwise. In *S. otites*, n = 121 for the first four boxplots, n = 90 for the fifth next, n = 154 for the four next and n = 110 for the fifth last.

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



ure 5.2a), suggesting a mild Y degeneration while Xm:AA was not significantly higher than Xf:AA (median Xf:AA = −0.8524, Wilcoxon signed-rank paired test, $V$ = 5 207, $p$ = 0.572, Figure 5.2a). When Xm:AAnut was plotted against Ym:AAnut, the latter being a proxy of the Y level of degeneration, no clear dosage compensation (i.e., higher Xm:AA relatively to Ym:AA when Ym:AA is low) could be detected (Figure 5.3a). In *S. otites*, the sex-specific chromosome has a reduced expression in the heterogametic sex for both the XY system (median Xm:AAnut = −0.8268, median Ym:AAnum = −1.1668, Wilcoxon signed-rank paired test, $V$ = 2 395, $p$ < 0.001) and the ZW system (median Zf:AAnut = −0.8093, median Wf:AAnum = −1.0327, Wilcoxon signed-rank paired test, $V$ = 4 041, $p$ = 0.003, Figure5.2b). While the X expression levels in the heterogametic sex was not significantly different from the homogametic sex (median (XXf/2):AAnut = −0.8215, $V$ = 1 608, $p$ = 0.807, Figure 5.2b), Xm:AAnut tended to be higher than Ym:AAnut when Ym:AAnut was low (Figure 5.3b). The Z expression levels in the heterogametic sex were significantly higher than in the homogametic sex (median (ZZm/2):AAnut = −0.9326, $V$ = 1 848, $p$ < 0.001, Figure 5.2b), with a higher level of Zf:AAnut relatively to Wf:AAnut when Wf:AAnut was low (i.e., when the W chromosome was degenerated, Figure 5.3c). This clearly shows dosage compensation in the ZW system.

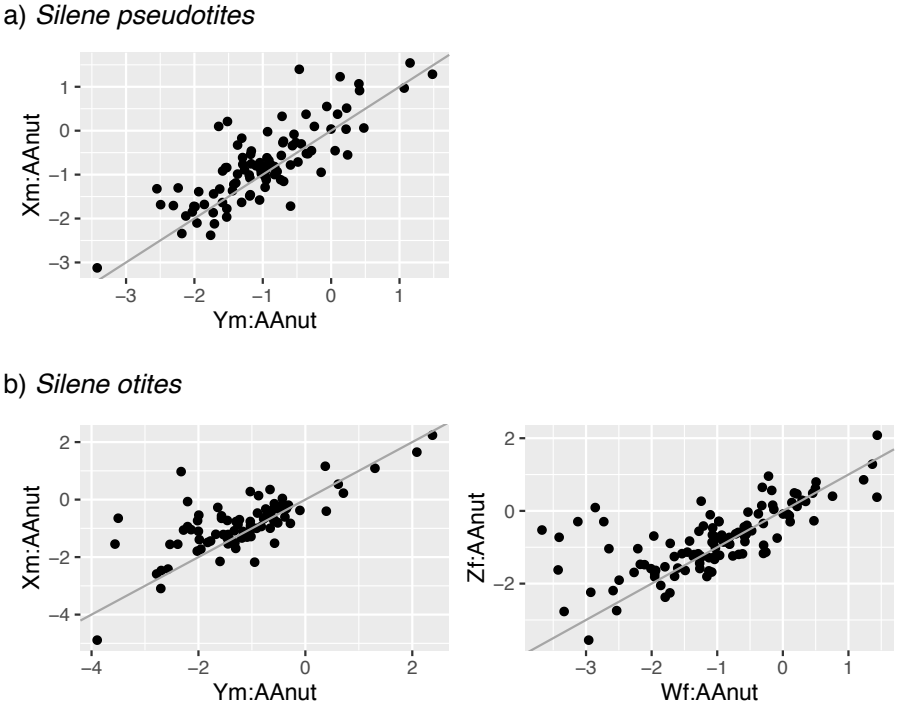



**Figure 5.3:** Gene expression ratio in *S. pseudotites* and *S. otites*. Xm = male X allele expression, Ym = male Y allele expression, Zf = female Z allele expression, Wf = female W allele expression and AAnut = genotype expression of orthologs in the non dioecious *S. nutans* that do not have sex chromosomes (therefore, orthologs are on autosomes). Diagonal lines across the plot show no change in expression.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



## 5.4 Discussion

### 5.4.1 Sex determination of *S. pseudotites* and *S. otites*

One of our main findings is that *S. pseudotites* is male heterogametic, as suggested by Sansome (1938). Its sex-linked genes are not orthologous to those of *S. latifolia*, but to the LG6-linked genes of *S. latifolia* (Figure 5.1a). These results support the idea that the sex chromosomes of *S. latifolia* and *S. pseudotites* have evolved from different pairs of autosomes, in agreement with the hypothesis of an independent origin of sex chromosomes (Desfeux et al., 1996, Mrackova et al., 2008, Marais et al., 2011, Slancarova et al., 2013). Mrackova et al. (2008) found that *S. colpophylla*, a close relative to *S. pseudotites* in the *Otites* subsection, is also male heterogametic and does not share sex-linked genes with *S. latifolia* either. Whether *S. colpophylla* and *S. pseudotites* share their sex-linked genes, and therefore share the same X-Y chromosomes, remains to be determined.

Sansome (1938) first identified the heterogametic sex in *S. otites* based on the study of the sex ratio in the progeny of controlled crosses. He first crossed a *S. pseudotites* female with a *S. otites* male and obtained mostly females. He also crossed a *S. otites* female with a *S. pseudotites* male and obtained an unbiased sex-ratio (Table 5.4). He proposed male heterogamety in *S. pseudotites* and female heterogamety in *S. otites* to explain that he obtained only one sex in the former cross, while he obtained an unbiased sex ratio in the latter cross (Table 5.4). However, our results on *S. otites* suggest that both male and female heterogamety occur in our controlled cross. XY-linked genes are orthologous to *S. latifolia* LG1-linked genes, whereas ZW-linked genes are orthologous to *S. latifolia* LG3-linked genes (Figure 5.1b). Thus, three sex chromosomes could segregate in a hybrid progeny. To explain the observed phenotype segregation (Table 5.4), Sansome (1938) advances that the sex chromosomes of *S. pseudotites* take over sex determinism. Nevertheless, we also could propose the presence of a cryptic CMS in *S. pseudotites* that male-sterelized the $A_1X_1A_3Z_3X_6A_6$ progeny.

On the other hand, Warmke (1942) crossed a tetraploid *S. otites* female with a diploid *S. otites* male, and the offspring consisted of 50% of females and 50% of males. To explain the presence of both sexes, he proposed male heterogamety: 4n-female XXXX crossed with 2n-male XY form both XXX and XXY genotypes in the progeny (leading to a female and a male phenotype, respectively), since a 4n-female ZWZW crossed with 2n-male ZZ would form only one genotype, ZZW, and thus only one phenotype (Warmke, 1942, Westergaard, 1958). These observed proge-

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



**Table 5.4:** Sansome (1938)'s experiment in the light of our results. P = *S. pseudotites*, O = *S. otites*, 1:1 = unbiased phenotype, A = autosome, X = X chromosome, Y = Y chromosome, Z = Z chromosome, W = W chromosome, subscript numbers refer to *S. latifolia* linkage group as defined by Papadopulos et al. (2015).

| Sansome's observation | | Sansome's deduction | | | | Hypothesis from our results | | | |
|---|---|---|---|---|---|---|---|---|---|
| Dame sp x Sire sp | Progeny phenotype | Mother genotype | Father genotype | Progeny genotype | Progeny phenotype | Mother genotype | Father genotype | Progeny genotype | Progeny phenotype |
| P x O | Female | XX | ZZ | XZ | Female | $A_1A_1A_3A_3X_6X_6$ | $X_1Y_1Z_3Z_3A_6A_6$ | $A_1X_1A_3Z_3X_6A_6$ | Female |
| | | | | | | | | $A_1Y_1A_3Z_3X_6A_6$ | Female |
| O x P | 1:1 | ZW | XY | ZX | Female | $X_1X_1Z_3W_3A_6A_6$ | $A_1A_1A_3A_3X_6Y_6$ | $X_1A_1Z_3A_3A_6X_6$ | Female |
| | | | | WX | Female | | | $X_1A_1W_3A_3A_6X_6$ | Female |
| | | | | ZY | Male | | | $X_1A_1Z_3A_3A_6Y_6$ | Male |
| | | | | WY | Male | | | $X_1A_1W_3A_3A_6Y_6$ | Male |

lilliad.univ-lille.fr

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



nies are difficult to explain in the light of our results on *S. otites*. Nevertheless, as Slancarova et al. (2013) suggest, one cannot reject the hypothesis that this experiment was in fact performed on *S. pseudotites* instead of *S. otites sensu stricto*. Indeed, many forms have been grouped under the name *S. otites*, such as *S. pseudotites*, but also *S. densiflora* or *S. exaltata* (Tutin et al., 1964, Yildiz and Çirpici, 2013).

Recently, Slancarova et al. (2013) found genetic evidence for female heterogamety in *S. otites* using amplified fragment length polymorphism (AFLP). Over 171 segregating AFLP markers, four were Z-linked, one was W-linked, and two were pseudoautosomal (Slancarova et al., 2013). Our RNA-seq data on one family confirms the presence of a linkage group that confers ZW determinism.

Finally, our results suggest that *S. otites* follows a polygenic sex determination (PSD). As pointed out by Moore and Roberts (2013), this phenomenon might be more frequent than usually believed, when involving autosomal or sex chromosomes. For example, in the African ciclid fish (*Metriaclima pyrsonotus*), both XY and ZW chromosomes derived from different autosomes (7 and 5, respectively) can segregate in a family (Ser et al., 2010). However, contrary to the *S. otites* case, the ZW system in *M. pyrsonotus* is epistatically dominant to the XY system, leading to XYZZ males and ZZXX, ZWXX and ZWXY females. In the *S. otites* family of the present study, the possibility to detect both systems using the SEX-DETector methodology implies that all females are $X_1X_1Z_3W_3$, and that all males are $X_1Y_1Z_3Z_3$ (numbers referring to *S. latifolia*'s LGs). Therefore, two possible genotypes are missing and possibly lethal: $X_1X_1Z_3Z_3$ and $X_1Y_1Z_3W_3$. Further understanding of the molecular mechanisms that prevent the formation or the survival of these genotypes is needed.

### 5.4.2 Molecular evolution of young sex chromosomes

In *S. pseudotites*, the X-Y divergence ($d_{XY}$ = 0.0192) is higher than both the average X-linked ($\pi_S$ = 0.0126) and autosomal ($\pi_S$ = 0.0135) polymorphisms, suggesting that recombination stopped between gametologs. As a result of selective events that occured in a non-recombining region, the effective population size of the Y-linked genes is expected to be reduced, weakening the efficacy of selection (Charlesworth and Charlesworth, 2000). Although we found a higher number of synonymous substitutions in the Y haplotype compared to the X (as previously described in birds, e.g., Kahn and Quinn, 1999; in plants, e.g., Filatov and Charlesworth, 2002; and in mammals, e.g., Goetting-Minesky and Makova, 2006, as males undergo a higher germline cell division), $d_N/d_S$ tended to be higher for Y-linked genes consistent with a reduced efficacy of purifying selection,

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



ultimately leading to the degeneration of the Y chromosomes (Charlesworth and Charlesworth, 2000). Nevertheless, the observed pattern is not as strong as the one observed in *S. latifolia* (Papadopulos et al., 2015), suggesting a much younger system in *S. pseudotites*.

*S. otites* might be an even younger system, as it is more parsimonious to postulate that *S. colpohylla* and *S. pseudotites*, both exhibiting a XY system that does not share sex-linked genes with *S. latifolia*, have the same sex chromosomes. Therefore, the transition between a unique XY determinism to a multiple sex chromosome system could have occured as recently as the differentiation between *S. otites* and *S. pseudotites*. The fact that the X-Y divergence in *S. pseudotites* is higher than both the X-Y and Z-W divergences in *S. otites* supports this hypothesis. Interestingly, in *S. otites*, we did not find a $dS_{XY}$ higher than $dS_{ZW}$, suggesting that the expected male mutation bias pattern of Y is not detectable. This is in agreement with a recent recruitment of autosomes into XY sex chromosomes. Nevertheless, we found a higher number of nonsynonymous substitution in W-linked genes compared to Z-linked genes, while it was similar between Y-linked and X-linked genes. This suggests that the effect of a reduced efficacy of selection on the W is already visible. We therefore hypothesize that the ZW system was the first to be recruited, given the observed patterns of molecular evolution clearer on W than on Y.

### 5.4.3 Levels of autosomal, X, and Y polymorphism

In both species, the genetic diversity is similar between the autosomal and the X-linked genes, while in *S. otites* the Z-linked genes exhibited a lower synonymous polymorphism than both autosomal and X-linked genes (Table 5.3). Due to the reduced effective size of sex chromosomes compared to autosomes, their genetic diversity should be lower. However, different evolutionary factors, such as the mating system, can impact the neutral expectation (reviewed in Ellegren, 2009). In particular, our results could suggest that polygyny occurs in both species. Indeed, if very few males reproduce, the expected X:A ratio approaches 1, and is higher than the Z:A ratio (Ellegren, 2009). While we could not find any information about the sex ratio in natural populations of *S. pseudotites*, the sex ratio of *S. otites* is generally male-biased (Correns, 1928 in Sansome, 1938) but can be changed to female-biased in small populations (Lauterbach et al., 2012) or under certain environmental conditions (high vegetation coverage or dry coverage, Soldaat et al., 1997, 2000). *S. otites* is predominantly insect-pollinated (Brantjes and Leemans, 1976, Dötterl et al., 2012) and emits floral odors to attract or repel mosquitoes, one of its pollinators Jhumur et al. (2008). Variations among males to attract pollinators could lead to variations in male mating success, and thus to polygyny (if pollinators then visit several females). Historically controversial,

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



sexual selection in dioecious plants is now interpreted as a major selective force in the evolution of flora diversity (Moore and Pannell, 2011).

