

N° d'ordre 40538



Etude de l'autoincompatibilité chez la chicorée (*Cichorium intybus* L., Asteraceae) : Cartographie génétique et physique du locus *S*

THÈSE

soutenue publiquement le 5 Juillet 2011

pour l'obtention du

Doctorat de l'Université de Lille 1
Ingénierie des fonctions biologiques

par

Lucy Gonthier

Composition du jury

<i>Président :</i>	Xavier VEKEMANS, Professeur	GEPV, Université de Lille 1
<i>Rapporteurs :</i>	Christine MÉZARD, Directeur de Recherche INRA Simon HISCOCK, Professeur	INRA, Versailles-Grignon University of Bristol, UK
<i>Examineurs :</i>	Catherine FEUILLET, Directeur de Recherche INRA Hélène BERGÈS, Chargé de recherche INRA	INRA, Clermont-Ferrand INRA CNRGV, Toulouse
<i>Directeurs :</i>	Theo HENDRIKS, Professeur Marie-Christine QUILLET, Maître de conférence	SADV, Université de Lille 1 SADV, Université de Lille 1

Remerciements

Ce travail de thèse n'a été possible que parce que je ne l'ai pas réalisé seule. Bien évidemment, d'un point de vue scientifique, de nombreuses personnes ont participé à ce projet. Mais les soutiens extérieurs se sont avérés tout aussi importants pour que ce projet puisse aboutir. Je tiens donc à remercier toutes les personnes qui m'ont permis de finaliser ce travail, en espérant n'oublier personne.

Merci à Marie-Christine Quillet qui m'a encadrée durant ces quelques années de thèse, pour m'avoir fait confiance dans ce travail, pour ses conseils et pour s'être toujours montrée disponible, notamment durant la période difficile qu'est la rédaction du manuscrit. Merci également à Jean-Louis Hilbert, directeur du laboratoire, et Theo Hendriks, directeur de thèse, pour m'avoir accueillie au SADV.

Merci aux membres de mon jury, Christine Mézard, Simon Hiscock et Catherine Feuillet pour avoir consacré du temps à la lecture de ce manuscrit et pour leurs remarques constructives lors de la soutenance.

Merci également à Hélène Bergès et Xavier Vekemans, membres du jury mais aussi de mon comité de thèse, qui m'ont fait part de leurs conseils tout au long de ma thèse. Je tiens aussi à remercier Hélène pour m'avoir accueillie durant 1 mois au CNRGV, afin de participer à la construction des banques BAC. Au-delà de l'aspect formateur, ce séjour a permis de mettre en place une vraie collaboration avec l'équipe du CNRGV. Je tiens donc aussi à remercier Arnaud Bellec, Elisa Prat, Nicolas Helmstetter, Sonia Vautrin, Genseric Beydon et Joëlle Fourment, pour leur implication dans ce projet.

Merci à Christelle Blassiau, pour m'avoir initiée aux joies des AFLP et aidée en BM et à la serre, mais surtout pour sa bonne humeur quotidienne, sa gentillesse, son soutien et son amitié. Merci également à Monika Mörchen pour son aide en BM, sa "mauvaise" humeur et sa franchise.

Merci à Angélique Bourceaux, Cédric Glorieux, Eric Schmitt et Nathalie Faure qui ont pris soin de mes "quelques" chicorées à la serre. Merci également à Najia Voedts qui a pris soin des quelques autres conservées *in vitro* et pour sa recette de pastilla (que je n'ai toujours pas essayée).

Merci à Sylvain Legrand et Pauline Goubet pour leurs précieux conseils en bioinformatique.

Merci à Nicolas Lance et Audrey Caillaud, les stagiaires qui ont participé à ce projet. Et merci à toutes les personnes du laboratoire qui m'ont, à un moment ou un autre, encouragée et /ou aidée.

Merci à tous les "diplômitifs", Estelle, Aline, Jovana, Ildephonse, Matthieu, Nadim, Max, etc, ainsi qu'à Thierry, Philippe et Aurélie pour les bons moments passés ensembles, les tablettes de chocolats, les paquets de bonbons, les repas et les bières

partagés, et pour toutes nos discussions “très” constructives.

Merci bien sûr à tous mes amis grâce à qui j’ai passé de bons, voir très bons moments.

Merci aux thésards et “assimilés”, JB narcoleptique de fin de soirée, Isa et son humour toujours de bon goût, Camilla qui ne sait pas dire non, Camillo “since the Darwin’s ideas...”, Benon & Meriem, où comment organiser un mariage tout en rédigeant une thèse chacun, Fafa Saturday-night-fever, Professeur Jean-François, et bien sûr Aude (que je n’aime pas), sans oublier Manue, Marilyne, Romain,... avec qui j’ai passé la majorité de mes soirées ces 4 dernières années, les fins de thèse de chacun, quelques grands moments et un tas de petits moments salutaires.

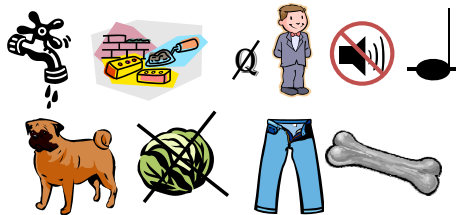
Merci aussi à tous mes amis plus anciens qui ne m’ont pas oubliée malgré cette grande parenthèse que représente la thèse, et avec qui j’ai aussi passé des moments mémorables, et plus particulièrement merci à Audrey, Julien (binôme choc de TP), Bob, Vincent, Julien et Sarah pour leur présence le jour de ma soutenance.

Merci également à toute la famille Laviale pour avoir pris de mes nouvelles régulièrement lors de la dernière ligne droite.

Merci à ma famille souvent bruyante, caustique, agaçante mais formidable. Une dédicace particulière à mon cher frère, que je remercie pour son soutien indéfectible.

Et surtout, un très grand merci à mes parents qui m’ont offert la chance de faire des études, m’ont toujours soutenue et fait confiance et sur lesquels je sais que je peux compter sans réserve.

Martin, pour m’avoir supportée, au sens propre comme au figuré, tout particulièrement durant cette longue et ingrate période de rédaction, pour m’avoir encouragée, aidée, ...



à Solange

Table des matières

Table des figures	9
Liste des tableaux	13
Introduction	17
1 Mécanismes d'autoincompatibilité	19
1.1 L'autoincompatibilité sporophytique hétéromorphe	21
1.2 L'autoincompatibilité gamétophytique	23
1.3 L'autoincompatibilité sporophytique homomorphe	29
1.4 Divergences et similitudes des systèmes d'AI homomorphe	34
2 L'autoincompatibilité sporophytique chez les Asteraceae : la chicorée comme modèle ?	36
1 Matériel et méthodes	45
1.1 Matériel végétal	45
1.1.1 Présentation des populations utilisées	45
1.1.2 Culture <i>in vitro</i>	46
1.1.3 Culture en serre	49
1.2 Outils de biologie moléculaire	50
1.2.1 Echantillons d'ADN	50
1.2.2 Marqueurs moléculaires spécifiques de séquence	51
1.2.3 Marqueurs AFLP	55
1.2.4 Transformation de marqueurs AFLP en marqueurs SCAR	59
1.2.5 Banques BAC	62
1.3 Cartographie génétique de la région du locus <i>S</i>	66

1.3.1	Calcul des distances génétiques	66
1.3.2	Etude de la ségrégation des marqueurs	66
1.3.3	Précision des distances génétiques en fonction de la taille de l'échantillon à cartographier	67
1.3.4	Détermination du génotype au locus S	70
I	Carte fine du locus S	73
2	Résultats préliminaires : Carte moyenne densité du locus S	75
2.1	Carte génétique	77
2.2	Cartographie du locus S	77
2.2.1	Identification des allèles S présents chez K28 et K59 et de leurs relations de dominance	77
2.2.2	Carte moyenne densité du locus S	79
2.3	Différence des fréquences de recombinaison...	80
3	Carte haute densité	85
3.1	BSA assistée par marqueurs	86
3.1.1	Intérêt de la BSA assistée par marqueurs	86
3.1.2	Construction des bulks	87
3.2	Recherche de nouveaux marqueurs liés au locus S	89
3.2.1	Criblage AFLP des bulks	89
3.2.2	Cartographie et validation des nouveaux marqueurs AFLP	90
3.2.3	Identification de nouveaux marqueurs microsatellites	92
3.3	Transformation des marqueurs AFLP en marqueurs SCAR	92
3.3.1	Intérêt de la démarche	92
3.3.2	Simplification des profils AFLP par miniséquençage	94
3.3.3	Spécificité des séquences correspondant aux fragments AFLP	95
3.3.4	Définition de nouvelles amorces et recherche de polymorphisme	96
4	Carte haute résolution	99
4.1	Détermination du nombre d'individus à cartographier	99

4.2	Sélection des populations pour l'établissement de la carte haute résolution	102
4.2.1	Caractéristiques des populations étudiées	102
4.2.2	Informations apportées par les nouveaux marqueurs	103
4.3	Criblage séquentiel des recombinants	105
4.3.1	Stratégie de génotypage en 3 étapes	105
4.3.2	Gestion des données manquantes	106
4.4	Construction de 2 cartes haute résolution	109
4.4.1	Effet de l'environnement : homogénéité des fréquences de recombinaison entre différents lots de graines pour les génotypes K28 et K59	110
4.4.2	Effet du génotype : homogénéité des fréquences de recombinaison au sein des méioses mâles et femelles	110
4.4.3	Effet du sexe sur la fréquence de recombinaison	112
4.4.4	Distorsion de ségrégation dans la région du locus S	112
4.4.5	Construction des cartes génétiques des parents et de 2 cartes consensus	114
4.4.6	Implication du locus S dans l'effet du sexe sur la fréquence de recombinaison ?	114
5	Discussion	119
II	Carte physique du locus S	129
6	Vers le clonage positionnel du locus S	131
6.1	Construction de banques BAC de Chicorée	131
6.2	Marche chromosomique	132
6.2.1	Stratégie mise en place afin de réaliser la marche chromosomique	132
6.2.2	Criblage des banques BAC	135
6.2.3	Caractéristiques des BES	138
6.2.4	Construction des contigs liés à l'haplotype S_2	140

7	Caractérisation de régions génomiques au voisinage du locus <i>S</i>	143
7.1	Séquençage des clones BAC	143
7.1.1	Caractéristiques générales des inserts séquencés	143
7.1.2	Organisation des contigs	146
7.2	Composition en éléments répétés	149
7.2.1	Elements répétés identifiés à partir de RepeatMasker et des Blastx	149
7.2.2	Identification d'autres séquences répétées	150
7.3	Etude de la fraction potentiellement codante à proximité du locus <i>S</i> .	151
7.3.1	Comparaison des séquences de chicorée et des banques d'EST d'Asteraceae	152
7.3.2	Composition en gènes des séquences de chicorée dans la région du locus <i>S</i>	154
7.3.3	Les logiciels de prédiction de gènes peuvent-ils être utilisés dans le cas de la chicorée?	163
7.3.4	Etude préliminaire de la conservation de la synténie	164
8	Discussion	173
	Conclusions et perspectives	185
	Bibliographie	193
	Annexes	213
A	Matériels et méthodes	215
A.1	Culture <i>in vitro</i>	215
A.2	Amorces PCR	217
B	Carte physique	221
B.1	Assemblage du supercontig	222

B.2	Annotation des gènes homologues d' <i>A. thaliana</i>	223
C	Gonthier <i>et al.</i> <i>BMC Research Notes</i> 2010	225
D	Synthesis and translations	237
D.1	Introduction	237
1.4	Difference and similarity of the homomorphic SI systems	237
2	Asteraceae sporophytic self-incompatibility : chicory as a model?	238
D.2	Part I : Fine mapping of the S-locus	240
Chapter 2	: Preliminary results : Moderate density map of the S-locus	240
Chapter 3	: High density map	241
Chapter 4	: High resolution mapping	241
Discussion	242
D.3	Part II : Physical map of the S-locus	243
Chapter 6	: Towards the positional cloning of the S-locus	243
Chapter 7	: Characterization of genomic region in the vicinity of the S-locus	244
Discussion	246
D.4	Conclusions and perspectives	247
E	Captions translation	251
Introduction	251
Chapter 1	: Materials and methods	252
Part I	: Fine mapping of the S-locus	253
Chapter 2	: Preliminary results : Moderate density map of the S-locus	253
Chapter 3	: High density map	254
Chapter 4	: High resolution mapping	255
Discussion	257
Part II	: Physical map of the S-locus	257
Chapter 6	: Towards the positional cloning of the S-locus	257
Chapter 7	: Characterization of genomic region in the vicinity of the S-locus	258
Discussion	260

Table des matières

Table des figures

1	Schéma interaction pollen-pistil	18
2	Répartition des différent types d'AI chez les Angiospermes	20
3	Schéma de la structure du locus <i>S</i> de l'AI homomorphique	21
4	Autoincompatibilité hétéromorphique	22
5	Autoincompatibilité gamétophytique	24
6	Modèle possible de l'interaction des S-RNases avec SLF/SFB	26
7	Interaction des protéines PrpS et PrsS	28
8	Autoincompatibilité sporophytique	29
9	Polymorphisme de structure au locus <i>S</i> chez <i>Brassica rapa</i>	31
10	Modèle moléculaire de la réponse autoincompatible chez les Brassicaceae	31
11	Chicorée (<i>Cichorium intybus</i>)	37
1.1	Différentes étapes du clonage positionnel et matériel végétal utilisé pour chacune d'entre elles.	46
1.2	Généalogie des différentes populations de chicorée utilisées	48
1.3	Culture <i>in vitro</i>	49
1.4	Culture en serre	50
1.5	Schéma des différentes étapes de la production de marqueurs AFLP . . .	55
1.6	Simplification du profil AFLP	60
2.1	Récapitulatif des résultats préliminaires	76
2.2	Seed-set et compatibilité	78
2.3	Relations de dominance entre haplotypes S	78
2.4	Carte moyenne densité du locus <i>S</i>	81
2.5	Comparaison des fréquences de recombinaison sur les 9 groupes de liaison de la chicorée	83

Table des figures

3.1	Cartes moyenne densité pour K28 e K59	86
3.2	Stratégie des bulks	88
3.3	Marqueurs AFLP obtenus par BSA	91
3.4	Génotypes graphiques et cartographie des marqueurs obtenus par BSA .	93
3.5	Cartes génétiques haute densité de K28 et K59	94
3.6	Exemple de simplification de profil AFLP	96
4.1	Représentation de l' i_r en fonction de r	100
4.2	Expression de n en fonction de r	102
4.3	Génotype graphique des parents des populations utilisées pour la carte haute résolution	104
4.4	Stratégie de génotypage en 3 étapes	107
4.5	Répartition des effectifs dans les différentes populations	108
4.6	Cartes génétiques de la région du locus S obtenues pour différentes popu- lations	115
4.7	Cartes génétiques consensus	116
6.1	Différentes étapes de la marche chromosomique vers le locus S	134
6.2	Macroarrays	138
6.3	Marche chromosomique vers le locus S	139
7.1	Organisation des Cg des clones séquencés	147
7.2	Critères de sélection utilisés pour les Blast	153
7.3	Fréquences des portions de séquence alignées sur les séquences de clones BAC	154
7.4	Distributions des différents types d'éléments chez les 6 clones séquencés .	155
7.4	Distributions des différents types d'éléments chez les 6 clones séquencés - suite	156
7.4	Distributions des différents types d'éléments chez les 6 clones séquencés - fin	157
7.5	Distribution des gènes identifiés sur les chromosomes d' <i>A. thaliana</i> et de la vigne	159
7.6	Arbre phylogénétique des Dicotylédones...	165
7.7	Distribution des gènes pO et H(*) du chromosome 18 de la vigne...	167
7.8	Conservation de la synténie...	169

8.1	Représentation schématique des blocs de gènes	183
-----	---	-----

Table des figures

Liste des tableaux

1	Différentes caractéristiques de chaque type d'autoincompatibilité	19
1.1	Différentes populations utilisées dans le cadre du clonage positionnel du locus <i>S</i> chez la chicorée	47
1.2	Mélange PCR pour l'amplification de marqueurs STS	52
1.3	Mélange de migration sur séquenceur ABI	54
1.4	Mélange PCR utilisé dans le cadre de la technique HRM	54
1.5	Mélange de digestion enzymatique	56
1.6	Mélange de ligation	56
1.7	Mélange de préamplification	57
1.8	Mélange d'amplification sélective	58
1.9	Amorces EcoRI-A+2 et MseI-C+2 utilisées pour l'amplification sélective	58
1.10	Amorces dégénérées	61
1.11	Composition d'un gel polyacrylamide	61
1.12	Bases de données EST des différentes espèces d'Asteraceae utilisées . . .	65
1.13	Bases de données peptidiques utilisées	65
1.14	Expression de i_r en fonction du type de marqueurs considéré	68
1.15	Exemples de fréquences attendues pour différents génotypes au sein d'une descendance...	69
1.16	Schémas de croisements	72
2.1	Groupes de compatibilité entre allèles <i>S</i> et stratégies de croisement . . .	79
3.1	Marqueurs AFLP identifiés avec la stratégie des bulks	90
3.2	Résultats du miniséquençage	95
3.3	Comparaison des marqueurs AFLP et SCAR	97

Liste des tableaux

4.1	Valeurs minimales de r pouvant être calculées	106
4.2	Tests d'homogénéité	111
4.3	Ségrégation des haplotypes S	113
4.4	Effet du sexe sur r pour différents intervalles de la région du locus S	117
5.1	Approches BSA et BSA assistées par marqueur	121
5.2	Localisation chromosomique du locus S et relation Mb/cM au voisinage du locus S pour différentes espèces	122
6.1	Caractéristiques des banques BAC de chicorée	133
6.2	Exemple de construction d'un contig à partir des marqueurs BES	133
6.3	Hybridations sur la banque CinS2S2	136
6.4	Hybridations sur la banque CinS1S4	137
6.5	Exemple de difficultés d'assemblage entre haplotypes S	137
7.1	Caractéristiques des séquences de clones issus de la banque BAC CinS2S2	145
7.2	Caractérisation des abords du locus S en éléments répétés	151
7.3	Gènes identifiés à proximité du locus S	161
8.1	Composition génomique en éléments répétés pour différentes espèces	177
A.1	Milieus d'induction de l'embryogénèse somatique et de culture <i>in vitro</i>	215
A.2	Macroéléments	215
A.3	Microéléments	216
A.4	Vitamines de Morel et Wetmore (1951)	216
A.5	Solution de Fe-EDTA	216
A.6	Séquences et caractéristiques des amorces	217
B.1	Schéma d'assemblage du supercontig	222
B.2	Annotation des gènes homologues d' <i>A. thaliana</i>	223