### 5.4.4 Dosage compensation in young systems

When focusing on sex-unbiased sex-linked genes in both species, we found Y and W alleles to be slightly less expressed than X or Z alleles in the heterogametic sex when compared to their autosomal orthologs in *S. nutans*, used as a proxy of ancestral expression (Figure 5.2). The difference becomes stronger as the level of degeneration of the expression of the sex-specific chromosomes in *S. otites* increases (Figure 5.3). Finally, Z allele expressions are slightly higher in the heterogametic sex than in the homogametic one in *S. otites* (Figure 5.2). These results suggest that dosage compensation occurs, even mildly, in both species, despite the youth of their sex chromosomes (Mank, 2013). Evidence for dosage compensation in young system is rare in plant as well as in animals. However, the 10 million years old *S. latifolia* sex chromosomes enabled *de novo* dosage compensation to be selected for as soon as Y expression started to decline (Muyle et al., 2012, Papadopulos et al., 2015). In *Rumex hastatulus*, sex chromosomes evolved within the past 15-16 million years and a neo-Y sex chromosome system recently derived (Navajas-Pérez et al., 2005). Both old and young sex-linked genes showed an overall trend of reduced Y expression relative to X-linked alleles, while male and female expression were not different (Hough et al., 2014). In animals, the neo-X chromosome of *Drosophila miranda*, which is at most two million years old, is partially compensated (Marín et al., 1996). Our results, together with previous studies, suggest that dosage compensation occurs rapidly after the recruitment of an autosome as a sex chromosome.

Nevertheless, sex chromosome dosage compensation does not seem to be universal, and when it occurs in a given species, it might only apply to a sub-set of sex-linked genes (Mank, 2013). Indeed, female heterogametic species seem to lack global dosage compensation when compared with male heterogametic species (e.g. in chicken, Ellegren et al., 2007; or in snakes, Vicoso et al., 2013). This inequality can be explained by sexual conflicts, together with stronger selection and greater reproductive variance in males, that delay the selection for dosage compensation in Z compared with X (Mullon et al., 2015). In *S. otites*, we found trends of dosage compensation in both XY and ZW systems, although the XY system seems more recent. The patterns of polymorphism described above suggest that polygyny occurs in this species, which could lead to a greater reproductive variance in males and therefore promote the conditions that are predicted to lead to a faster selection of X over Z for dosage compensation.

© 2016 Tous droits réservés.



lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



### 5.4.5 Turnover of sex chromosomes in sub-section otites

We found evidence of 3 different sex chromosome pairs between *S. pseudotites* and *S. otites*. With *S. latifolia* and *S. diclinis*, this rises the number of sex chromosomes to at least 5 in the *Silene* genus (Howell et al., 2009). The diversity due to sex chromosome turnover is well known in fishes (Ross et al., 2009), amphibians (Dufresnes et al., 2015), or diptera (Vicoso and Bachtrog, 2015). In plants, this phenomenon has been described in *Salicaceae* (Hou et al., 2015). Several mechanisms have been suggested to explain the lability of genetic sex determination, such as sex ratio bias (Ogata et al., 2003, Vuilleumier et al., 2007), X-Y recombination (Dufresnes et al., 2015), or sex-antagonistic selection (van Doorn and Kirkpatrick, 2007, 2010). Interestingly, on theoretical grounds, turnover is facilitated when the sex-limited chromosome is not too much degenerated (explaining that turnover is observed in species with homomorphic sex-chromosomes), which is the case in *S. pseudotites* XY system, if we consider it to be ancestral to the XY/ZW system found in *S. otites*. However, given this rapid turnover that seems to occur in this sub-section, composed of more than ten species that are all dioecious (Oxelman et al., 2013), investigation of sex-determination systems at the scale of the sub-section is necessary to have a better picture of sex chromosome evolution, using the methodology used in the present study as a very efficient first step. In addition, the striking case of *S. otites* exhibiting a polygenic sex determination (PGD) with both XY and ZW chromosomes unrelated to the XY system found in *S. pseudotites*, opens the question of the fixation of this complex system at the species level on a larger geographical scale, especially considering the fitness cost involved in the maintenance of both systems leading to the lethality of half the progeny. If PGD is confirmed in *S. otites*, it will also offer a unique opportunity to investigate the concomitant evolution of two heterogametic systems within a single species and disentangle the evolutionary process acting on sex chromosomes, all things being equal (Ellegren, 2009).

## Acknowledgements

We are grateful to Eric Schmitt for taking care of the *Silene* collection, Sylvain Santoni for his expertise on NGS strategies, Chantal Griveau (Conservatoire Botanique National du Bassin Parisien) and Jos Kafer for sharing seeds, Jonathan Aceituno for technical support with R, and Bohuslav Janousek for discussions. Numerical results presented in this paper were carried out using the HPC service of the Centre de Ressources Informatiques (CRI) of Lille 1 University. We thank the technical staff for providing the technical support and infrastructure. This work



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



was supported by the Agence Nationale de la Recherche (ANR-14-CE19-0021 (NGSex), ANR-11-BSV7-013-03 (TRANS)) to PT and GAB and a PhD fellowship from the French Research Ministry to HM.

## 5.5 Bibliography

Bachtrog, D. (2006). A dynamic view of sex chromosome evolution. *Current Opinion in Genetics and Development*, 16:578–585.

Bachtrog, D., Kirkpatrick, M., Mank, J. E., McDaniel, S. F., Pires, J. C., Rice, W. R., and Valenzuela, N. (2011). Are all sex chromosomes created equal? *Trends in Genetics*, 27:350–357.

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T.-L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J. C., and the tree of sex consortium (2014). Sex determination: why so many ways of doing it? *PLoS biology*, 12:e1001899.

Bergero, R. and Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends in Ecology and Evolution*, 24:94–102.

Bernasconi, G., Antonovics, J., Biere, A., Charlesworth, D., Delph, L. F., Filatov, D., Giraud, T., Hood, M. E., Marais, G. A. B., McCauley, D., Pannell, J. R., Shykoff, J. A., Vyskot, B., Wolfe, L. M., and Widmer, A. (2009). *Silene* as a model system in ecology and evolution. *Heredity*, 103:5–14.

Brantjes, N. B. M. and Leemans, J. A. A. M. (1976). *Silene otites* (Caryophyllaceae) pollinated by nocturnal lepidoptera and mosquitoes. *Acta Botanica Neerlandica*, 25:281–295.

Burgarella, C., Gayral, P., Ballenghien, M., Bernard, A., David, P., Jarne, P., Correa, A., Hurtrez-Boussès, S., Escobar, J., Galtier, N., and Glémin, S. (2015). Molecular evolution of freshwater snails with contrasting mating systems. *Molecular Biology and Evolution*, 32:2403–2416.

Charlesworth, B. and Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 355:1563–1572.

Charlesworth, D. and Mank, J. E. (2010). The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. *Genetics*, 186:9–31.

Correns, C. (1928). Bestmmung, Vererbung und Verteilung des Geschlechts bei der höheren Pflanzen. *Handb. Vererbungw*, 2.

Desfeux, C., Maurice, S., Henry, J.-P., Lejeune, B., and Gouyon, P.-H. (1996). Evolution of reproductive systems in the genus *Silene*. *Proceedings of the Royal Society B: Biological Sciences*, 263:409–414.

Dötterl, S., Jahreiß, K., Jhumur, U. S., and Jürgens, A. (2012). Temporal variation of flower scent in *Silene otites* (Caryophyllaceae): a species with a mixed pollination system. *Botanical Journal of the Linnean Society*, 169:447–460.

Dufresnes, C., Borzée, A., Horn, A., Stöck, M., Ostini, M., Sermier, R., Wassef, J., Litvinchuck, S. N., Kosch, T. A., Waldman, B., Jang, Y., Brelsford, A., and Perrin, N. (2015). Sex-chromosome homomorphy in palearctic tree frogs results from both turnovers and X-Y recombination. *Molecular Biology and Evolution*, 32:2328–2337.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Ellegren, H. (2009). The different levels of genetic diversity in sex chromosomes and autosomes. *Trends in Genetics*, 25:278–284.

Ellegren, H. (2011). Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nature Reviews Genetics*, 12:157–166.

Ellegren, H., Hultin-Rosenberg, L., Brunström, B., Dencker, L., Kultima, K., and Scholz, B. (2007). Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. *BMC Biology*, 5:40–52.

Filatov, D. A. and Charlesworth, D. (2002). Substitution rates in the X- and Y-linked genes of the plants, *Silene latifolia* and *S. dioica*. *Molecular Biology and Evolution*, 19:898–907.

Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., Lourenco, J. M., Alves, P. C., Ballenghien, M., Faivre, N., Belkhir, K., Cahais, V., Loire, E., Bernard, A., and Galtier, N. (2013). Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genetics*, 9:e1003457.

Geraldes, A., Hefer, C. A., Capron, A., Kolosova, N., Martinez-Nuñez, F., Soolanayakanahally, R. Y., Stanton, B., Guy, R. D., Mansfield, S. D., Douglas, C. J., and Cronk, Q. C. B. (2015). Recent y chromosome divergence despite ancient origin of dioecy in poplars (*Populus*). *Molecular Ecology*, 24:3243–3256.

Goetting-Minesky, M. P. and Makova, K. D. (2006). Mammalian male mutation bias: impacts of generation time and regional variation in substitution rates. *Journal of Molecular Evolution*, 63:537–544.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., and Regev, A. (2013). De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocols*, 8:1494–1512.

Hou, J., Ye, N., Zhang, D., Chen, Y., Fang, L., Dai, X., and Yin, T. (2015). Different autosomes evolved into sex chromosomes in the sister genera of *Salix* and *Populus*. *Scientific Reports*, 5:9076.

Hough, J., Hollister, J. D., Wang, W., Barrett, S. C. H., and Wright, S. I. (2014). Genetic degeneration of old and young Y chromosomes in the flowering plant *Rumex hastatulus*. *Proceedings of the National Academy of Sciences*, 111:7713–7718.

Howell, E. C., Armstrong, S. J., and Filatov, D. A. (2009). Evolution of neo-sex chromosomes in *Silene diclinis*. *Genetics*, 182:1109–1115.

Huang, X. and Madan, A. (1999). CAP 3: a DNA sequence assembly program. *Genome Research*, 9:868–877.

Jhumur, U. S., Dötterl, S., and Jürgens, A. (2008). Floral odors of *Silene otites*: their variability and attractiveness to mosquitoes. *Journal of Chemical Ecology*, 34:14–25.

Kahn, N. W. and Quinn, T. W. (1999). Male-driven evolution among Eoaves? A test of the replicative division hypothesis in a heterogametic female (ZW) system. *Journal of Molecular Evolution*, 49:750–759.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9:357–359.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Lauterbach, D., Ristow, M., and Gemeinholzer, B. (2012). Population genetics and fitness in fragmented populations of the dioecious and endangered *Silene otites* (Caryophyllaceae). *Plant Systematics and Evolution*, 298:155–164.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, . G. P. D. P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25:2078–2079.

Mank, J. E. (2013). Sex chromosome dosage compensation: definitely not for everyone. *Trends in Genetics*, 29:677–683.

Marais, G. A. B., Forrest, A., Kamau, E., Käfer, J., Daubin, V., and Charlesworth, D. (2011). Multiple nuclear gene phylogenetic analysis of the evolution of dioecy and sex chromosomes in the genus *Silene*. *PloS One*, 6:e21915.

Marín, I., Franke, A., Bashaw, G. J., and Baker, B. S. (1996). The dosage compensation system of *Drosophila* is co-opted by newly evolved X chromosomes. *Nature*, 383:160–163.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17:10–12.

Ming, R., Bendahmane, A., and Renner, S. S. (2011). Sex chromosomes in land plants. *Annual Review of Plant Biology*, 62:485–514.

Moore, E. C. and Roberts, R. B. (2013). Polygenic sex determination. *Current Biology*, 23:510–512.

Moore, J. C. and Pannell, J. R. (2011). Sexual selection in plants. *Current Biology*, 21:176–182.

Mrackova, M., Nicolas, M., Hobza, R., Negrutiu, I., Monéger, F., Widmer, A., Vyskot, B., and Janousek, B. (2008). Independent origin of sex chromosomes in two species of the genus *Silene*. *Genetics*, 179:1129–1133.

Mullon, C., Wright, A. E., Reuter, M., Pomiankowski, A., and Mank, J. E. (2015). Evolution of dosage compensation under sexual selection differs between X and Z chromosomes. *Nature Communications*, 6:7720.

Muyle, A., Käfer, J., Zemp, N., Mousset, S., Picard, F., and Marais, G. A. B. (2016). SEX-DETector : a probabilistic approach to study sex chromosomes in non-model organisms. *Genome Biology and Evolution*, 8:2530–2543.

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

Navajas-Pérez, R., de la Herrán, R., González, G. L., Jamilena, M., Lozano, R., Rejón, C. R., Rejón, M. R., and Garrido-Ramos, M. A. (2005). The evolution of reproductive systems and sex-determining mechanisms within *Rumex* (polygonaceae) inferred from nuclear and chloroplastidial sequence data. *Molecular Biology and Evolution*, 22:1929–1939.

Ogata, M., Ohtani, H., Igarashi, T., Hasegawa, Y., Ichikawa, Y., and Miura, I. (2003). Change of the heterogametic sex from male to female in the frog. *Genetics*, 164:613–620.

Oxelman, B., Rautenberg, A., Thollesson, M., Larsson, A., Frajman, B., Eggens, F., Petri, A., Aydin, Z., Töpel, M., and Brandtberg-Falkman, A. (2013). *Sileneae* taxonomy and systematics. *http://www.sileneae.info*.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Papadopulos, A. S. T., Chester, M., Ridout, K., and Filatov, D. A. (2015). Rapid Y degeneration and dosage compensation in plant sex chromosomes. *Proceedings of the National Academy of Sciences*, 112:13021–13026.

Ranwez, V., Harispe, S., Delsuc, F., and Douzery, E. J. P. (2011). MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One*, 6:e22594.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26:139–140.

Ross, J. A., Urton, J. R., Boland, J., Shapiro, M. D., and Peichel, C. L. (2009). Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *PLoS Genetics*, 5:e1000391.

Sansome, F. W. (1938). Sex determination in *Silene otites* and related sepecies. *Journal of Genetics*, 35:387–396.

Schmieder, R. and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27:863–864.

Ser, J. R., Roberts, R. B., and Kocher, T. D. (2010). Multiple interacting loci control sex determination in Lake Malawi cichlid fish. *Evolution*, 64:486–501.

Slancarova, V., Zdanska, J., Janousek, B., Talianova, M., Zschach, C., Zluvova, J., Siroky, J., Kovacova, V., Blavet, H., Danihelka, J., Oxelman, B., Widmer, A., and Vyskot, B. (2013). Evolution of sex determination systems with heterogametic males and females in *Silene*. *Evolution*, 67:3669–3677.

Soldaat, L. L., Lorenz, H., and Treffiich, A. (2000). The effect of drought stress on the sex ration variation of *Silene otites*. *Folia Geobotanica*, 35:203–210.

Soldaat, L. L., Vetter, B., and Klotz, S. (1997). Sex ratio in populations of *Silene otites* in relation to vegetation cover, population size and fungal infection. *Journal of Vegetation Science*, 8:697–702.