Introduction

Introduction

Introduction

Le groupe des Angiospermes est largement prédominant chez les végétaux terrestres et est caractérisé par de nombreuses stratégies de reproduction sexuée. Comme 75% des Angiospermes sont hermaphrodites, ces stratégies sont généralement définies en fonction du taux d'autofécondation et vont de l'autogamie à l'allogamie strictes, en passant par des systèmes de reproduction mixte (Gaude et Cabrillac, 2001 ; Gaude *et al.*, 2001 ; Goodwillie *et al.*, 2005). Si l'autofécondation confère un avantage sélectif en augmentant le nombre de copies de génomes qu'un individu peut transmettre à sa descendance par rapport à une espèce strictement allogame, elle entraîne également une perte de la diversité génétique (Charlesworth et Charlesworth, 1987). On observe aussi un grand nombre de mécanismes qui limitent ou empêchent l'autofécondation et donc la consanguinité. Il existe des systèmes d'évitement spatial (distylie, monoécie) ou temporel (protandrie, protogynie), de séparation des sexes (dioécie), mais c'est l'autoincompatibilité qui est la stratégie la plus répandue chez les plantes à fleurs : on la retrouve chez plus de 50% des espèces d'Angiospermes (Gaude et Cabrillac, 2001 ; Hiscock, 2000a ; Hiscock et Tabah, 2003 ; Silva et Goring, 2001).

L'autoincompatibilité se définit comme l'incapacité d'une plante hermaphrodite fertile à produire des zygotes après autopollinisation (de Nettancourt, 1977). Les mécanismes d'autoincompatibilité permettent aux individus d'une espèce de ne pas être fécondés par leur propre pollen et/ou le pollen d'individus apparentés. Cela implique une interaction pollen-pistil permettant la reconnaissance d'un phénotype commun, comme cela est schématisé dans la figure 1. Les données moléculaires dont on dispose indiquent que ces mécanismes sont apparus plusieurs fois de façons indépendantes (Hiscock, 2000b) et qu'ils se sont maintenus au cours du temps (Igic *et al.*, 2004). Ils doivent donc avoir une grande importance évolutive et l'acquisition de l'autoincompatibilité par les premiers Angiospermes serait une des clés de leur importante radiation adaptative (Hiscock et Tabah, 2003).

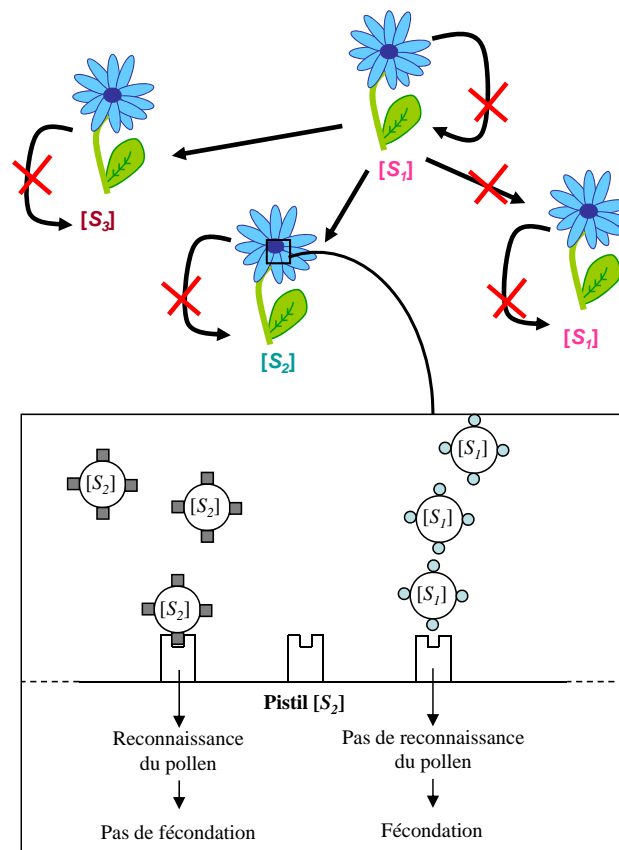


FIGURE 1 – Schéma représentant l'interaction pollen-pistil intervenant dans les mécanismes d'autoincompatibilité. Lorsque le pollen et le pistil présentent le même phénotype, un mécanisme de reconnaissance empêche la fécondation. Par contre, si le pollen et le pistil présentent des phénotypes différents, la fécondation a lieu.

Un grand nombre d'articles, revues et livres détaillent des différents mécanismes d'autoincompatibilité identifiés. L'objectif de cette introduction n'est donc pas d'en faire un compte-rendu exhaustif. Il s'agit plutôt d'offrir une vue d'ensemble afin de mettre en évidence la diversité ainsi que les similitudes qui existent entre ces stratégies.

1 Mécanismes d'autoincompatibilité

L'autoincompatibilité peut être classée en 3 catégories : sporophytique hétéromorphe (AI Het), gamétophytique homomorphe (AIG) et sporophytique homomorphe (AIS) (Gibbs, 1986 ; Gibbs et Bianchi, 1999). Les caractéristiques de chacune de ces catégories sont récapitulées dans le tableau 1. Ayant été respectivement identifiées dans 17 (Franklin-Tong et Franklin, 2003 ; Hiscock, 2000b) et 28 familles de plantes (Barrett et Shore, 2008), l'AIG et l'AI Het sont les plus communes (Figure 2). L'AIS a, quant à elle, été observée dans 6 familles (Figure 2, Hiscock et Tabah, 2003).

TABLEAU 1 – Récapitulatif des différentes caractéristiques de chaque type d'autoincompatibilité. a)Hiscock (2000b), b)Franklin-Tong et Franklin (2003), c)Hiscock et Tabah (2003), d)Barrett et Shore (2008), e)Allen et Hiscock (2008).

Autoincompatibilité		Locus	Nombre Allèles	Familles	Référence
Homomorphe	Gamétophytique	1	multiallèlique	16	a
		2	multiallèlique	1	b
	Sporophytique	1	multiallèlique	6	c
Hétéromorphe	Sporophytique	1 (distylie)	2	24	d
		2 (tristylic)	2 par locus	4	e

L'AI Het est associée à un polymorphisme floral entre individus, généralement sous le contrôle d'un seul gène diallèlique (McCubbin, 2008). L'AIG et l'AIS présentent des fleurs de morphologie identique. Dans le cas de l'AIG, le phénotype d'incompatibilité du grain de pollen est déterminé par son propre génotype haploïde, alors que pour les systèmes sporophytiques, le phénotype d'incompatibilité du grain de pollen est déterminé par le génotype diploïde de la plante dont il est issu (de Nettancourt, 2001). Dans la plupart des cas, l'autoincompatibilité homomorphe est sous contrôle d'un seul locus multiallèlique, appelé locus *S*, qui regroupe les gènes fortement liés des déterminants mâle et femelle de l'incompatibilité (Franklin-Tong et Franklin, 2003 ; Silva et Goring, 2001 ; Takayama et Isogai, 2005). Afin de décrire la combinaison d'allèles de ces différents gènes, on utilise le terme d'haplotype *S* (Nasrallah et Nasrallah, 1993, Figure 3). Il

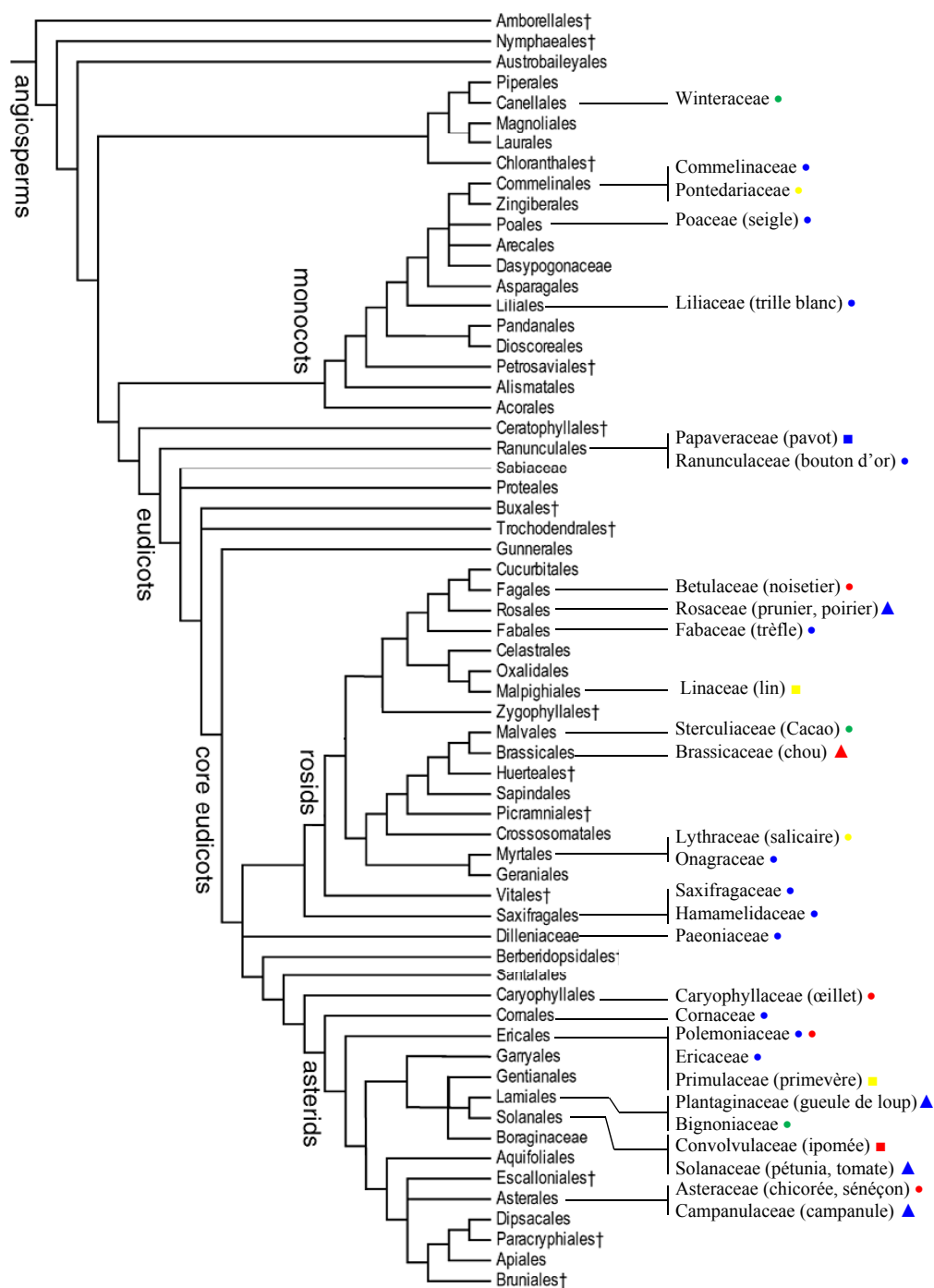


FIGURE 2 – Répartition des différents types d’AI chez les Angiospermes (phylogénie d’après AGP III, Bremer *et al.*, 2009). AI G : ● mécanisme indéterminé, ▲ S-RNases, ■ PrpS-PrsS; AIS : ● mécanisme indéterminé, ▲ SRK-SCR/SP11, ■ SE1/SE2/SEA-AB2; AI Het (liste non-exhaustive) : ■ distylie, ● tristylie; AI ovarienne (liste non-exhaustive) : ● (Allen et Hiscock, 2008 ; Barrett et Shore, 2008 ; de Nettancourt, 1997 ; Eenink, 1981a ; Franklin-Tong et Franklin, 2003 ; Gibbs et Bianchi, 1999 ; Hiscock, 2000a ; Klein *et al.*, 2009 ; Labonne *et al.*, 2008 ; Rahman *et al.*, 2007b ; Takayama et Isogai, 2005).

existe également des cas de systèmes multi-locus, comme chez les Poaceae où 2 locus interviennent (Kakeda *et al.*, 2008 ; Yang *et al.*, 2008). Dans certaines espèces, des mécanismes faisant intervenir jusqu'à 4 locus ont été proposés (par exemple : *Lilium martagon*, *Beta vulgaris*, Allen et Hiscock, 2008 ; Yang *et al.*, 2008). Ceci dit, il s'agit d'espèces où peu d'études ont été menées et où un système à 2 locus pourrait également être proposé (Richards, 1997). Une quatrième forme d'autoincompatibilité a été identifiée pour laquelle la réaction se produit une fois que le tube pollinique a atteint l'ovaire, avant ou après la fécondation : l'autoincompatibilité ovarienne ou tardive (Gibbs et Bianchi, 1999). Pour cette forme d'AI, il est difficile de déterminer s'il s'agit d'un mécanisme génétique ou de l'expression d'une dépression de consanguinité.

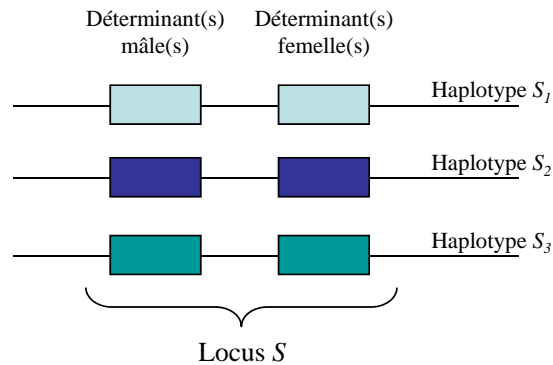


FIGURE 3 – Schéma de la structure du locus S de l'AI homomorphique. Il existe au moins 2 gènes génétiquement liés, l'un codant le déterminant mâle et l'autre le déterminant femelle. Chaque combinaison d'allèles constitue un haplotype différent.

1.1 L'autoincompatibilité sporophytique hétéromorphique

L'AI Het est polyphylétique et a été observée dans 28 familles appartenant à des sous classes et ordres différents (Figure 2) : cela suggère que ce mécanisme est apparu plusieurs fois durant la diversification des Angiospermes (Figure 2, Allen et Hiscock, 2008 ; Gibbs, 1986 ; McCubbin, 2008). Ce mécanisme concerne les espèces pour lesquelles on observe, entre les fleurs des individus, des différences morphologiques associées au mécanisme d'autoincompatibilité. Les morphes diffèrent par la longueur des étamines et des styles. On peut également observer du polymorphisme au niveau du pollen et du stigmate (Barrett *et al.*, 2000 ; de Nettancourt, 2001). La majorité des familles AI Het sont distyles (comme par exemple *Primula sp.*) : on y rencontre ainsi des individus à styles longs et anthères courtes (morphe L) et d'autres à styles courts et anthères longues (morphe C)

Introduction

(Figure 4, Barrett et Shore, 2008). La tristylie demeure un mécanisme d'AI Het relativement peu fréquent en comparaison de la distylie : 4 familles seulement (Allen et Hiscock, 2008) se caractérisent par des individus à styles longs, moyens ou courts (morphes L, M et C) (Figure 4, Barrett et Shore, 2008). En plus de l'évitement spatial entre pollen et stigmate qu'engendrent les différentes longueurs de styles, le mécanisme d'autoincompatibilité empêche également les croisements entre individus de morphes identiques : seuls les génotypes ayant des styles et des anthères de même taille sont compatibles (Barrett et Shore, 2008).

Le déterminisme génétique de l'AI Het est caractérisé par un seul locus S dans le cas de la distylie et de 2 locus, S et M , pour la tristylie. Il s'agit généralement de locus diallèles qui déterminent l'ensemble des traits qui distinguent les différentes morphologies (Barrett et Shore, 2008). Le site de la réaction d'incompatibilité varie selon la famille, l'espèce et la morphologie considérées et peut avoir lieu au niveau du stigmate, du style ou de l'ovaire (de Nettancourt, 2001).

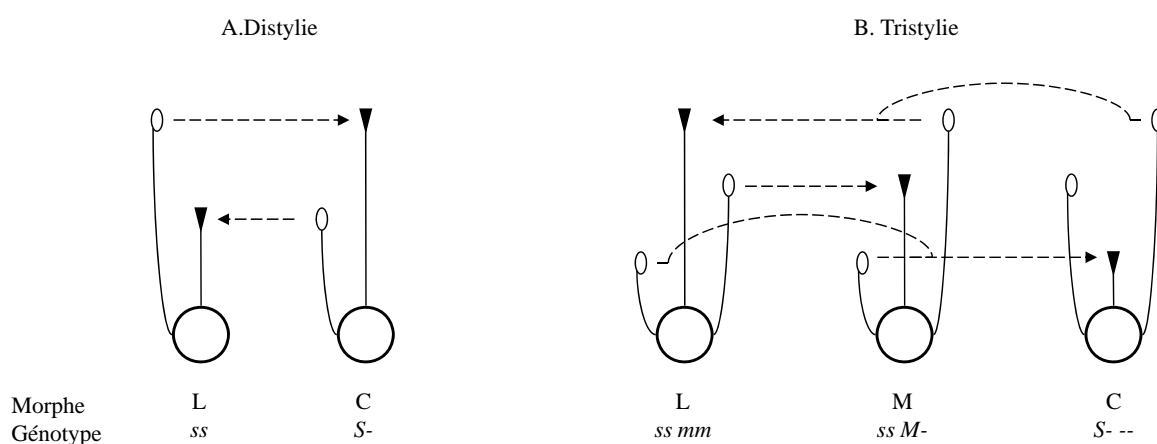


FIGURE 4 – Autoincompatibilité hétéromorphique. L, M et C correspondent aux morphes long, moyen et court du style. L'allèle S du locus S est dominant et gouverne le morphe C. Dans le cas de la tristylie, les individus ss ont soit un morphe L si leur génotype au locus M est mm , soit un morphe M s'ils possèdent au moins un allèle M . Les anthères sont représentées par un ovale blanc et les stigmates par un triangle noir. Les croisements compatibles sont indiqués par des flèches en tirets (d'après Barrett *et al.*, 2000).