Tsagkogeorga, G., Cahais, V., and Galtier, N. (2012). The population genomics of a fast evolver: high levels of diversity, functional constraints, and molecular adaptation in the Tunicate *Ciona intestinalis*. *Genome Biology and Evolution*, 4:852–861.

Tutin, T., Heywood, V., Burges, N., Valentine, D., Walters, S., and Webb, D., editors (1964). *Flora europaea*. Cambridge University Press, Cambridge.

van Doorn, G. S. and Kirkpatrick, M. (2007). Turnover of sex chromosomes induced by sexual conflict. *Nature*, 449:909–912.

van Doorn, G. S. and Kirkpatrick, M. (2010). Transitions between male and female heterogamety caused by sex-antagonistic selection. *Genetics*, 186:629–645.

Vicoso, B. and Bachtrog, D. (2015). Numerous transitions of sex chromosomes in *Diptera*. *PLoS Biology*, 13:e1002078.

Vicoso, B., Emerson, J. J., Zektser, Y., Mahajan, S., and Bachtrog, D. (2013). Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biology*, 11:e1001643.

Vuilleumier, S., Lande, R., Van Alphen, J. J. M., and Seehausen, O. (2007). Invasion and fixation of sex-reversal genes. *Journal of Evolutionary Biology*, 20:913–920.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Warmke, H. (1942). A new method for determining the sex heterozygote in species with morphologically undifferentiated sex chromosomes, and its application to *Silene otites*. *Genetics*, 27:174.

Westergaard, M. (1958). The mechanism of sex determination in dioecious flowering plants. *Advances in Genetics*, 9:217–281.

Yildiz, K. and Çirpici, A. H. (2013). Taxonomic revision of *Silene* (Caryophyllaceae) sections *Siphonomorpha*, *Lasiostemones*, *Sclerocalycinae*, *Chloranthae*, *Tataricae*, and *Otites* in Turkey. *Turkish Journal of Botany*, 37:191–218.

## 5.6 Appendix

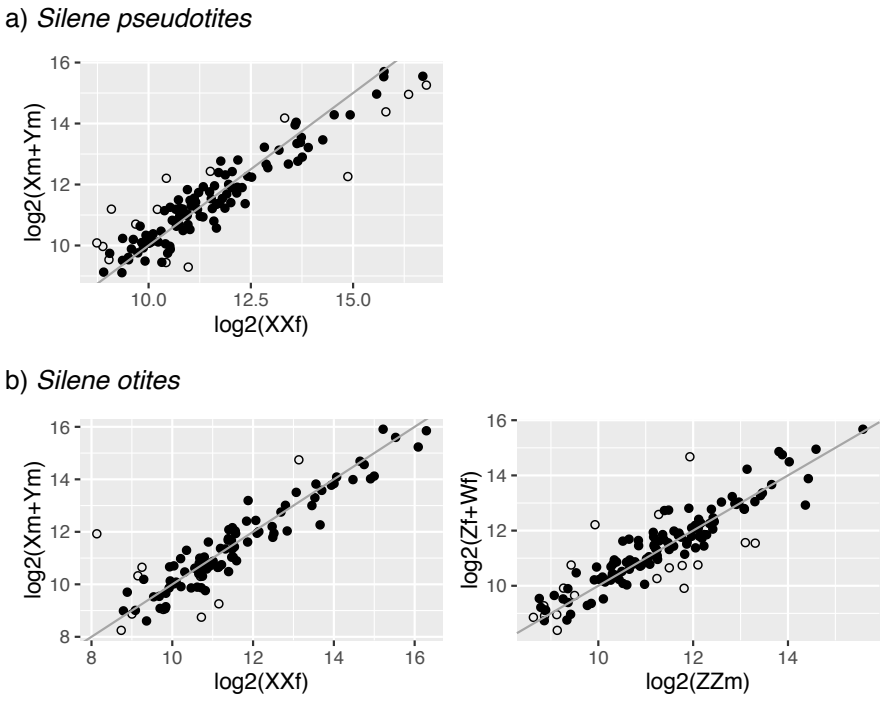



**Figure S5.1:** Sex-linked contigs expression in male and female in *S. pseudotites* and *S. otites*. Mean expression values in males are plotted against those in females. Grey circle represent sex-biased contigs and black circle represent sex-unbiased contigs. Sex-biased expression was identified using edgeR.

© 2016 Tous droits réservés.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

# **Chapitre 5** : Effet du système de reproduction sur l'efficacité de la sélection

**Aline Muyle*, Hélène Martin*, Sylvain Glémin, Pascal Touzet** et Gabriel Marais****

*, ** les auteurs ont contribué de manière égale

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



L'objectif de ce chapitre était de tester l'impact du système de reproduction sur l'efficacité de la sélection et notamment de tester l'hypothèse du cul-de-sac évolutif associée à la dioécie. Nous avons utilisé des données RNA-seq de huit espèces dans le genre *Silene* : 2 hermaphrodites, 3 gynodioïques et 3 dioïques, provenant de la section *Behenantha* et de la section *Silene* et nous avons estimé leurs niveaux de polymorphisme et de divergence. Les résultats présentés ici sont des résultats préliminaires, mais ils montrent une corrélation entre l'efficacité de la sélection, le système de reproduction et la diversité génétique. En particulier, nous avons constaté que les espèces dioïques étaient moins efficaces pour fixer des mutations bénéfiques que les espèces gynodioïques pour des niveaux de diversité équivalents.

Une partie des données RNA-seq utilisées dans ce chapitre a été générée avant cette thèse et proviennent de différents laboratoires (LBBE de Lyon et le laboratoire d'Alex Widmer de Zurich). Ce chapitre a été fait en étroite collaboration avec Aline Muyle et Gabriel Marais du Laboratoire de Biométrie et Biologie Evolutive (Lyon). J'ai réalisé les analyses dans la section *Silene* pendant qu'Aline s'occupait de la section *Behenantha*. Nous avons co-écrit la section materiel et méthodes et une partie des résultats. Et j'ai rédigé le reste du chapitre en incorporant les corrections et remarques des co-auteurs.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



# Breeding system and effective population size influence the efficacy of selection in the *Silene* genus

## 6.1 Introduction

The evolutionary consequences of the transition from hermaphroditism to dioecy remain unclear. As species richness is generally lower in dioecious taxa compared to their hermaphroditic or monoecious sister taxa, dioecy is traditionally associated with lower diversification (and/or higher extinction rate, Heilbuth, 2000) and considered an evolutionary dead-end. However, recent studies have questioned the dark fate of dioecious lineages and suggested only a weak (Sabath et al., 2016) or, on the contrary, a positive effect (Käfer and Mousset, 2014, Käfer et al., 2014) of dioecy on diversification rate. Two main ecological causes could alter population dynamics by reducing effective population size, ultimately leading to the extinction of dioecious populations. First, the "seed-shadow handicap" hypothesis predicts that, as only 50% of the population produces seeds, seed dispersal efficiency is reduced, leading to higher local competition (Heilbuth et al., 2001). Second, sexual selection can lead to sexual dimorphism and to an unequal attractiveness to pollinators between sexes. If pollinators are rare, only one sex can be visited and no or few seeds are produced (Vamosi and Otto, 2002).

Reductions in seed dispersal and production are expected to cause a reduction in effective population size ($N_e$) and thus in genetic diversity (Ellegren and Galtier, 2016). As a result, purifying selection is weakened and the rate of deleterious substitutions is expected to increase (Charlesworth, 2009), as observed in small-$N_e$ animal species (Galtier, 2016). On the contrary, in large-$N_e$ species, when evolution is not mutation-rate limited, one would expect a higher rate of adaptive substitution (Lanfear et al., 2014, Ellegren and Galtier, 2016). Although this correlation has been observed in *Helianthus* (Strasburg et al., 2011), it is not always detected (Galtier, 2016), as other predictors could impact the rate of adaptive substitution (Lourenço et al., 2013).

However, few empirical studies have studied the effects of the transition to dioecy at the genomic level (Castric et al., 2014, Muyle and Marais, 2016). Käfer et al. (2013) found evidence for relaxed purifying selection in dioecious *S. latifolia* compared to its gynodioecious close relative *S. vulgaris*, in agreement with the above predictions. However, they did not detect differences between the dioecious *S. otites* and its gynodioecious close relative *S. nutans*. They explained

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



these disparate results by a difference in time scale: the dioecious clade of *S. otites*, being younger than the one of *S. latifolia*, did not have enough time to show a reduced $N_e$ at the genomic level. Nevertheless, without polymorphism data, they were not able to test for a reduction of diversity of dioecious species, as expected in a small-$N_e$ species.

There is a high diversity of sexual systems in *Silene*: hermaphroditism is the most common, followed by gynodioecy (comprising gynodioecy and gynodioecy-gynomonoecy) and dioecy (Casimiro-Soriguer et al., 2015). Dioecious species are present in distinct subgenera of *Silene*. Thus, dioecy probably evolved independently several times from hermaphroditism, possibly via gynodioecy (Desfeux et al., 1996, Marais et al., 2011, Slancarova et al., 2013, Dufay et al., 2014). In this study, we focused on the genomic effects of the transition to dioecy in two subgenera, *Behenantha* and *Silene*. In each subgenus, we sequenced the transcriptome of a hermaphroditic, a gynodioecious and a dioecious species from several natural populations. As the studied species have distinct breeding systems and distribution ranges, thus possibly distinct effective sizes, we tested the above expectations using polymorphism and divergence data.

## 6.2 Material and Methods

### 6.2.1 Sampling and sequencing

We investigated patterns of polymorphism and divergence of three mating systems (hermaphroditism, gynodioecy and dioecy) in two subgenera of the *Silene* genus: subgenus *Behenantha* and subgenus *Silene* (Figure 6.1). For the subgenus *Behenantha*, we sampled fourteen females of the dioecious species *S. latifolia*, ten hermaphrodites and four females of the gynodioecious species *S. vulgaris*, and five hermaphrodites of the hermaphroditic species *S. viscosa*. For the subgenus *Silene*, we sampled four females and four males of the dioecious species *S. otites*, two females and two males of its dioecious sister species *S. pseudotites*, seven hermaphrodites and four females of the gynodioecious species *S. nutans* E (Eastern lineage, Martin et al., 2016), eight hermaphrodites, two gynomonoecious and one females of the gynodioecious species *S. nutans* W (Western lineage, Martin et al., 2016) and twelve hermaphrodites of the hermaphroditic species *S. paradoxa*. Two samples of *Dianthus chinensis* were added to root the phylogeny.

Seeds were sampled around the geographic distribution of each species (Supplementary Figures S6.1). They were sown and grown in controlled greenhouse conditions. In case plants did not flower, young leaves were sampled instead. Total RNA was extracted through the Spectrum

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



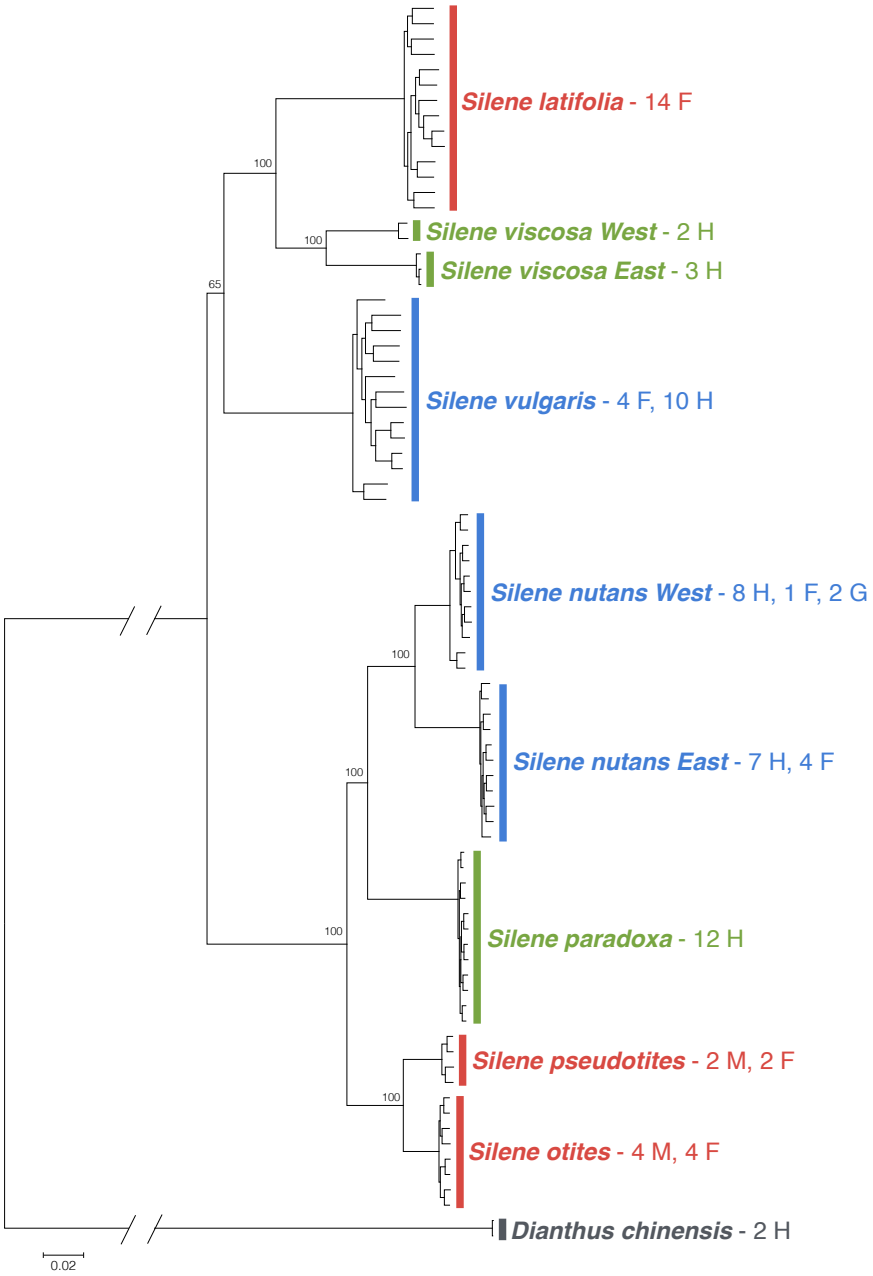



**Figure 6.1:** Sampled species and their phylogenetic relationships. Dioecious species are shown in red, gynodioecious species in blue and hermaphroditic species in green. Sample size and and sex are indicated: F for female, H for hermaphrodite, G for gynomonoecious and M for male. Bootstrap support values were generated using 1 000 replications and only between species values are indicated.