Plusieurs approches ont été utilisées pour identifier les déterminants responsables de l'AI Het mais malgré les nombreuses études réalisées, les gènes responsables ne sont toujours pas identifiés (Barrett et Shore, 2008 ; McCubbin, 2008). L'approche biochimique a permis d'identifier des protéines spécifiques d'un morphe par rapport à un autre mais leur fonction n'a pas été définie et aucune n'a pu être liée au locus S . Par exemple, jusqu'à

177 protéines variant d'un morphe à l'autre ont été identifiées chez *Lythrum salicaria*. Ces différences peuvent s'expliquer par le nombre important de protéines différentes mises en jeu pour obtenir les différentes morphologies. Dans le cas de *Primula vulgaris*, une banque soustractive d'ADNc a été construite, mais aucun des ADNc exprimés différentiellement selon les morphes n'a pu être lié au locus *S*. Enfin, le locus *S* a été cartographié dans des intervalles de 0,6 et 0,57 cM, respectivement chez *Fagopyrum sp* et *Primula vulgaris*, ce qui permet d'envisager l'identification des déterminants de l'AI Het par clonage positionnel (Barrett et Shore, 2008 ; McCubbin, 2008).

1.2 L'autoincompatibilité gamétophytique

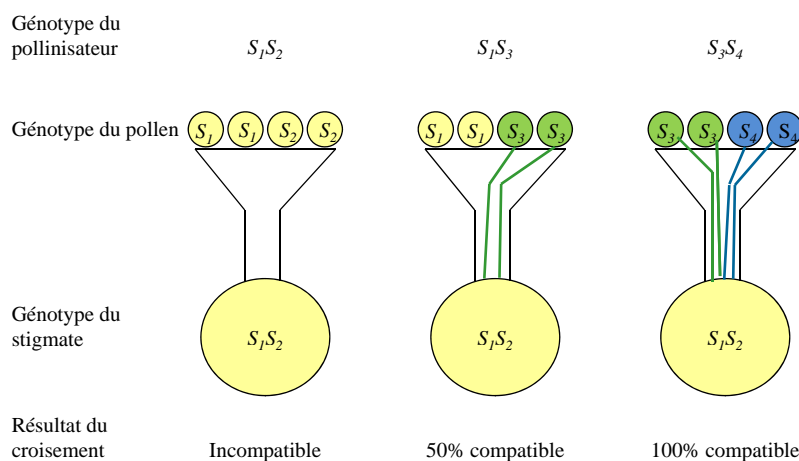
L'AIG est le système d'autoincompatibilité homomorphique le plus répandu. Il est étudié au niveau moléculaire principalement chez les Solanaceae, les Rosaceae, les Plantaginaceae (anciennement Scrophulariaceae, Olmstead *et al.*, 2001), les Papaveraceae et dans une moindre mesure, chez les Poaceae (Figure 2). Comme le phénotype des grains de pollen est défini par le génotype haploïde du gamétophyte, il n'existe pas de relations de dominance entre haplotypes *S*. Par conséquent, si le génotype correspond à l'un des 2 allèles *S* (ou des 4 allèles dans le cas des Poaceae) exprimés au niveau du pistil, alors la réaction d'autoincompatibilité a lieu (Figures 5).

1.2.1 Système gamétophytique des Solanaceae, des Rosaceae et des Plantaginaceae

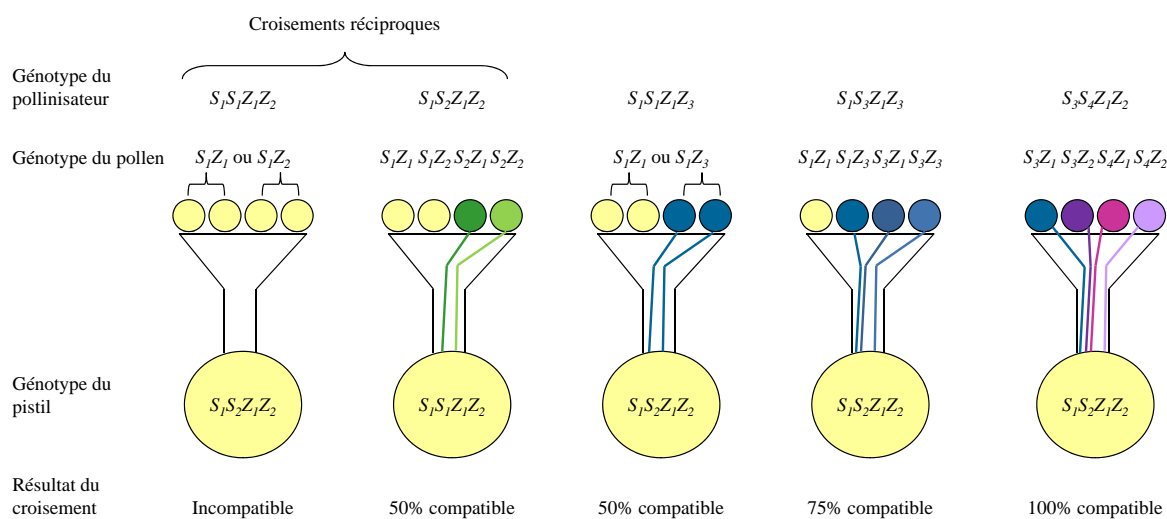
Dans ces familles, la réaction d'incompatibilité a lieu dans le style. Les grains de pollen, compatibles et incompatibles, commencent indifféremment leur germination. Dans le cas du pollen incompatible, on observe une croissance anormale du tube avant son arrêt au niveau du premier tiers du style (Takayama et Isogai, 2005). L'incompatibilité est ici contrôlée par un locus *S* multiallélique (jusqu'à 28 allèles différents chez *Physalis crassifolia*) où se trouvent 2 gènes codant respectivement les déterminants mâle et femelle (Lawrence, 2000 ; McClure et Franklin-Tong, 2006 ; Takayama et Isogai, 2005).

L'approche protéomique par IEF (isoélectrofocalisation) réalisée chez *Nicotiana glauca* a permis d'identifier des protéines spécifiques de différents allèles *S*, présentes uniquement dans le stigmate et le style (Bredemeijer et Blaas, 1981). L'ADNc correspondant à la protéine spécifique de l'allèle *S*₂ de *N. glauca* a été séquencé (Anderson *et al.*, 1986). Le gène identifié comme le déterminant femelle code une protéine ayant une activité

Introduction



(a) Autoincompatibilité gamétophytique à un seul locus, le locus S .



(b) Autoincompatibilité gamétophytique à 2 locus, les locus S et Z .

FIGURE 5 – Autoincompatibilité gamétophytique. (a) : La réaction d'incompatibilité survient quand l'allèle S exprimé par le pollen correspond à l'un des 2 allèles S exprimés au niveau du pistil (d'après Silva et Goring, 2001). (b) : La réaction d'autoincompatibilité a lieu lorsque les 2 allèles S et Z exprimés par le pollen correspondent à 2 des 4 allèles exprimés au niveau du pistil (d'après Yang *et al.*, 2008).

ribonucléase, celle-ci étant appelée S-RNase (Zhang et Xue, 2008). Le niveau d'expression de ce gène augmente progressivement dans le style avec la maturité de la fleur (Franklin-Tong et Franklin, 2003 ; Silva et Goring, 2001). Les séquences alléliques sont composées de régions conservées et de régions hypervariables entre spécificités S, ce qui se traduit par une faible similitude des séquences d'acides aminés (entre 40% et 70% d'identité entre allèles *S* chez *N. alata*, Ioerger *et al.*, 1990 ; Silva et Goring, 2001).

Afin d'identifier le déterminant mâle, une étude de l'expression différentielle des ARN de différents haplotypes a permis d'identifier 10 groupes d'ADNc spécifiques du pollen et liés au locus *S*. Cependant, le déterminant pollen n'a pu être identifié par cette voie (McCubbin *et al.*, 2000). Dans un second temps, la région génomique à proximité du gène *S-RNase* a été séquencée et sa structure étudiée, d'abord chez *Antirrhinum majus* (Lai *et al.*, 2002) puis chez *Prunus mume* et *Prunus dulcis* (Entani *et al.*, 2003 ; Ushijima *et al.*, 2003). Cette stratégie a permis d'identifier le gène *S-Locus F-box* (*SLF*), également appelé *S haplotype-specific F-box* (*SFB*), qui code le déterminant pollen (Entani *et al.*, 2003 ; Ushijima *et al.*, 2003). Tout comme les S-RNases, les protéines SLF/SFB présentent un polymorphisme élevé (entre 68 et 76% d'identité entre les séquences d'acides aminés chez *P. dulcis*, Ushijima *et al.*, 2003).

En plus d'identifier le gène *SLF/SFB*, l'étude génomique du locus *S* chez *Prunus dulcis* a montré que les déterminants de l'AI sont localisés dans une région fortement divergente de 70 kb encadrée par des régions conservées (Entani *et al.*, 2003). Une étape de validation fonctionnelle par transformation a confirmé que le gène codant les S-RNases est suffisant pour le rejet du pollen incompatible, et que le gène *SLF/SFB* est bien le déterminant pollen (Zhang et Xue, 2008).

La S-RNase détermine la spécificité du pistil et est responsable du rejet du pollen incompatible (Entani *et al.*, 2003 ; McClure et Franklin-Tong, 2006). Elle passe de la matrice extra-cellulaire du style au cytoplasme du tube pollinique en cours de germination. Si le pollen est incompatible, elle dégrade les ARN, ce qui provoque un arrêt de croissance du tube pollinique en empêchant l'expression des gènes du pollen. SLF/SFB appartient à la famille des protéines F-box dont la principale caractéristique est de lier un substrat spécifique à un complexe ligase E3-ubiquitine pour qu'il soit dégradé. Une première hypothèse considère que les S-RNases reconnues comme 'non-soi' par SLF/SFB sont dégradées par la voie des ubiquitines alors que les S-RNases reconnues comme 'soi' par SLF/SFB ne le sont pas (Figure 6).

Un second modèle suggère que les S-RNases, qui pénètrent sans discrimination dans

Introduction

le tube pollinique, pourraient être stockées dans une vacuole. Dans le cas d'un pollen compatible elles seraient éliminées du tube pollinique par exocytose et, si le pollen est incompatible, ces S-RNases seraient libérées dans le cytoplasme. Cependant, dans ce modèle, le rôle de SLF/SFB et son mode d'interaction avec les S-RNases ne sont pas encore définis (McClure et Franklin-Tong, 2006).

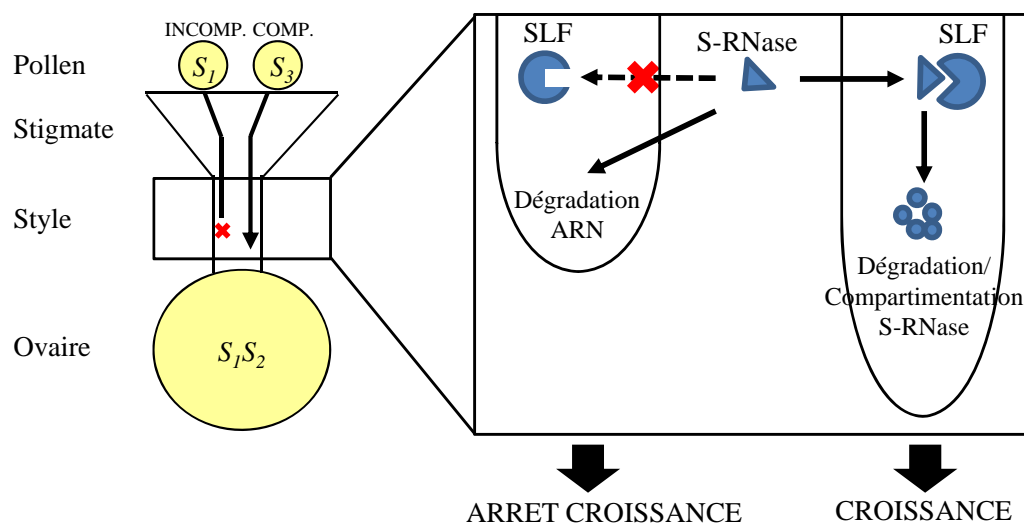


FIGURE 6 – Modèle possible de l'interaction des S-RNases avec SLF/SFB, chez les Solanaceae, Rosaceae et Plantaginaceae (d'après Takayama et Isogai, 2005).

1.2.2 Système gamétophytique des Papaveraceae

Chez les Papaveraceae, l'autoincompatibilité se caractérise par une inhibition rapide de la croissance du tube pollinique à la surface du stigmate (Silva et Goring, 2001 ; Thomas *et al.*, 2006). L'autoincompatibilité est contrôlée par un locus *S* multiallèlique (66 allèles connus pour *Papaver rhoeas*, Silva et Goring, 2001 ; Takayama et Isogai, 2005). Les gènes des déterminants mâle et femelle ont été identifiés chez *Papaver rhoeas* et codent des petites protéines respectivement appelées PrpS (*Papaver rhoeas* pollen S, 20kDa) et PrsS (*Papaver rhoeas* stigma S, 15kDa) (McClure et Franklin-Tong, 2006 ; Wheeler *et al.*, 2009).

PrsS est une protéine sécrétée par le stigmate. Elle a été identifiée, comme chez les Solanaceae, grâce à une approche protéomique par IEF puis par clonage et séquençage de l'ADNc correspondant (Foote *et al.*, 1994). Le gène codant PrsS est très polymorphe (40 à 46% d'homologie entre les séquences d'acides aminés des différents allèles *S*). Il

appartient à une famille de gènes codant des protéines de fonctions inconnues, les SPH (S-protein homologue), dont on retrouve 84 membres chez *A. thaliana* (Ride *et al.*, 1999 ; Wheeler *et al.*, 2010). L'inhibition de la germination du pollen S_1 placé sur un milieu de culture contenant la protéine Prs S_1 a permis la validation fonctionnelle de PrsS en tant que déterminant femelle (Foote *et al.*, 1994).

Le gène *PrpS* a été identifié grâce au séquençage de la région génomique du locus *S*, à proximité du déterminant femelle (Wheeler *et al.*, 2009). On observe une divergence de 40 à 50% entre les séquences d'acides aminés des différents allèles *S* (Wheeler *et al.*, 2010). PrpS ne présente aucune homologie avec les protéines déjà identifiées. Il s'agit vraisemblablement d'une protéine transmembranaire localisée au niveau du tube pollinique qui interagit avec PrsS et est impliquée dans l'inhibition du pollen incompatible (Franklin-Tong et Franklin, 2003 ; Takayama et Isogai, 2005 ; Wheeler *et al.*, 2009). De même que pour PrsS, la validation fonctionnelle de PrpS en tant que déterminant mâle a été réalisée dans un système de bioessai *in vitro*. Les auteurs ont montré que, lorsque du pollen S_1 était déposé sur un milieu de culture contenant des oligonucléotides antisens spécifiques des transcrits du gène *PrpS_1*, la production de cette protéine était inhibée de façon significative. En présence de la protéine Prs S_1 , ils ont observé une réduction significative de l'inhibition de la germination du pollen S_1 , montrant ainsi que Prp S_1 était le déterminant pollinique de l'AI chez les Papaveraceae (Wheeler *et al.*, 2009).

Chez *P. rhoeas*, la réaction d'incompatibilité déclenche une augmentation de la concentration cytosolique d'ions Ca^{2+} dans le tube pollinique et une diminution dans son extrémité. Cette augmentation déclenche une cascade réactionnelle causant une mort cellulaire programmée (PCD : programmed cell death) du tube pollinique (Figure 7). Même si plusieurs protéines sont activées ou inhibées, la cause principale de cette mort cellulaire est la dépolymérisation de plus de 60% de l'actine-F qui entraîne une réorganisation du cytosquelette et l'arrêt de la croissance du tube pollinique (Franklin-Tong et Franklin, 2003 ; Thomas *et al.*, 2006).

1.2.3 Système gamétophytique des Poaceae

Pour la plupart des Poaceae autoincompatibles (par exemple *Secale cereale* ou *Gaussia fragilis*), la réaction d'incompatibilité est rapide et se produit à la surface du stigmate (Langridge et Baumann, 2008 ; Yang *et al.*, 2008). Dans cette famille, 2 locus sont impliqués : le locus *S* et le locus *Z* (Figure 5). Ces locus sont indépendants et localisés sur des groupes de liaisons différents (Langridge et Baumann, 2008). Une diversité allélique

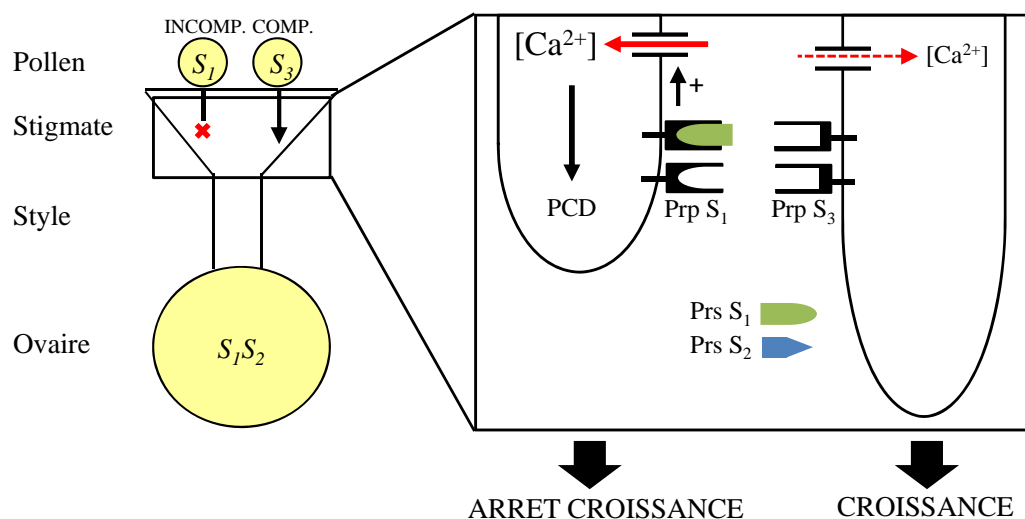


FIGURE 7 – Cascade de réactions induite par l'interaction des protéines PrpS et PrsS chez les Papaveraceae (d'après Franklin-Tong et Franklin, 2003).

élevée (par exemple, 17 allèles *S* et *Z* ont été identifiés chez *Lolium perenne*) associée à la combinaison des 2 gènes font de l'AIG des Poaceae un système particulièrement complexe (Langridge et Baumann, 2008 ; Lawrence, 2000). Les déterminants génétiques et les mécanismes moléculaires associés à l'incompatibilité des Poaceae ne sont pas encore identifiés.