Plant Total RNA kit (Sigma, Inc., USA), following the manufacturer's protocol, and treated with a DNAse. Libraries were prepared with the TruSeq RNA sample Preparation v2 kit (Illumina Inc., USA). Each 2 nM cDNA library was sequenced using a paired-end protocol on a HiSeq2000 sequencer, producing 100 bp reads. Demultiplexing was performed using CASAVA 1.8.1 (Illumina) to produce paired sequence files containing reads for each sample in the Illumina FASTQ format. RNA extraction and sequencing were done by the sequencing platform in the AGAP laboratory (Montpellier, France).

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



### 6.2.2 Reference transcriptome assemblies

Reference transcriptomes were assembled *de novo* for all species. Reads from all individuals of subgenus *Behenantha* were pooled for each species. 100% identical reads were assumed to be PCR duplicates, and thus filtered using the ConDeTri trimming software (Smeds and Künstner, 2011). Illumina reads were then filtered for sequencing adapters and low read quality using the ea-utils FASTQ processing utilities (Aronesty, 2011). Filtered reads were assembled using trinity with the default settings (Haas et al., 2013). Poly-A tails were removed using PRINSEQ (Schmieder and Edwards, 2011), with parameters `-trim_tail_left 5 -trim_tail_right 5`. rRNA-like sequences were removed using riboPicker version 0.4.3 (Schmieder et al., 2012), with parameters `-i 90 -c 50 -l 50` and the following databases: the SILVA Large subunit reference database, the SILVA Small subunit reference database, the GreenGenes database, and the Rfam database. Obtained transcripts were further assembled, inside of trinity components, using cap3 (Huang and Madan, 1999), with parameter `-p 90`, and home-made Perl scripts.

We filtered raw data of the five taxa of subgenus *Silene* using Cutadapt (Martin, 2011) to remove sequencing adapters, as well as PRINSEQ (Schmieder and Edwards, 2011) to remove low-quality reads and poly-A tails. We assembled *de novo* reference transcriptomes by mixing all members of a family For each focal species of the subgenus, reference transcriptomes were assembled *de novo* by mixing filtered reads from two individuals from the same population using Trinity (Haas et al., 2013) and a subsequent CAP3 run (Huang and Madan, 1999) with the default settings. Finally, in both subgenera, we used TransDecoder (Haas et al., 2013) with PFAM domain searches as ORF retention criteria to predict coding sequences.

### 6.2.3 Mapping, genotyping, and alignment

We mapped reads from each member of a family onto their reference transcriptome using BWA version 0.7.12 with parameter `-n 5` (Li and Durbin, 2009) for subgenus *Behenantha*, and using Bowtie2 (Langmead and Salzberg, 2012) for subgenus *Silene*. SAM files were compressed and sorted using SAMTOOLS (Li et al., 2009). We genotyped individuals using reads2snp_2.0 (Tsagkogeorga et al., 2012, Gayral et al., 2013), with parameters `-par 1 -min 3` (`10` for subgenus *Silene*) `-aeb -bqt 20 -rqt 10`; in other words: stringent filtering of paralogous positions, minimum coverage of 3 (or 10) reads for calling a genotype, accounting for allelic expression bias, minimum read mapping quality threshold of 10, and minimum base quality threshold of 20. The output we obtained was composed of two unphased sequences for each

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



individual and each ORFs.

Orthologous ORFs were identified using OrthoMCL Basic Protocol 2 (Li et al., 2003), with default parameters. Best reciprocal blast hits were considered orthologs among species, and orthologous ORFS were aligned for all individuals using MACSE (Ranwez et al., 2011). Species from subgenus *Behenantha* were aligned separately with *S. paradoxa*, which is considered as an outgroup of this section, and species from subgenus *Silene* were aligned with *S. viscosa*, considered as an outgroup of that section.

### 6.2.4 Phylogenetic reconstruction

Orthologous ORFs among species were obtained using best reciprocal blast hits among the nine species (all species from the *Silene* genus and *Dianthus chinensis*). For each orthologous ORF, the nine species were added one by one into the alignment using MACSE (Ranwez et al., 2011). The data were filtered so that a maximum of 20% of missing data per site and per species was allowed, totaling 47,075 SNPs spread over 1,345 ORFs. We used the maximum of ORFs we had, so we did not restrict our analysis to autosomal ORFs identified in dioecious species. The phylogenetic tree was reconstructed using the GTRGAMMA model in RaxML (Stamatakis, 2006). Bootstrap support values were generated using 1,000 replications.

### 6.2.5 Data filtering

We only analyzed autosomal genes, in order to avoid the confounding effects of sex chromosomes and mating systems on selection efficacy. Autosomal genes were predicted using SEXDETector (Muyle et al., 2016), thanks to the study of a cross (parents and progeny) sequenced using RNA-seq (data not shown). 11,951 contigs were inferred as autosomal in *S. latifolia*, 4,162 in *S. otites*, and 7,342 in *S. pseudotites*.

We further filtered the data in order to remove ORFs with frameshifts. Only biallelic sites (sites with two or less alleles) were kept for analyses. Also, we only retained ORFs longer than 100 codons and with a maximum of 20% of missing individuals of the focal species in the alignment, after removing incomplete codons and gaps. Finally, we only analyzed cleaned ORFs present in all species of a section, so that statistics were estimated on the same gene, across species of a section. Therefore, 188 homologous ORF were analysed for subgenus *Silene*, and 1,947 for subgenus *Behenantha*.

© 2016 Tous droits réservés.

 lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



### 6.2.6 Polymorphism, divergence levels, and statistics on the efficacy of selection

We computed and averaged population statistics on the filtered dataset for each ORF using the PopPhyl tool dNdSpiNpiS_1.0 (Gayral et al., 2013), with parameter kappa=2. The following statistics were computed for each contig: mean fixation index $F_{IT}$ (Weir and Cockerham, 1984), per-individual heterozygosity (proportion of heterozygote positions), per-site synonymous ($\pi_S$) and nonsynonymous ($\pi_N$) diversity, and number of synonymous ($d_S$) and nonsynonymous ($d_N$) fixed differences between the focal species and the outgroup. For each species, we averaged statistics across ORFs, weighting by the number of complete sites per gene, thus giving equal weight to every SNP. For each *Silene* species, $\pi_N$, $\pi_S$, $d_N$, and $d_S$ were averaged first, and $\pi_N/\pi_S$ and $d_N/d_S$ were subsequently computed. Confidence intervals were obtained with 10 000 bootstraps over ORFs. The PopPhyl tool also outputs neutrality index NI = $(\pi_N/\pi_S)/(d_N/d_S)$ (Rand and Kann, 1996), which was computed over all SNPs (NI) and on SNPs filtered for having a minor allele frequency higher than 0.2 (NI_0.2, Fay et al., 2001). A neutrality index higher than 1 indicates negative selection, whereas a neutrality index lower than 1 indicates positive selection.

### 6.2.7 Effects of expression levels on the efficacy of selection

As purifying selection is found to be more efficient in highly expressed contigs (Drummond et al., 2005, Nabholz et al., 2014, Burgarella et al., 2015), we controlled for the effect of expression on $\pi_N/\pi_S$. Expression levels were computed in RPKM (reads per kilobase per million mapped reads) by summing read coverage for each contig on all positions and dividing by the length of the contigs and the library size (number of mapped reads) of the individual. Expression levels were then averaged among individuals for each contig. ORFs were grouped into classes of expression, each class containing a similar number of SNPs, and mean $\pi_N/\pi_S$ were computed for each class. To test for differences in $\pi_N/\pi_S$ among breeding systems, controlling for expression, an ANCOVA was performed using R (R Core Team Development, 2014) after the log-transform of $\pi_N/\pi_S$ and expression.

### 6.2.8 Inference of adaptation

The unfolded synonymous and nonsynonymous site frequency spectra (SFS, the distribution of derived allele counts across SNPs) were computed for each species with the sfs program provided

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



by Nicolas Galtier. These SFS were used to estimate $\alpha$, the proportion of amino-acid substitutions that occurred adaptively, according to a modified version of the method of Eyre-Walker and Keightley (Eyre-Walker and Keightley, 2009, Galtier, 2016). A population genetic model was fitted to SFS with parameters including population mutation rate $\theta$, demography r, neutral divergence T, and DFE of nonsynonymous mutation using the program dfem provided by Nicolas Galtier. Four distinct DFE were used: the classical Neutral and Gamma models, which only consider negative selection coefficients, and the GammaExpo and ScaledBeta models, which account for beneficial mutations (see Galtier, 2016). The goodness of fit to the data of each model was compared using the Akaike Information Criterion (AIC), and the best model was chosen to infer $\alpha$. The (per synonymous substitution) rate of adaptive nonsynonymous substitutions, $\omega_A = \alpha d_N/d_S$, was also computed, along with the rate of nonadaptive nonsynonymous substitutions $\omega_{NA} = (1 - \alpha) d_N/d_S$.

## 6.3 Results

### 6.3.1 Dataset

Eight *Silene* species were sampled and sequenced by RNA-seq (Figure 6.1). Transcriptomes were *de novo* assembled in each species separately and ORFs were annotated (the total number of ORFs can be found in Table 6.1). To infer the phylogenetic relationship among *Silene* species, orthologous ORFs were aligned among *Silene* species and *Dianthus chinensis*, which was used to root the tree. All samples from one species but *S. viscosa* ended grouped together and formed relevant species (Figure 6.1). In the case of *S. viscosa*, samples were distributed between two highly divergent lineages: a western lineage that groups samples from Sweden and Czech Republic, and an eastern lineage that groups samples from Bulgaria, Russia, and Kirgizstan (Figure 6.1). These two lineages probably belong to two sub-species and were therefore analysed separately in the rest of this chapter.

### 6.3.2 Patterns of polymorphism

The level of synonymous diversity, $\pi_S$, was used as a proxy of the effective size $N_e$. In both sections, the lowest $\pi_S$ was found in hermaphroditic species (0.0073 for *S. paradoxa* and 0.0029 for *S. viscosa E*) while the highest level was found in gynodioecious species (0.0130 for *S. nutans W* and 0.0286 for *S. vulgaris*, Table 6.1). The deviation from the Hardy-Weinberg expectation

© 2016 Tous droits réservés.

 lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



**Table 6.1:** Characteristics of the data sets, coding sequence polymorphism, divergence patterns, rate of adaptive Evolution ($\alpha$), adaptive (and non-adaptive) proportion of $^{d_N/d_S}$ ($\omega_A$, $\omega_{NA}$) for all studied species

| Outgroup | *S. viscosa* | | | | | *S. paradoxa* | | | |
|---|---|---|---|---|---|---|---|---|---|
| Focal species | *S. paradoxa* | *S. otites* | *S. pseudotites* | *S. nutans* E | *S. nutans* W | *S. vulgaris* | *S. viscosa* E | *S. viscosa* W | *S. latifolia* |
| Breeding system | hermaphrodite | dioecious | dioecious | gynodioecious | gynodioecious | gynodioecious | hermaphrodite | hermaphrodite | dioecious |
| # of samples | 12 | 8 | 4 | 11 | 11 | 14 | 3 | 2 | 14 |
| # of clean contigs | 188 | 188 | 188 | 188 | 188 | 1947 | 1947 | 1947 | 1947 |
| # of SNPs | 2223 | 2717 | 2390 | 3262 | 4094 | 97097 | 4184 | 8809 | 81903 |
| $F_{IT}$ | 0.148 [0.110;0.187] | 0.005 [-0.026;0.036] | 0.061 [0.015;0.104] | 0.226 [0.194;0.257] | 0.235 [0.198;0.272] | 0.210 [0.202; 0.217] | 0.786 [0.725; 0.843] | 0.086 [0.033; 0.138] | 0.181 [0.174; 0.188] |
| $\pi_S$ | 0.0073 [0.0061;0.0087] | 0.0114 [0.0101;0.0129] | 0.0125 [0.0111;0.0141] | 0.0118 [0.0099;0.0137] | 0.0130 [0.0113;0.0151] | 0.0286 [0.0278; 0.0295] | 0.0029 [0.0026; 0.0032] | 0.0076 [0.0070; 0.0082] | 0.0234 [0.0225; 0.0242] |
| $\pi_N$ | 0.0012 [0.0010;0.0014] | 0.0019 [0.0016;0.0022] | 0.0021 [0.0018;0.0024] | 0.0014 [0.0012;0.0017] | 0.0018 [0.0016;0.0021] | 0.0035 [0.0034; 0.0036] | 0.0005 [0.0005; 0.0006] | 0.0013 [0.0012; 0.0014] | 0.0028 [0.0027; 0.0030] |
| $\pi_N/\pi_S$ | 0.163 [0.138;0.194] | 0.165 [0.141 ;0.192] | 0.168 [0.145 ;0.194] | 0.121 [0.102 ;0.144] | 0.140 [0.119 ;0.164] | 0.122 [0.117; 0.128] | 0.180 [0.164; 0.197] | 0.172 [0.161; 0.183] | 0.121 [0.116; 0.127] |
| $d_S$ | 0.141 [0.132 ;0.151] | 0.136 [0.128 ;0.145] | 0.146 [0.137 ;0.155] | 0.118 [0.111 ;0.126] | 0.127448 [0.118 ;0.138] | 0.090 [0.088; 0.093] | 0.142 [0.138; 0.146] | 0.121 [0.118; 0.125] | 0.096 [0.093; 0.099] |
| $d_N$ | 0.024 [0.021 ;0.027] | 0.024 [0.021 ;0.027] | 0.026 [0.023 ;0.029] | 0.020 [0.018 ;0.023] | 0.023 [0.020 ;0.026] | 0.019 [0.018; 0.020] | 0.026 [0.025; 0.027] | 0.020 [0.019; 0.021] | 0.020 [0.019; 0.021] |
| $d_N/d_S$ | 0.171 [0.152 ;0.193] | 0.175 [0.155 ;0.197] | 0.180 [0.160 ;0.202] | 0.172 [0.153 ;0.194] | 0.178 [0.158 ;0.200] | 0.210 [0.201; 0.218] | 0.184 [0.177; 0.191] | 0.166 [0.160; 0.173] | 0.205 [0.197; 0.213] |
| $\alpha$ | 0.580 [0.546;0.613] | 0.507 [0.475;0.540] | 0.526 [0.493;0.559] | 0.745 [0.711;0.780] | 0.655 [0.622;0.689] | 0.749 [0.738; 0.762] | 0.477 [0.466; 0.488] | 0.562 [0.549; 0.575] | 0.644 [0.632; 0.656] |
| $\omega_A$ | 0.099 [0.093;0.104] | 0.089542 [0.084;0.095] | 0.094 [0.088;0.100] | 0.126 [0.120;0.132] | 0.115 0.109;0.120] | 0.119 [0.116; 0.120] | 0.075 [0.073; 0.077] | 0.079 [0.077; 0.081&] | 0.103 [0.101; 0.105] |
| $\omega_{NA}$ | 0.072 [0.066;0.077] | 0.087 [0.081;0.093] | 0.085 [0.079;0.091] | 0.043 [0.037;0.049] | 0.060 [0.054;0.066] | 0.040 [0.038; 0.041] | 0.082 [0.081; 0.084] | 0.062 [0.060; 0.063] | 0.057 [0.055; 0.059] |



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



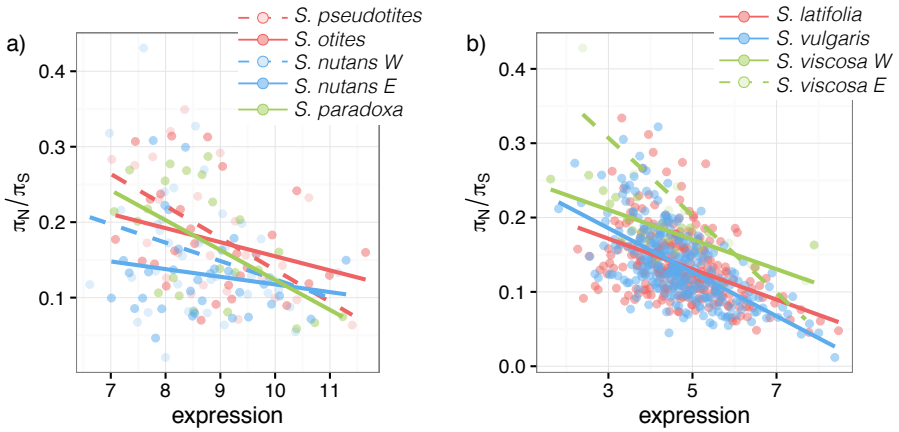



**Figure 6.2:** $\pi_N/\pi_S$ was plotted as a function of log transformed gene expression levels in a) the subgenus *Silene* and b) the subgenus *Behenantha*. Dioecious species are shown in red, gynodioecious species in blue and hermaphroditic species in green.

was the highest in *S. viscosa E*, with a mean fixation index $F_{IT} = 0.7862$ (Table 6.1). In subgenus *Silene*, mean $F_{IT}$ values ranked from 0.0046 (in *S. otites*) to 0.2349 (in *S. nutans W*). Finally, mean per-individual heterozygosity (computed as the proportion of heterozygote positions) ranked from 0.0003 in *S. viscosa E* to 0.0080 in *S. vulgaris* for subgenus *Behenantha*, and from 0.0024 in *S. paradoxa* to 0.0045 in *S. pseudotites* for subgenus *Silene* (Table 6.1).

### 6.3.3 Inference of the efficacy of selection using $\pi_N/\pi_S$ and $d_N/d_S$

The global strength of purifying selection was estimated using the $\pi_N/\pi_S$ ratio. The lowest value of the $\pi_N/\pi_S$ was found in the gynodioecious species *S. nutans E* in the *Silene* subgenus ($\pi_N/\pi_S = 0.121$) and in the dioecious *S. latifolia* for subgenus *Behenantha* ($\pi_N/\pi_S = 0.121$, Table 6.1). In subgenus *Silene*, the covariate (average contig expression) was significantly related to the mean $\pi_N/\pi_S$ ($F_{(1,143)} = 21.628$, $p < 0.001$, partial $\eta^2 = 13.137\%$), but the breeding system, as well as the interaction, had no significant effect on mean $\pi_N/\pi_S$ after controlling for the effect of average expression ($F_{(2,143)} = 2.539$, $p = 0.082$, partial $\eta^2 = 3.429\%$ and $F_{(2,143)} = 1.933$, $p = 0.0149$, partial $\eta^2 = 2.632\%$ respectively, Figure 6.2a). In subgenus *Behenantha*, the covariate (average contig expression) was significantly related to the mean $\pi_N/\pi_S$ ($F_{(1,143)} = 159.433$, $p < 0.001$, partial $\eta^2 = 21.9\%$), as well as the breeding system ($F_{(2,143)} = 4.371$, $p = 0.013$, partial $\eta^2 = 1.5\%$) and the interaction ($F_{(2,143)} = 7.281$, $p < 0.001$, partial $\eta^2 = 2.5\%$, Figure 4.2b). Overall, with both subgenera taken together, we did not find a significant effect of the breeding system ($F_{(2,5)} = 1.852$, $p = 0.250$) or of the average level of synonymous diversity $\pi_S$ ($F_{(1,5)} = 3.838$, $p = 0.107$, Figure 4.4a).

We also estimated selective pressure using the ratio $d_N/d_S$. In subgenus *Silene*, $d_N/d_S$ was similar across species and was around 0.17 on average (Table 6.1). On the contrary, in subgenus

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



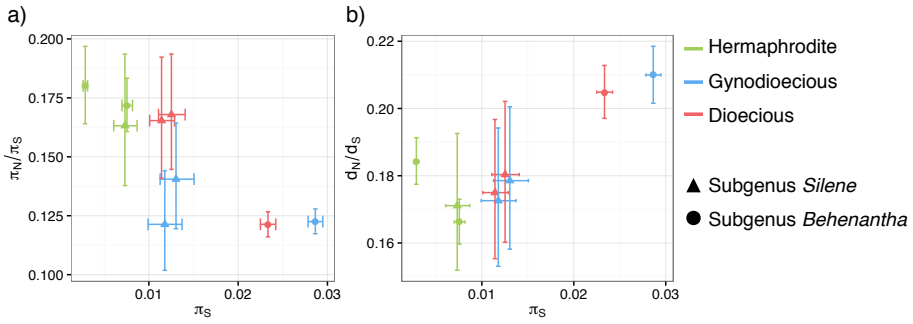



**Figure 6.3:** a) $\pi_N/\pi_S$ and b) $d_N/d_S$ plotted as a function of $\pi_S$.

*Behenantha*, $d_N/d_S$ ranged from 0.168 in hermaphroditic *S. viscosa W* to 0.210 in gynodioecious *S. vulgaris*. Overall, taking both subgenera together, we did not find a significant effect of the breeding system ($F_{(2,5)} = 0.469$, $p = 0.651$), but we found a significant effect of the average level of synonymous diversity $\pi_S$ ($F_{(1,5)} = 11.643$, $p = 0.019$, partial $\eta^2 = 70.0\%$, Figure 6.3b).

### 6.3.4 Inference of the efficacy of selection using a McDonald-Kreitman approach

For each species, a DFE of non-synonymous mutation was fitted to the unfolded synonymous and non-synonymous site frequency spectra and divergence data. The estimated DFE were roughly consistent across species (Figure 6.4). Although most mutations were strongly deleterious ($N_e s < 100$), between 13% and 35% of mutations were slightly to moderately deleterious ($-100 < N_e s < 0$), and a very small proportion of mutations (up to 4% for *S. pseudotites*) were neutral to slightly adaptive ($0 < N_e s < 100$).

Taking into account the DFE of each species (thus controlling the effects of slightly deleterious and adaptive mutations), we estimated the proportion of adaptive nonsynonymous substitutions $\alpha$. During the divergence between the species of the *Silene* subgenus and their outgroup *S. viscosa E*, the highest level of $\alpha$ was in gynodioecious *S. nutans E* ($\alpha = 0.74$) and the lowest level was in dioecious *S. otites* ($\alpha = 0.51$, Table 6.1). During the divergence between the species of the *Behenantha* subgenus and their outgroup *S. paradoxa*, the highest level of $\alpha$ was in gynodioecious *S. vulgaris* ($\alpha = 0.75$) and the lowest level was in hermaphroditic *S. viscosa E* ($\alpha = 0.48$). We found a higher rate of adaptive nonsynonymous substitutions ($\omega_A$) and a lower rate of nonadaptive nonsynonymous substitutions ($\omega_{NA}$) in the gynodioecious species in both sections compared to hermaphrodite and dioecious species (Table 6.1).

Synonymous diversity barely impacted $\alpha$ ($F_{(1,5)} = 4.615$, $p = 0.084$, partial $\eta^2 = 48.0\%$), but the breeding system had a significant effect on $\alpha$ after controlling for the effect of $\pi_S$ ($F_{(2,5)} = 6.5206$,


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



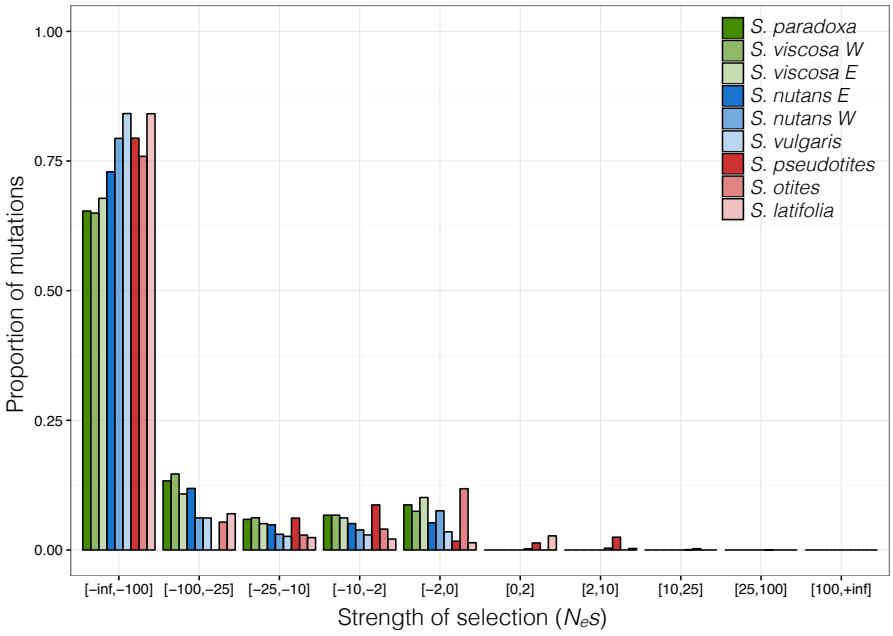



**Figure 6.4:** Distribution of fitness effect of new mutations computed by the program dfem (Galtier, 2016). Dioecious species are shown in red, gynodioecious species in blue and hermaphroditic species in green. Negative values indicate purifying selection and positive values indicate positive selection.

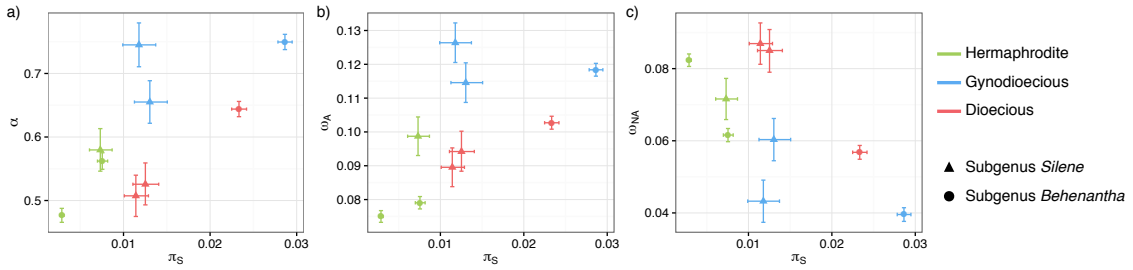



**Figure 6.5:** Values obtained for a) $\alpha$ (proportion of amino-acid substitutions that occurred adaptively), b) $\omega_A$ (rate of adaptive non-synonymous substitution) and c) $\omega_{NA}$ (rate of nonadaptive non-synonymous substitution) in all *Silene* species using program dfem (Galtier, 2016). Dioecious species are shown in red, gynodioecious species in blue and hermaphroditic species in green.

$p = 0.040$, partial $\eta^2 = 72.3\%$, Figure 6.5a). Synonymous diversity did not impact $\omega_A$ ($F_{(1,5)} = 0.357$, $p = 0.576$) while the breeding system had a significant effect on $\omega_A$ after controlling for the effect of $\pi_S$ ($F_{(2,5)} = 6.230$, $p = 0.044$, partial $\eta^2 = 71.4\%$, Figure 6.5b). Finally, synonymous diversity slightly impacted $\omega_{NA}$ ($F_{(1,5)} = 5.882$, $p = 0.060$, partial $\eta^2 = 54.0\%$) as well as the breeding system ($F_{(2,5)} = 5.506$, $p = 0.054$, partial $\eta^2 = 68.8\%$, Figure 6.5c).

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



## 6.4 Discussion

### 6.4.1 Patterns of polymorphism and adaptation in the *Silene* genus

Synonymous diversity ($\pi_S$) varied widely across our data set (from 0.0029 to 0.0286). Interestingly, $\pi_S$ seemed to be roughly positively correlated with the geographic distribution area (Figure S6.1). The highest levels of $\pi_S$ were found in *S. latifolia* and *S. vulgaris*, two widespread species known for their successful post-glacial expansion and invasiveness (Taylor and Keller, 2007). These high observed levels of genetic diversity could be the result of sampling distinct lineages that arose from distinct glacial refugia, or from a single structured refugium (Taylor and Keller, 2007). On the contrary, the geographical distribution of other taxa was smaller, and these taxa more likely came from a single refugium. *S. nutans* used to be described as having a distribution as large as *S. latifolia* (Figure S6.1), but our phylogenetic tree suggests the occurence of at least two taxa that probably represent two reproductively isolated species (Van Rossum et al., 1996, Martin et al., 2016). The geographical distributions of *S. nutans E* and *S. nutans W* were therefore reduced compared to what was previously described. Sub-structures within each of the studied species are in agreement with the low levels of $F_{IT}$, except for *S. viscosa E*, which exhibited very high levels of $F_{IT}$ and low levels of individual heterozygosity, suggesting a high proportion of inbreeding.

Our results provide evidence for a significant role of adaptive evolution in the plant species we investigated, with estimates of $\alpha$ ranging from 0.502 to 0.768. Such a high level of adaptive evolution is on par with those found in animals (Galtier, 2016), but is higher than those previously found in plants (mean estimate across ten species is -0.18, standard error = 0.47, Gossmann et al., 2010), even when considering plant species with a large effective size and low population structure (Ingvarsson, 2010, Slotte et al., 2010). However, $\alpha$ being a ratio, a high estimate can be due to increased fixation of adaptive mutations or by reduced fixation of slightly deleterious mutations (Lourenço et al., 2013), hence $\omega_A$ is preferred (Gossmann et al., 2012). Our estimates of $\omega_A$ are less different from other plants, on par with the sunflower or the lettuce (Gossmann et al., 2010, Strasburg et al., 2011). But, as argued by Strasburg et al. (2011), the way $\omega_A$ is computed, as well as the outgroup used, greatly change the value of the estimators, making between-study comparisons difficult. In this study, we used a pipeline similar to the one used in Galtier (2016) on an animal dataset. Interestingly, our estimates were similar to what he found in animals, in contradiction with the idea that plants exhibit less adaptive evolution than animals (Gossmann et al., 2010).