Plusieurs approches (gènes candidats et approches biochimiques) ont été mises en place afin d'identifier les déterminants de l'AI mais aucune n'a donné de résultats probants à ce jour (Langridge et Baumann, 2008 ; Yang *et al.*, 2008). La conservation de la synténie pour les locus *S* et *Z* a été mise en évidence entre les espèces *Secale cereale*, *Phalaris coerulea* et *Lolium perenne* (Yang *et al.*, 2008). La cartographie de marqueurs dans la région des locus *S* et *Z* dans l'une de ces espèces est donc transposable aux autres (Langridge et Baumann, 2008). Le locus *S* a été cartographié chez *Lolium* et *Phalaris* dans une région de 0,1 cM qui correspond à une région de 2 Mb chez le riz, ce qui témoigne d'une importante suppression de la recombinaison dans cette région. Le locus *Z* a été cartographié dans une région de 1,5 cM mais qui correspond cette fois-ci à une région de 380 kb chez le riz. Avec un taux de recombinaison près de 1000 fois plus important que pour le locus *S*, le locus *Z* semble donc une meilleure cible dans le cadre d'une approche par clonage positionnel (Langridge et Baumann, 2008).

1.3 L'autoincompatibilité sporophytique homomorphique

L'AIS se distingue de l'AIG par des relations variables de dominance ou codominance qui peuvent exister entre les différents haplotypes S . En effet, le phénotype du grain de pollen est dépendant du génotype diploïde de la plante mère. Les 2 allèles du déterminant pollen portés par celle-ci peuvent être soit codominants, soit dominant et récessif. Il existe également des relations de dominance ou codominance entre les allèles du déterminant femelle, ces relations n'étant pas toujours les mêmes que pour le déterminant mâle. De ce fait, les phénotypes du pollen et du stigmate ne reflètent que partiellement le génotype des plantes mères. D'une part, ces relations de dominance augmentent le nombre de croisements compatibles au sein d'une population, d'autre part, elles permettent la génération d'individus homozygotes pour les allèles S récessifs (Figure 8). Ce mode d'autoincompatibilité est étudié au niveau moléculaire dans 3 familles : les Brassicaceae, les Convolvulaceae et les Asteraceae (Figure 2, Hiscock et Tabah, 2003).

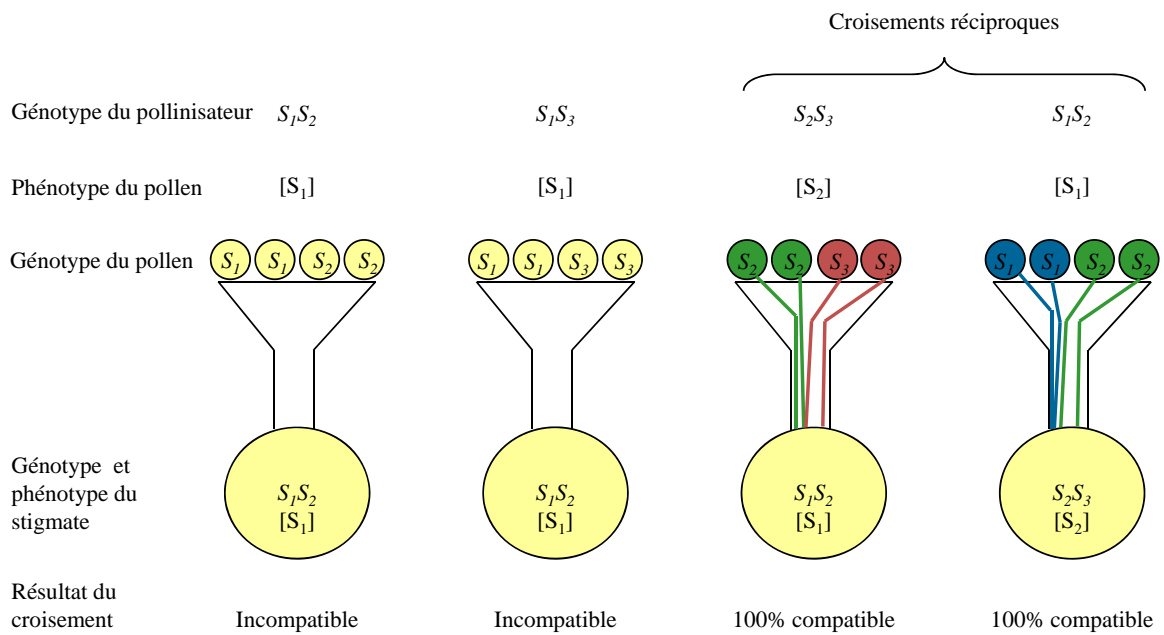


FIGURE 8 – Autoincompatibilité sporophytique. Le phénotype du pollen est déterminé par le génotype de la plante mère. Dans cet exemple, l'haplotype S_1 est dominant sur S_2 et S_3 et l'haplotype S_2 est dominant sur S_3 (d'après Silva et Goring, 2001).

1.3.1 Système sporophytique des Brassicaceae

La réaction d'incompatibilité a lieu à la surface du stigmate, la reconnaissance du pollen incompatible entraînant l'inhibition de l'hydratation du pollen ou bien l'arrêt rapide de la croissance du tube pollinique. Un seul locus polymorphe (30 haplotypes chez *Brassica campestris*) contrôle l'AI des Brassicaceae : le locus *S* (Silva et Goring, 2001 ; Takayama et Isogai, 2005).

Une approche biochimique similaire à celle menée chez les Solanaceae a permis d'identifier une première protéine comme étant le déterminant femelle de l'AI, la protéine S-locus-glycoprotéin (SLG), protéine abondante au niveau du stigmate (Nishio et Hinata, 1977 ; Watanabe *et al.*, 2008). Le gène *SRK* (S locus receptor kinase), codant une protéine appartenant à la famille des récepteurs kinases et ayant un niveau d'expression très bas, a été identifié grâce à sa similarité de séquence avec le gène *SLG* (Sherman-Broyles et Nasrallah, 2008 ; Silva et Goring, 2001 ; Watanabe *et al.*, 2008). Les protéines SLG et SRK sont exprimées dans le stigmate, principalement dans les papilles exposées à la surface. Les expériences de validation fonctionnelle par transformation ont montré que seule la protéine SRK détermine la spécificité du pistil. Le rôle de la protéine SLG est variable : il a été montré qu'elle facilite les événements de reconnaissance pour certains haplotypes (Takayama et Isogai, 2005). La protéine SRK est une protéine transmembranaire constituée d'une région extracellulaire (très similaire à SLG), appelée domaine S, d'une région transmembranaire et d'un domaine kinase sérine/thréonine intracellulaire (Silva et Goring, 2001).

Le déterminant mâle est une petite protéine (< 10 kDa) de l'enveloppe du pollen (pollen coat protein) appelée soit SCR (S-locus Cystein-Rich) soit SP11 (S-locus Protein 11), ayant été identifiée par 2 laboratoires différents grâce à l'étude génomique de la région du locus *S* (Schopfer *et al.*, 1999 ; Silva et Goring, 2001 ; Suzuki *et al.*, 1999 ; Takayama et Isogai, 2005). Elle présente une similarité avec les défensines (Sherman-Broyles et Nasrallah, 2008). Cette protéine est sécrétée par les cellules du tapis de l'anthere et est transférée à la surface des grains de pollen, ce qui explique la nature sporophytique de ce type d'AI (Hiscock et Tabah, 2003 ; Takayama et Isogai, 2005). SCR/SP11 est également exprimée par le gamétophyte (Hiscock et Tabah, 2003). La transformation d'individus par gain de fonction a permis de valider les gènes *SCR/SP11* et *SRK* comme étant les déterminants mâle et femelle de l'AI chez les Brassicaceae (Takayama et Isogai, 2005 ; Watanabe *et al.*, 2008).

1. Mécanismes d'autoincompatibilité

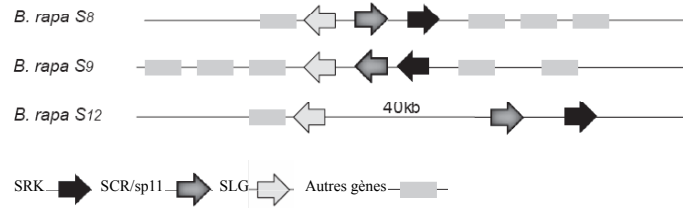


FIGURE 9 – Polymorphisme de structure au locus *S* chez *Brassica rapa* (d'après Sherman-Broyles et Nasrallah, 2008).