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



### 6.4.2 $N_e$ and efficacy of selection

We found that synonymous nucleotide diversity $\pi_S$ had only a marginal effect on the proportion of adaptive nonsynonymous substitutions. In agreement with previous studies (in plants, Hough et al., 2013; or in animals, Galtier, 2016), $\alpha$ tends to be lower with small population sizes, probably because deleterious mutations are fixed at a higher rate than in large populations. Indeed, a lower $N_e$ implies a greater effect of genetic drift, leading to a higher probability of fixation of slightly deleterious mutations (Akashi et al., 2012). This mechanically decreases $\alpha$. But while we found a negative effect of $\pi_S$ on the nonadaptive nonsynonymous substitution rate $\omega_{NA}$, we did not find an effect of $\pi_S$ on $\pi_N/\pi_S$. At least two explanations can be suggested to solve this apparent paradox. First, although our data set encompasses a wide range of synonymous diversity values, we lacked large-$N_e$ species ($\pi_S > 0.015$). It would be worthwhile to enlarge our sampling to other *Silene* plant species to get more statistical power. Second, our sampling strategy was to cover the distribution range of the species. Therefore, we plausibly sampled populations from distinct environments. Due to local adaptation, a mutation could be fixed locally in a population, but stay polymorphic at the species level. Sampling individuals over a species range could inflate $\pi_N$ and thus $\pi_N/\pi_S$, in particular in *S. vulgaris* and *S. latifolia* because of their large sampling area. In such a situation, it is important to discriminate adaptive mutations from deleterious ones.

Contrary to previous studies (Strasburg et al., 2011, Gossmann et al., 2012), we did not find any effect of $\pi_S$ on the adaptive nonsynonymous substitution rate $\omega_A$, but we did find a positive effect of $\pi_S$ on the $d_N/d_S$ ratio, suggesting that large-$N_e$ species have a higher fixation rate of adaptive mutations. Nevertheless, this relationship seems to be mostly influenced by *S. latifolia* and *S. vulgaris* (Figure 6.3), which, aside from having a large $N_e$, are invasive and in expansion. Had we removed them from the analysis, the positive relationship would likely not hold anymore. $\omega_A$ is not always expected to correlate with $\pi_S$. Also, $\omega_A$ can be much less influenced by $N_e$ than by factors like the rate of environmental change or organism complexity (Lourenço et al., 2013). In our case, organism complexity does not apply (even if we were able to quantify it) because our study focused on one plant genus. However, environmental change could positively influence $\omega_A$, as a population far from its optimal fitness has a higher proportion of mutations that increase its fitness, while a well-adapted population close to its optimal fitness would have a lower proportion of such mutations (Lanfear et al., 2014). The same hypothesis could hold for a small-$N_e$ species: as it accumulates more slightly deleterious mutations, a higher proportion of compensatory adaptive mutations would appear (Galtier, 2016).

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



### 6.4.3 Breeding system and efficacy of selection: the dead-end hypothesis

Under the dead-end hypothesis of dioecy, dioecious species are expected to have a reduced effective size due to a reduction in seed dispersal and seed production compared to non-dioecious species. Consequently, deleterious mutations would accumulate faster while adaptive mutations would have a lower fixation rate (Castric et al., 2014, Muyle and Marais, 2016). In our study, contrary to what one would expect under the dead-end hypothesis, dioecious species exhibit similar levels of $\pi_S$ (which we used as a proxy of effective size) to gynodioecious species, and higher levels than hermaphroditic species.

Patterns of polymorphism observed in *S. viscosa E* suggested that self-fertilization occurred in the lineage. The reduction of $\pi_S$ in selfers compare to outcrossers is expected to increase the rate of deleterious mutations (Charlesworth and Wright, 2001, Glémin, 2007). Although evidence for such relationships have been recently found in snails (Burgarella et al., 2015), only weak differences between selfers and outcrossers have been reported in plants (Wright et al., 2002, 2008, Haudry et al., 2008), as expected in theoretical models (Glémin, 2007). If *S. viscosa E* has become predominantly selfing, this event should be as recent as the divergence with *S. viscosa W*, which did not seem to self-fertilize. Probably because of this recent origin, no striking differences were found between *S. viscosa E* and the obligated outcrossers *S. otites* and *S. pseudotites* in nonadaptive nonsynonymous substitution rates (Figure 6.5c). On the other hand, while selfers accumulate faster recessive adaptive mutations, outcrossers perform better when adaptive mutations are dominant or co-dominant and selection is weak (Glémin, 2007). We found that *S. viscosa E* has the lowest rate of adaptive nonsynonymous substitution (Figure 6.5b), and most of positively selected mutations were only weakly selected (Figure 6.4). Altogether, these findings suggest that dominant or codominant adaptive mutations arose in the genus *Silene*.

Käfer et al. (2013) found evidence for the lower efficacy of purifying selection in the dioecious species *S. latifolia* compared to gynodioecious *S. vulgaris*. They interpreted this as a sign of a lower effective size, in agreement with a dead-end hypothesis. Using a McDonald-Kreitman-based approach, we found an effect of the breeding system on $\omega_A$ and marginally on $\omega_{NA}$. Estimates of adaptive evolution tend to be higher in gynodioecious species, probably due to a higher fixation of adaptive nonsynonymous mutations and a lower fixation of deleterious mutations compared to the other breeding systems. Nevertheless, these differences between dioecious and gynodioecious species are not due to a lower effective size in the former, as we found similar levels of $\pi_S$ between both breeding systems.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Adaptive evolution could be impacted by something more specific to breeding systems than effective size, such as genomic conflicts (Wright et al., 2008, Glémin and Galtier, 2012). In our gynodioecious species, the coexistence of hermaphrodite and female individuals in populations is due to cytonuclear interactions. Cytoplasmic male sterility (CMS) on the mitochondrial genome increases female fitness while nuclear restorers bring back male fertility. To maintain a coevolution between the mitochondrial and the nuclear genomes, a higher mutation accumulation in the mitochondrial genome might induce a higher molecular evolution in nuclear-encoding mitochondrial proteins (Sloan, 2015). The coevolutionary arms race, as revealed by the higher level of molecular evolution in the mitochondrial genome of gynodioecious species in Silene (Touzet and Delph, 2009, Lahiani et al., 2013) and by the observed nuclear compensatory mutations (Sloan et al., 2014), should be hindered in dioecious species, as the transition to dioecy is expected to fix a mitochondrial genotype, thereby reducing diversity in the cytoplasmic genome (Lahiani et al., 2013).

Nevertheless, the co-occurrence of males and females in dioecious lineages leads to conflicts within and between sexes (Glémin and Galtier, 2012). Genes involved in sexual reproduction are known to evolve rapidly through adaptive evolution (Swanson and Vacquier, 2002, Käfer et al., 2013). Sexual dimorphism and the diverging interests of males and females involve sexually antagonistic genes assumed to have a sex-biased expression (Mank and Ellegren, 2009). Such genes are preferentially sex-linked and are found on sex chromosomes (Mank and Ellegren, 2009, Zemp et al., 2016). As we only focused on autosomes, we were not able to detect these genes. This could explain why we found reduced adaptive evolution in dioecious plant species compared to gynodioecious ones.

## Acknowledgements

We would like to thank Jos Kafer, Raquel Tavares, Alexander Widmer, Niklaus Zemp, J. J. Wieringa, Deborah Charlesworth, Bohuslav Janousek, Tatiana Giraud, Honor C. Prentice, Salvatore Cozzolino, and Kerstin Kustas for sampling *Silene* seeds. We are grateful to Jonathan Aceituno for technical support in R and to Céline Poux for discussion on phylogenetic inference.

We are grateful to Sylvain Santoni and his staff for the extraction and sequencing of RNAseq. This work was supported by the Agence Nationale de la Recherche (ANR-11-BSV7-013-03, TRANS) to SG, GM, and PT and a PhD fellowship from the French Research Ministry to HM. Numerical results presented in this paper were carried out using the HPC service of the

© 2016 Tous droits réservés.

 lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Centre de Ressources Informatiques (CRI) of Lille 1 University. We thank the technical staff for providing the technical support and infrastructure.

## 6.5 Bibliography

Akashi, H., Osada, N., and Ohta, T. (2012). Weak selection and protein evolution. *Genetics*, 192:15–31.

Aronesty, E. (2011). ea-utils: command-line tools for processing biological sequencing data. *http://code.google.com/p/ea-utils*.

Burgarella, C., Gayral, P., Ballenghien, M., Bernard, A., David, P., Jarne, P., Correa, A., Hurtrez-Boussès, S., Escobar, J., Galtier, N., and Glémin, S. (2015). Molecular evolution of freshwater snails with contrasting mating systems. *Molecular Biology and Evolution*, 32:2403–2416.

Casimiro-Soriguer, I., Buide, M. L., and Narbona, E. (2015). Diversity of sexual systems within different lineages of the genus *Silene*. *Annals of Botany Plants*, 7:plv037.

Castric, V., Billiard, S., and Vekemans, X. (2014). Trait transitions in explicit ecological and genomic contexts: plant mating systems as case studies. In Landry, C. R. and Aubin-Horth, N., editors, *Ecological genomics; ecology and the evolution of genes and genomes*.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10:195–205.

Charlesworth, D. and Wright, S. I. (2001). Breeding systems and genome evolution. *Current Opinion in Genetics and Development*, 11:685–690.

Desfeux, C., Maurice, S., Henry, J.-P., Lejeune, B., and Gouyon, P.-H. (1996). Evolution of reproductive systems in the genus *Silene*. *Proceedings of the Royal Society B: Biological Sciences*, 263:409–414.

Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O., and Arnold, F. H. (2005). Why highly expressed proteins evolve slowly. *Proceedings of the National Academy of Sciences*, 102:14338–14343.

Dufay, M., Champelovier, P., Käfer, J., Henry, J.-P., Mousset, S., and Marais, G. A. B. (2014). An angiosperm-wide analysis of the gynodioecy-dioecy pathway. *Annals of Botany*, 114:539–548.

Ellegren, H. and Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics*, 17:422–433.

Eyre-Walker, A. and Keightley, P. D. (2009). Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Molecular Biology and Evolution*, 26:2097–2108.

Fay, J. C., Wyckoff, G. J., and Wu, C.-I. (2001). Positive and negative selection on the human genome. *Genetics*, 158:1227–1234.

Galtier, N. (2016). Adaptive protein evolution in animals and the effective population size hypothesis. *PLoS Genetics*, 12:e1005774.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., Lourenco, J. M., Alves, P. C., Ballenghien, M., Faivre, N., Belkhir, K., Cahais, V., Loire, E., Bernard, A., and Galtier, N. (2013). Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genetics*, 9:e1003457.

Glémin, S. (2007). Mating systems and the efficacy of selection at the molecular level. *Genetics*, 177:905–916.

Glémin, S. and Galtier, N. (2012). Genome evolution in outcrossing versus selfing versus asexual species. *Methods in Molecular Biology*, 855:311–335.

Gossmann, T. I., Keightley, P. D., and Eyre-Walker, A. (2012). The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biology and Evolution*, 4:658–667.

Gossmann, T. I., Song, B. H., Windsor, A. J., Mitchell-Olds, T., Dixon, C. J., Kapralov, M. V., Filatov, D. A., and Eyre-Walker, A. (2010). Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Molecular Biology and Evolution*, 27:1822–1832.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., and Regev, A. (2013). De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocols*, 8:1494–1512.

Haudry, A., Cenci, A., Guilhaumon, C., Paux, E., Poirier, S., Santoni, S., David, J., and Glémin, S. (2008). Mating system and recombination affect molecular evolution in four *Triticeae* species. *Genetics Research*, 90:97–109.

Heilbuth, J. C. (2000). Lower species richness in dioecious clades. *The American Naturalist*, 156:221–241.

Heilbuth, J. C., Ilves, K. L., and Otto, S. P. (2001). The consequences of dioecy for seed dispersal: modeling the seed-shadow handicap. *Evolution*, 55:880–888.

Hough, J., Williamson, R. J., and Wright, S. I. (2013). Patterns of selection in plant genomes. *Annual Review of Ecology, Evolution, and Systematics*, 44:31–49.

Huang, X. and Madan, A. (1999). CAP 3: a DNA sequence assembly program. *Genome Research*, 9:868–877.

Ingvarsson, P. K. (2010). Natural selection on synonymous and nonsynonymous mutations shapes patterns of polymorphism in *Populus tremula*. *Molecular Biology and Evolution*, 27:650–660.

Käfer, J., de Boer, H. J., Mousset, S., Kool, A., Dufay, M., and Marais, G. A. B. (2014). Dioecy is associated with higher diversification rates in flowering plants. *Journal of Evolutionary Biology*, 27:1478–1490.

Käfer, J. and Mousset, S. (2014). Standard sister clade comparison fails when testing derived character states. *Systematic Biology*, 63:601–609.

Käfer, J., Talianová, M., Bigot, T., Michu, E., Guéguen, L., Widmer, A., Žlůvová, J., Glémin, S., and Marais, G. A. B. (2013). Patterns of molecular evolution in dioecious and non-dioecious *Silene*. *Journal of Evolutionary Biology*, 26:335–346.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Lahiani, E., Dufaÿ, M., Castric, V., Le Cadre, S., Charlesworth, D., Van Rossum, F., and Touzet, P. (2013). Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity*, 111:157–164.

Lanfear, R., Kokko, H., and Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends in Ecology and Evolution*, 29:33–41.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9:357–359.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25:1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, . G. P. D. P. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25:2078–2079.

Li, L., Stoeckert Jr, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, 13:2178–2189.

Lourenço, J. M., Glémin, S., and Galtier, N. (2013). The rate of molecular adaptation in a changing environment. *Molecular Biology and Evolution*, 30:1292–1301.

Mank, J. E. and Ellegren, H. (2009). Sex-linkage of sexually antagonistic genes is predicted by female, but not male, effects in birds. *Evolution*, 63:1464–1472.

Marais, G. A. B., Forrest, A., Kamau, E., Käfer, J., Daubin, V., and Charlesworth, D. (2011). Multiple nuclear gene phylogenetic analysis of the evolution of dioecy and sex chromosomes in the genus *Silene*. *PloS One*, 6:e21915.

Martin, H., Touzet, P., Van Rossum, F., Delalande, D., and Arnaud, J.-F. (2016). Phylogeographic pattern of range expansion provides evidence for cryptic species lineages in *Silene nutans* in Western Europe. *Heredity*, 116:286–294.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17:10–12.

Muyle, A., Käfer, J., Zemp, N., Mousset, S., Picard, F., and Marais, G. A. B. (2016). SEX-DETector : a probabilistic approach to study sex chromosomes in non-model organisms. *Genome Biology and Evolution*, 8:2530–2543.

Muyle, A. and Marais, G. (2016). Mating systeme, genome evolution and. In Kliman, R. M., editor, *Encyclopedia of Evolutionary Biology*, volume 2, pages 480–492. Oxford: Academic Press.

Nabholz, B., Sarah, G., Sabot, F., Ruiz, M., Adam, H., Nidelet, S., Ghesquière, A., Santoni, S., David, J., and Glémin, S. (2014). Transcriptome population genomics reveals severe bottleneck and domestication cost in the African rice (*Oryza glaberrima*). *Molecular Ecology*, 23:2210–2227.

R Core Team Development (2014). R: a language and environment for statistical computing. *R Foundation for Statistical Computing Vienna, Austria*, https://www.r–project.org/.

Rand, D. M. and Kann, L. M. (1996). Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Molecular Biology and Evolution*, 13:735–748.

Ranwez, V., Harispe, S., Delsuc, F., and Douzery, E. J. P. (2011). MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. *PLoS One*, 6:e22594.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Sabath, N., Goldberg, E. E., Glick, L., Einhorn, M., Ashman, T.-l., Ming, R., Otto, S. P., Vamosi, J. C., and Mayrose, I. (2016). Dioecy does not consistently accelerate or slow lineage diversification across multiple genera of angiosperms. *New Phytologist*, 209:1290–1300.

Schmieder, R. and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27:863–864.

Schmieder, R., Lim, Y. W., and Edwards, R. (2012). Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics*, 28:433–435.

Slancarova, V., Zdanska, J., Janousek, B., Talianova, M., Zschach, C., Zluvova, J., Siroky, J., Kovacova, V., Blavet, H., Danihelka, J., Oxelman, B., Widmer, A., and Vyskot, B. (2013). Evolution of sex determination systems with heterogametic males and females in *Silene*. *Evolution*, 67:3669–3677.

Sloan, D. B. (2015). Using plants to elucidate the mechanisms of cytonuclear co-evolution. *New Phytologist*, 205:1040–1046.

Sloan, D. B., Triant, D. A., Wu, M., and Taylor, D. R. (2014). Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Molecular Biology and Evolution*, 31:673–682.

Slotte, T., Foxe, J. P., Hazzouri, K. M., and Wright, S. I. (2010). Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. *Molecular Biology and Evolution*, 27:1813–1821.

Smeds, L. and Künstner, A. (2011). ConDeTri - a content dependent read trimmer for illumina data. *PloS One*, 6:e26314.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22:2688–2690.

Strasburg, J. L., Kane, N. C., Raduski, A. R., Bonin, A., Michelmore, R., and Rieseberg, L. H. (2011). Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. *Molecular Biology and Evolution*, 28:1569–1580.

Swanson, W. J. and Vacquier, V. D. (2002). The rapid evolution of reproductive proteins. *Nature Reviews Genetics*, 3:137–144.

Taylor, D. R. and Keller, S. R. (2007). Historical range expansion determines the phylogenetic diversity introduced during contemporary species invasion. *Evolution*, 61:334–345.

Touzet, P. and Delph, L. F. (2009). The effect of breeding system on polymorphism in mitochondrial genes of *Silene*. *Genetics*, 181:631–644.

Tsagkogeorga, G., Cahais, V., and Galtier, N. (2012). The population genomics of a fast evolver: high levels of diversity, functional constraints, and molecular adaptation in the Tunicate *Ciona intestinalis*. *Genome Biology and Evolution*, 4:852–861.

Vamosi, J. C. and Otto, S. P. (2002). When looks can kill: the evolution of sexually dimorphic floral display and the extinction of dioecious plants. *Proceedings of the Royal Society B: Biological Sciences*, 269:1187–1194.

Van Rossum, F., De Bilde, J., and Lefèbvre, C. (1996). Barriers to hybridization in calcicolous and silicolous populations of *Silene nutans* from Belgium. *Belgian Journal of Botany*, 129:13–18.

Weir, B. S. and Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, 38:1358–1370.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Wright, S. I., Lauga, B., and Charlesworth, D. (2002). Rates and patterns of molecular evolution in inbred and outbred *Arabidopsis*. *Molecular Biology and Evolution*, 19:1407–1420.

Wright, S. I., Ness, R. W., Foxe, J. P., and Barrett, S. C. H. (2008). Genomic consequences of outcrossing and selfing in plants. *International Journal of Plant Sciences*, 169:105–118.

Zemp, N., Tavares, R., Muyle, A., Charlesworth, D., Marais, G. A. B., and Widmer, A. (2016). Evolution of sex-biased gene expression in a dioecious plant. *Nature Plants*, accepted.

## 6.6 Appendix

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

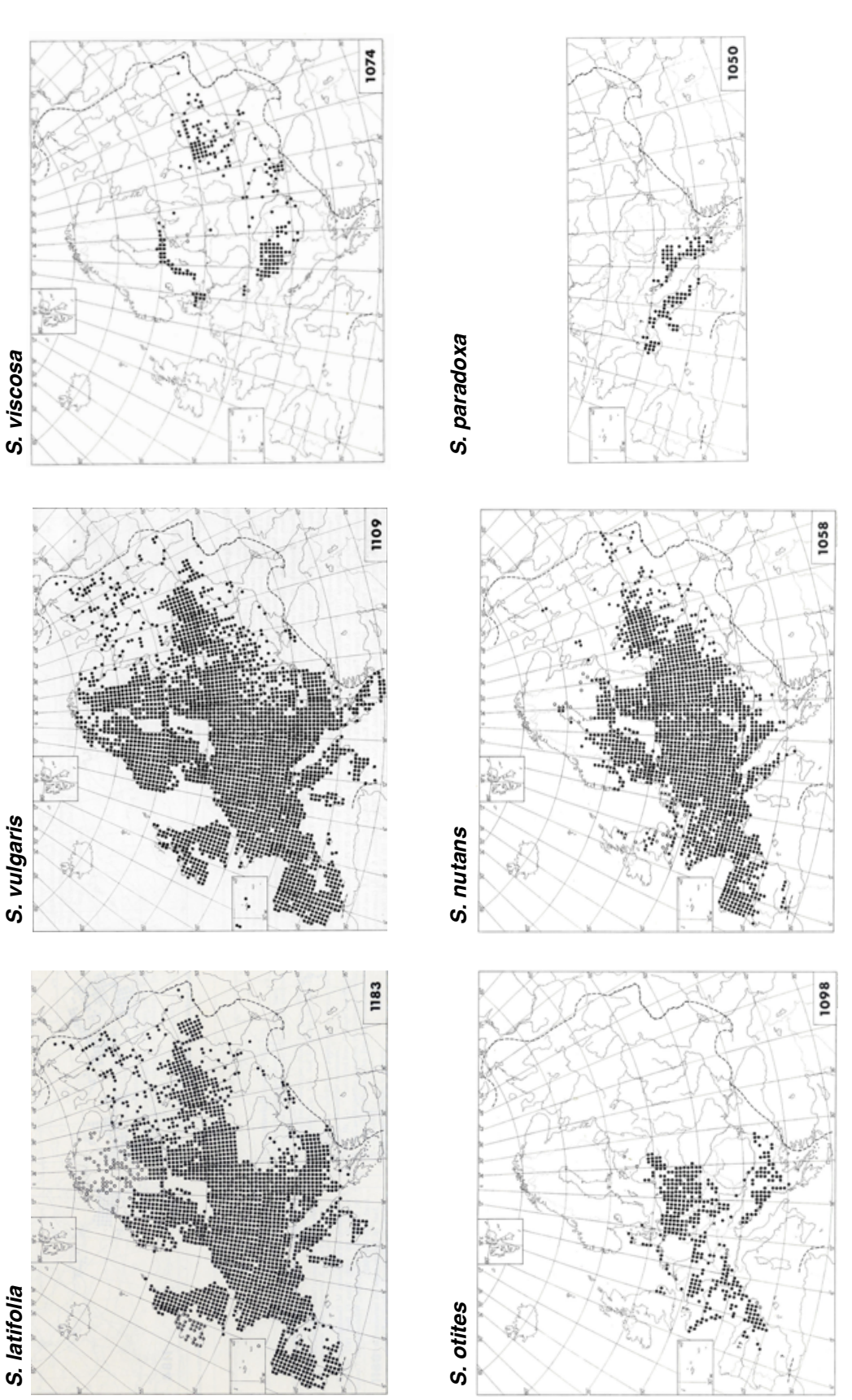



**Figure S6.1:** Geographic distribution of sampled *Silene* species (Jalas & Suominen, 1986). *Silene pseudotites* is widespread in Italy and adjacent parts of France and ex-Yugoslavia.

© 2016 Tous droits réservés.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

# Discussion

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



## 7.1 Résumé des chapitres

Chez *S. nutans*, des écotypes édaphiques génétiquement différenciés ont été décrits en Belgique (Van Rossum et al., 1997). Dans le **chapitre 1**, nous avons voulu déterminer si cette différenciation génétique est le résultat d'un contact primaire, associé à l'adaptation à la nature du sol, ou d'un contact secondaire suite à la remise en contact de lignées génétiques qui ont divergé en allopatrie. Nous avons utilisé une approche de phylogéographie et de génétique des populations sur un échantillonnage d'individus couvrant cinq pays d'Europe occidentale et génotypés sur treize marqueurs microsatellites et six SNP chloroplastiques. Sur la zone d'étude, nous avons décrit deux lignées nucléo-cytoplasmiques géographiquement structurées : une lignée *Ouest*, formée suite à une recolonisation postglaciaire vers le nord à partir de trois refuges glaciaires de l'ouest de l'Europe (dans le sud de la France, dans la péninsule Ibérique et en Italie), et une lignée *Est*, formée suite à une recolonisation postglaciaire vers le nord-ouest à partir de refuges glaciaires situé probablement en Europe de l'Est. Les deux lignées sont actuellement en contact en Angleterre, en Belgique ou en France, mais aucun flux de gènes n'a été détecté dans ces trois régions. Nous avons donc conclu que la différenciation génétique des écotypes belges était le résultat d'un contact secondaire entre deux lignées génétiques.

Ces écotypes sont isolés reproductivement : les hybrides entre ces écotypes sont en grande partie chlorotiques et la majorité d'entre eux ne survit pas au-delà du stade plantule (Van Rossum et al., 1996). Dans le **chapitre 2**, nous avons voulu déterminer si cet isolement reproducteur est le résultat d'un sélection disruptive liée à la nature du sol ou lors de la divergence des lignées en allopatrie. En serre, nous avons réalisé une expérience de croisements contrôlés avec des populations provenant des trois zones, en Angleterre, en Belgique et en France. Nous avons observé une importante mortalité au stade plantule (probablement liée à la chlorose observée dès la germination) chez les descendants des croisements interlignées par rapport aux descendants des croisements intralignées, que l'on croise des parents provenant de la même région ou de régions différentes. Nous avons donc conclu que l'isolement reproducteur postzygotique observé en Belgique est associé à la divergence en allopatrie des lignées *Ouest* et *Est*.

Si des flux de gènes ont eu lieu au cours de la divergence des lignées Ouest et Est, un scan génomique pourrait nous permettre d'identifier des régions candidates impliquées dans l'isolement reproducteur. Dans le **chapitre 3**, nous avons testé différents scénarios de divergence avec une approche ABC afin d'interpréter au mieux le patron de différenciation génétique observé le long du transcriptome. Le meilleur scénario démographique obtenu est un scénario d'isolement


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



strict des lignées *Ouest* et *Est* depuis leur séparation (estimé à 300.000 ans) avec une importante réduction de la taille efficace pour une partie du transcriptome. Cette réduction de taille efficace concerne les mêmes locus dans les deux lignées et la population ancestrale. Sur nos données observées, les locus présentant une valeur de $F_{ST}$ ntre les lignées *Ouest* et *Est* supérieure au 95ème centile (*les pics de* $F_{ST}$, présentent une réduction de diversité génétique par rapport au reste du transcriptome dans les lignées *Ouest* et *Est*, de même que leurs homologues chez S. otites et S. paradoxa. Ces résultats, associés à une hypothèse de synthénie entre les génomes des trois espèces, suggèrent que la sélection et la recombinaison ont façonné le patron hétérogène de différenciation génétique entre les lignées de S. nutans. Le scan transcriptomique ne nous permet donc pas de proposer des régions candidates impliquées dans l'isolement reproducteur entre les deux lignées.

Alors que les deux lignées de *S. nutans* manifestent un isolement reproducteur quasi-complet, les croisements entre deux espèces dioïques *S. otites* et *S. pseudotites* génèrent des hybrides viables (Sansome, 1938). Sansome (1938) propose une base génétique de cet isolement reproducteur incomplet entre *S. otites* et *S. pseudotites*. Pour expliquer le biais de sex-ratio, associés à de la chlorose et de la mortalité chez les hybrides entre *S. otites* et *S. pseudotites*, il propose un déterminisme du sexe différent entre les espèces. Dans le **chapitre 4**, nous avons voulu tester son hypothèse. Nous avons utilisé des données RNAseq d'une famille de *S. otites* et d'une famille de *S. pseudotites* et tenter de retrouver le type de ségrégation (autosomale, XY ou ZW) des marqueurs. Nous avons identifié, pour *S. pseudotites*, des locus liés au sexe sur le chromosome 6 de *S. latifolia*, avec une ségrégation de type XY. Pour *S. otites*, nous avons identifié des locus liés au sexe sur le chromosome 1 de *S. latifolia*, avec une ségrégation de type XY et sur le chromosome 3, avec une ségrégation de type ZW. Nous proposons d'expliquer les résultats de sex-ratio de Sansome (1938) par un conflit entre les différentes paires de chromosomes sexuels qui serait *remporté* par les chromosomes sexuel de *S. pseudotites*. Si *S. pseudotites* est utilisé comme parent maternel (donc avec les chromosomes sexuels XX) dans les croisements avec *S. otites*, la descendance est 100% femelle. Au contraire, si *S. pseudotites* est utilisé comme parent paternel (donc avec les chromosomes sexuels XY), les individus portant le chromosome X de *S. pseudotites* seraient femelles et ceux portant le chromosome Y seraient mâles.

Enfin, les systèmes de reproduction peuvent influencer les forces évolutives impliquées dans la fixation de nouvelles mutations et donc potentiellement de nouvelles incompatibilités. Notamment, le cul-de-sac évolutif de la dioécie décrit par Heilbuth (2000) peut être expliqué par une réduction de la taille efficace des taxons dioïques (Heilbuth et al., 2001, Vamosi and Otto, 2002), ce qui diminue l'efficacité de la sélection, pouvant mener à la fixation d'allèles délétères

© 2016 Tous droits réservés.

Thèse de Hélène Martin, Lille 1, 2016



et l'extinction des populations (Käfer et al., 2013). Dans le **chapitre 5**, nous avons voulu tester l'effet du système de reproduction sur l'efficacité de la sélection. Nous avons utilisé des données RNAseq de huit espèces dans le genre *Silene* : 2 hermaphrodites, 3 gynodioïques et 3 dioïques et estimé leurs niveaux de polymorphisme et de divergence. Nos résultats préliminaires montrent une corrélation entre l'efficacité de la sélection, le système de reproduction et la diversité génétique. En particulier, nous avons constaté que les espèces dioïques étaient moins efficaces pour fixer des mutations bénéfiques que les espèces gynodioïques pour des niveaux de diversité équivalents. Toutefois, nous n'avons pas encore d'explication à ce résultat.

## 7.2 Perspectives

### Spéciation chez *Silene nutans* ... et gynodioécie ?

Les trois premiers chapitres de cette thèse ont permis de montrer que la différenciation génétique et l'isolement reproducteur entre les écotypes belges ne sont pas en lien avec l'adaptation à la nature du sol, mais qu'il s'agit du produit de la divergence en allopatrie de lignées génétiques. Parce qu'aucun flux de gènes entre les lignées *Ouest* et *Est* n'a été détecté depuis leur séparation, nous ne pouvons pas utiliser le scan génomique ni l'étude de clines en zone hybride pour proposer des régions candidates impliquées dans l'isolement reproducteur. Parce que l'isolement reproducteur postzygotique est quasi complet, il est également impossible d'utiliser des approches de type QTL qui impliquent de faire des croisements hybrides sur plusieurs générations.

Un isolement reproducteur postzygotique aussi important mis en place en si peu de temps pose question. Les résultats du chapitre 5, bien que préliminaires, suggèrent que les lignées de *S. nutans* sont efficaces dans la fixation de mutations adaptatives. Or, la sélection positive accélère le temps de fixation de mutations bénéfiques (Kimura and Ohta, 1969, Li and Nei, 1977). Il est donc envisageable que les gènes impliqués dans l'isolement reproducteur ont été (ou sont) sous sélection. Parce que *S. nutans* est une espèce gynodioïque et donc le siège de conflit nucléo-cytoplasmique, on peut proposer que la *course aux armements* qui en découle est impliquée dans la divergence rapide de certains locus créant des incompatibilités génétiques de type nucléo-cytoplasmique chez l'individu hybride. Dans ce cadre, une piste à explorer serait l'étude de la coévolution des gènes impliqués dans les interactions nucléo-cytoplasmiques (Burton et al., 2013, Sloan, 2015). Toutefois, le polymorphisme cytoplasmique chez *S. nutans* semble être le résultat d'une sélection fréquence-dépendante qui maintient des haplotypes anciens (Touzet and Delph,


© 2016 Tous droits réservés.
lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



2009, Lahiani et al., 2013), ce qui contredit l'idée d'une coévolution nucléo-cytoplasmique indépendante entre les lignées puisqu'elles partageraient du polymorphisme cytoplasmique. Mais ces études ne se sont concentré que sur 2 gènes mitochondriaux et les forces évolutives qu'ils subissent ne sont pas forcément représentatives de celles exercées sur le reste des génomes (Sloan, 2015). Au sein de la lignée *Ouest*, trois groupes génétiques ont été identifiés : *Orange* (WI, provenant d'un refuge glaciaire dans le sud de la France), *Rouge* (W2, provenant d'un refuge glaciaire dans la péninsule Ibérique) et *Jaune* (W3, provenant d'un refuge glaciaire en Italie). Les résultats des croisements entre W1 et W3 ne montrent pas d'isolement postzygotique lors des cinq premières semaines du cycle de vie. Toutefois, en population naturelle, nous n'avons pas observé de flux de gènes entre les deux groupes génétiques (de même qu'entre W1 et W2). D'autres barrières à la reproduction sont peut-être impliquées : des barrières postzygotiques intrinsèques qui interviendraient plus tard dans le cycle de vie (stérilité), des barrières postzygotiques extrinsèques (les hybrides pourraient être moins compétitifs que les descendants intra-groupes génétiques), des barrières prézygotiques intrinsèques (une mauvaise interaction entre le grain de pollen et le stigmate) ou des barrières prézygotiques extrinsèques (des pollinisateurs différents). Si des barrières à la reproduction postzygotiques sont détectées, et si les hybrides sont néanmoins viables (et fertiles ?), il serait alors possible d'explorer l'architecture génétique de l'isolement reproducteur par des approches QTL. En parallèle, on pourrait également retracer l'histoire démographique des flux de gènes au cours du temps par une approche ABC. Si des flux de gènes sont inférés, il serait alors possible d'interpréter *les pics de* $F_{ST}$ omme des régions associées à l'isolement reproducteur. On pourrait alors confronter ces régions fortement différenciées avec les résultats de cartographie de QTL. Au cours de cette thèse nous avons présenté la lignée *Est* comme relativement homogène (nous n'avons pas détecté différents groupes génétiques). Toutefois, il faut garder en tête que l'aire de répartition de *S. nutans* est bien plus large que notre zone d'étude. En particulier, en Scandinavie, de la structure génétique avec des marqueurs allozymiques a été décrite (Van Rossum and Prentice, 2004, Van Rossum et al., 2016). En Europe de l'Est, une sous-espèce de *S. nutans* a également été décrite (*S. nutans* subsp *dubia*, Tutin et al., 1964). Sur l'ensemble de son aire de distribution, *S. nutans* apparaît alors comme un complexe d'espèces.

Enfin, rappelons que les études que nous avons réalisé sur *S. nutans* étaient motivées par la présence d'écotypes édaphiques reproductivement isolés en Belgique (De Bilde, 1973, Van Rossum et al., 1996, 1997). Bien que nous ayons retracé en partie l'histoire de ces écotypes, il nous manque la fin : comment (et pourquoi) l'adaptation édaphique s'est produite. Il est possible que les lignées soient pré-adaptées à un type de sol et que par exclusion compétitive, des écotypes

© 2016 Tous droits réservés.

 lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



se soient formés en Belgique. Pour explorer cette piste il serait possible de placer des individus provenant des deux lignées mais en dehors de la zone de contact, sur les deux types de sol en condition contrôlées et évaluer leur valeur sélective.

### Dioécie … et spéciation ?

Dans le cadre de l'étude de la spéciation entre *S. otites* et *S. pseudotites*, nous avons agi totalement différemment par rapport à ce que l'on a fait chez *S. nutans*. En se basant sur les données de Sansome (1938), nous avons cherché à tester son hypothèse d'un déterminisme du sexe différent entre les deux espèces expliquant les biais observés de sex-ratio dans les croisements réciproques. Nous nous sommes donc intéressés à l'architecture génétique de l'isolement reproducteur et non à l'histoire démographique des populations. Chez *S. otites*, le système polygénique du déterminisme du sexe que nous avons décrit suggère la présence de génotypes incompatibles (les femelles seraient XX ZW, les mâles seraient XY ZZ, donc les génotypes XY ZW et XX ZZ seraient absents). Ce déterminisme est donc très coûteux car si les chromosomes ségrègent indépendamment, la moitié des génotypes d'une génération F1 est absente (pour des raisons encore inconnues, probablement liées à une incompatibilité entre les paires de chromosomes sexuels). Pour vérifier ce biais dans les génotypes des descendants, il faudrait déterminer des marqueurs liés au sexe et génotyper toute une descendance (une extraction d'ADN au stade adulte pour vérifier l'absence de certains génotypes et dans un second temps au stade graine pour vérifier quels génotypes sont formés). Si le biais de génotypes est vérifié, il est peu probable que ce système soit stable.

Si le système du déterminisme du sexe chez *S. otites* est une étape transitoire, on devrait observer du polymorphisme en populations naturelles. Il devient donc nécessaire d'avoir une vision géographique de la structuration génétique à des marqueurs neutres (approche phylogéographique), mais aussi à des marqueurs liés au sexe. Cependant, parce que le système semble récent, un allèle lié au sexe dans une population ne l'est peut-être pas dans une autre. Cela ne veut pas forcément dire que le déterminisme du sexe n'est pas fixé. En effet un marqueur lié au sexe n'est pas forcément le marqueur du déterminisme du sexe. Si le système est récent, les marqueurs liés au sexe peuvent donc se retrouver à l'état polymorphe entre les populations parce que le polymorphisme ancestral est encore présent. Pour vérifier si le déterminisme du sexe est fixé ou non, il faudrait réaliser des croisements intra-populations dans différentes populations et regarder la ségrégation des marqueurs dans la descendance. Pour restreindre le nombre de marqueurs liés au sexe et cibler au mieux la(es) région(s) responsable(s) du déterminisme du sexe

© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



chez *S. otites*, il faudrait ajouter des évènements de recombinaison (au moins faire des générations F2). L'approche phylogéographique proposée plus haut permettrait également de clarifier la relation entre *S. otites* et *S. pseudotites*. Certes, les deux espèces présentent des morphologies différentes et apparemment un déterminisme du sexe différent, mais les croisements de Sansome (1938) montrent qu'il est possible d'obtenir des individus hybrides viables. En populations naturelles, il serait donc possible d'observer des flux de gènes entre les deux espèces. En particulier, si le déterminisme du sexe est polymorphe entre les populations attribuées à *S. otites*, dans quelle mesure *S. otites* et *S. pseudotites* ne seraient pas simplement le reflet d'un polymorphisme intra-spécifique ? Pour le savoir, il faudrait retracer l'histoire des populations de ces deux taxons et reproduire les expériences de croisement pour clairement établir les barrières postzygotiques entre les taxons pour estimer dans quelle mesure les paires de chromosomes sexuels sont réellement impliquées dans l'isolement reproducteur.

Si le système du déterminisme du sexe chez *S. otites* est une étape transitoire, encore faut-il déterminer *entre quoi et quoi*. A défaut de pouvoir prédire le futur, on pourrait estimer l'état ancestral. La sous-section *Otites* du genre *Silene* contient de nombreuses espèces dioïques (Jenkins and Keller, 2011, Slancarova et al., 2013). On pourrait dans un premier temps réaliser une phylogénie détaillée du système et identifier le déterminisme du sexe de chaque espèce par une approche similaire à celle développée dans le chapitre 4. En parallèle, une cartographie génétique des espèces pourrait être réalisée pour retracer de potentiels réarrangements chromosomiques ou des translocations génétiques potentiellement responsables d'un renouvellement du déterminisme du sexe.

## 7.3 Bibliographie

Burton, R. S., Pereira, R. J., and Barreto, F. S. (2013). Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematics*, 44 :281–302.

De Bilde, J. (1973). Etude génécologique du *Silene nutans* L. en Belgique : populations du *Silene nutans* L. sur substrats siliceux et calcaires. *Revue Générale de Botanique*, 80 :161–176.

Heilbuth, J. C. (2000). Lower species richness in dioecious clades. *The American Naturalist*, 156 :221–241.

Heilbuth, J. C., Ilves, K. L., and Otto, S. P. (2001). The consequences of dioecy for seed dispersal : modeling the seed-shadow handicap. *Evolution*, 55 :880–888.

Jenkins, C. and Keller, S. R. (2011). A phylogenetic comparative study of preadaptation for invasiveness in the genus *Silene* (Caryophyllaceae). *Biological Invasions*, 13(6) :1471–1486.

Käfer, J., Talianová, M., Bigot, T., Michu, E., Guéguen, L., Widmer, A., Žlůvová, J., Glémin, S., and Marais, G. A. B. (2013). Patterns of molecular evolution in dioecious and non-dioecious *Silene*. *Journal of Evolutionary Biology*, 26 :335–346.



© 2016 Tous droits réservés.

lilliad.univ-lille.fr

Thèse de Hélène Martin, Lille 1, 2016



Kimura, M. and Ohta, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics*, 61 :763–771.

Lahiani, E., Dufaÿ, M., Castric, V., Le Cadre, S., Charlesworth, D., Van Rossum, F., and Touzet, P. (2013). Disentangling the effects of mating systems and mutation rates on cytoplasmic diversity in gynodioecious *Silene nutans* and dioecious *Silene otites*. *Heredity*, 111 :157–164.

Li, W.-H. and Nei, M. (1977). Persistence of common alleles in two related populations or species. *Genetics*, 86 :901–914.

Sansome, F. W. (1938). Sex determination in *Silene otites* and related sepecies. *Journal of Genetics*, 35 :387–396.

Slancarova, V., Zdanska, J., Janousek, B., Talianova, M., Zschach, C., Zluvova, J., Siroky, J., Kovacova, V., Blavet, H., Danihelka, J., Oxelman, B., Widmer, A., and Vyskot, B. (2013). Evolution of sex determination systems with heterogametic males and females in *Silene*. *Evolution*, 67 :3669–3677.

Sloan, D. B. (2015). Using plants to elucidate the mechanisms of cytonuclear co-evolution. *New Phytologist*, 205 :1040–1046.

Touzet, P. and Delph, L. F. (2009). The effect of breeding system on polymorphism in mitochondrial genes of *Silene*. *Genetics*, 181 :631–644.

Tutin, T., Heywood, V., Burges, N., Valentine, D., Walters, S., and Webb, D., editors (1964). *Flora europaea*. Cambridge University Press, Cambridge.

Vamosi, J. C. and Otto, S. P. (2002). When looks can kill : the evolution of sexually dimorphic floral display and the extinction of dioecious plants. *Proceedings of the Royal Society B : Biological Sciences*, 269 :1187–1194.

Van Rossum, F., De Bilde, J., and Lefèbvre, C. (1996). Barriers to hybridization in calcicolous and silicolous populations of *Silene nutans* from Belgium. *Belgian Journal of Botany*, 129 :13–18.

Van Rossum, F. and Prentice, H. C. (2004). Structure of allozyme variation in Nordic *Silene nutans* (Caryophyllaceae) : population size, geographical position and immigration history. *Biological Journal of the Linnean Society*, 81 :357–371.

Van Rossum, F., Vekemans, X., Meerts, P., Gratia, E., and Lefèbvre, C. (1997). Allozyme variation in relation to ecotypic differentiation and population size in marginal populations of *Silene nutans*. *Heredity*, 78 :552–560.

Van Rossum, F., Weidema, I. R., Martin, H., Le Cadre, S., Touzet, P., Prentice, H. C., and Philipp, M. (2016). The structure of allozyme variation in *Silene nutans* (Caryophyllaceae) in Denmark and in north-western Europe. *Plant Systematics and Evolution*, 302 :23–40.

© 2016 Tous droits réservés.

lilliad.univ-lille.fr