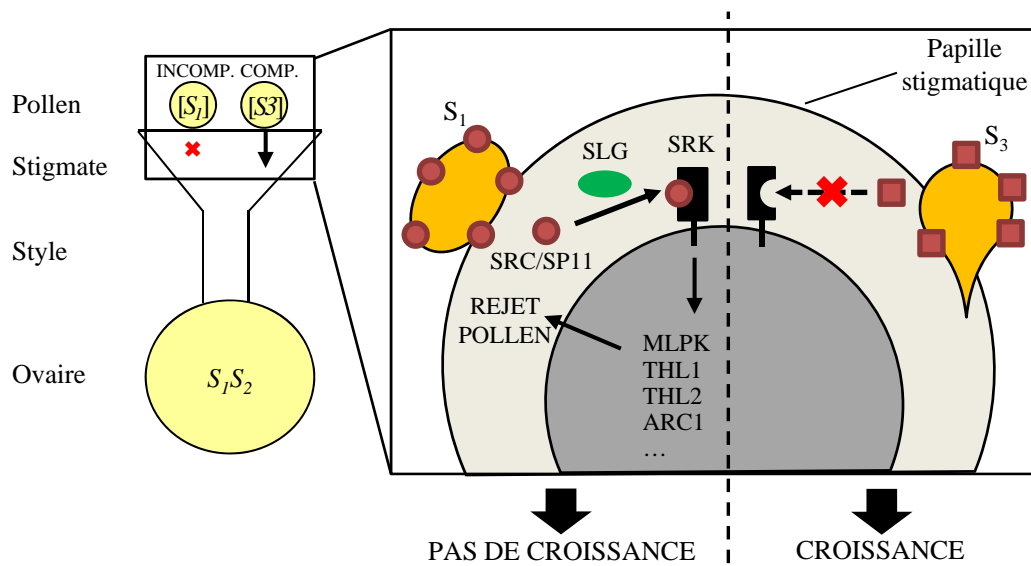


FIGURE 10 – Modèle moléculaire de la réponse autoincompatible chez les Brassicaceae avec S_1 dominant ou codominant avec S_2 (d'après Takayama et Isogai, 2005).

L'identification des déterminants mâle et femelle a ainsi mis en évidence une forte diversité de structure (taille des régions intergéniques, orientation des gènes) et de séquence au locus *S* d'un haplotype à l'autre (Figure 9, Sherman-Broyles et Nasrallah, 2008). Ainsi, une importante diversité allélique a été observée que ce soit pour le gène *SRK* (51% de diversité chez *Arabidopsis lyrata*) ou le gène *SCR/SP11* (35% d'identité chez *Brassica sp.*).

D'autres protéines du pistil ont été identifiées comme impliquées dans la cascade réactionnelle déclenchée par l'autoincompatibilité (Figure 10). MLPK (M locus protein kinase) est une protéine kinase, médiateur positif de la voie de signalisation de l'AI. THL1 et THL2 (thiorodoxin-h proteins) sont des régulateurs négatifs de l'autophosphorylation de SRK. Lorsque la protéine SCR/SP11 interagit de façon spécifique avec le récepteur

SRK, cela entraîne la phosphorylation de son domaine kinase. Celui-ci phosphoryle à son tour ARC1 (Armadillo-repeat-containing 1). Cette protéine ayant une activité ligase E3-ubiquitine, une des hypothèses envisagées est qu'elle déclencherait l'ubiquitination et la dégradation protéasomique des protéines stigmatiques impliquées dans la germination ou la croissance du tube pollinique. Si sa fonction n'est pas entièrement définie, il est certain qu'ARC1 est un maillon du mécanisme de rejet du pollen incompatible (voir pour revue Hiscock et Tabah, 2003 ; Takayama et Isogai, 2005).

1.3.2 Système sporophytique des Convolvulaceae

L'autoincompatibilité des Convolvulaceae est étudiée dans le genre *Ipomoea* (ipomée). Elle est caractérisée par l'absence de germination du grain de pollen à la surface du stigmate et est contrôlée par un locus multiallèlique (Hiscock et Tabah, 2003). Pour *Ipomoea trifida*, près de 50 allèles *S* différents ont été observés en populations naturelles (Kowyama *et al.*, 2000).

Une approche utilisant *SRK* comme gène candidat a montré que le polymorphisme de gènes présentant une forte homologie chez l'ipomée n'était pas associé aux haplotypes *S* étudiés, suggérant que les déterminants de l'AI étaient différents dans cette famille (Kowyama *et al.*, 1996). Une approche de clonage positionnel a été entreprise afin d'identifier les déterminants de l'autoincompatibilité au sein de cette espèce. Dans un premier temps, la cartographie fine de la région du locus *S* chez *I. trifida* a permis d'identifier des marqueurs moléculaires fortement liés ou coségrégeant avec le locus *S* (Tomita *et al.*, 2004a). Ces marqueurs ont ensuite été utilisés pour cribler différentes banques génomiques (BAC, λ , fosmid) construites à partir d'un individu homozygote au locus *S*. La marche sur le chromosome a permis d'identifier une région de 313 kb contenant le locus *S* (Tomita *et al.*, 2004b). Cette séquence a ensuite été comparée à celle d'un second haplotype et une région hautement divergente encadrée par des régions présentant au contraire une forte similarité a été mise en évidence (Rahman *et al.*, 2007a). C'est dans cette région divergente entre les haplotypes que 4 déterminants potentiels ont été identifiés, en fonction de leur diversité entre haplotypes et leur spécificité d'expression : *AB2*, *SE1*, *SE2* et *SEA*.

La protéine *AB2* est exprimée dans le tapis de l'anthere et présente des similarités avec des défensines. Les séquences protéiques déduites pour différents haplotypes présentent de 43 à 54% d'identité. Les protéines *SE1*, *SE2* et *SEA* sont de structures similaires mais ne présentent pas d'homologie avec des protéines connues. Les pourcentages d'identité

protéiques estimés entre haplotypes varient de 53 à 76% (Rahman *et al.*, 2007b). L'implication de ces gènes dans la réaction d'autoincompatibilité chez *I. trifida* doit encore être validée par transformation (Kakeda *et al.*, 2009).

1.3.3 Système sporophytique des Asteraceae

C'est chez les Asteraceae que le mécanisme d'autoincompatibilité sporophytique a été caractérisé pour la première fois (Gerstel, 1950 ; Hughes et Babcock, 1950), mais peu d'informations existent quant aux déterminants ou aux mécanismes moléculaires de l' AIS dans cette famille.

Chez le tournesol, autoincompatible à l'état sauvage, l'étude de l'autofertilité¹ a été réalisée dans un contexte d'analyse génétique du syndrome de domestication et des processus de post-domestication (Burke *et al.*, 2002 ; Gandhi *et al.*, 2005 ; Wills et Burke, 2007). Pour chacune des études réalisées, la carte génétique a été construite à partir de descendants issus de croisements entre des individus sélectionnés pour leur autofertilité et des individus sauvages autoincompatibles. Ces différents travaux ont mis en évidence de 1 à 4 QTL associés à la production de graines par autofécondation. Un QTL à effet fort a été retrouvé de façon récurrente sur le groupe de liaison 17 et a été associé au locus *S* (Gandhi *et al.*, 2005). Cependant, ce QTL, identifié à partir de la capacité des individus à s'autoféconder, pourrait également correspondre à un locus distinct du locus *S* et responsable de l'autofertilité.

Les études les plus approfondies ont été réalisées chez l'espèce *Senecio squalidus* pour laquelle on a montré que l'arrêt de la croissance du tube pollinique avait lieu à la surface du stigmate (Hiscock *et al.*, 2002). Un seul locus *S* multiallèlique est impliqué dans la réaction et une dizaine d'haplotypes ont été identifiés chez *S. squalidus* (Hiscock *et al.*, 2003).

Deux types d'approches ont été menées afin d'identifier les déterminants génétiques de l' AIS : une approche gène candidat à partir du gène *SRK* des Brassicaceae et une approche biochimique. Trois gènes de type *SRK* ont été identifiés chez *S. squalidus*. Cependant, ceux-ci s'expriment à divers degrés dans les différents tissus de la plante et aucun ne présente du polymorphisme en corrélation avec le phénotype d'autoincompati-

1. Le terme autofertile désigne ici des individus capables de s'autoféconder, mais pour lesquels le déterminisme du caractère n'est pas connu (gène(s) agissant(s) en amont ou en aval du locus *S* ou rupture de fonction au locus *S*). Le terme autocompatible sera employé uniquement lorsque la capacité d'autofécondation est effectivement due à une rupture de fonction du locus *S*.

bilité. Ces résultats suggèrent que chez *S. squalidus*, il ne s'agit pas du même mécanisme d'AIS que chez les Brassicaceae (Hiscock *et al.*, 2003 ; Tabah *et al.*, 2004). La seconde approche, un crible des protéines stigmatiques, a été menée à partir de la technique IEF pour des individus homozygotes S_1S_1 et S_2S_2 , ceci afin d'identifier les protéines exprimées spécifiquement chez chacun de ces 2 haplotypes. Deux protéines, codées par 2 allèles (SSP_1 et SSP_2) d'un même gène SSP , ont été identifiées comme étant *a priori* liées aux haplotypes S_1 et S_2 (Hiscock *et al.*, 2003). Une analyse menée sur un plus grand nombre d'individus a montré que ces allèles étaient également présents en l'absence des haplotypes S_1 et S_2 : SSP ne code pas le déterminant femelle de l'autoincompatibilité chez *S. squalidus*, mais est un gène exprimé spécifiquement dans les cellules stigmatiques, codant une peroxydase (McInnis *et al.*, 2005).

Une étude récente relate la construction et l'analyse d'une banque d'ADNc soustractive enrichie en transcrits extraits de style de *S. squalidus* (Allen *et al.*, 2010). Les auteurs ont identifié des ADNc correspondant à des gènes conservés et exprimés à la fois dans les pistils d'espèces dicotylédones et monocotylédones. Ces gènes seraient potentiellement impliqués dans les mécanismes généraux d'interaction pollen-pistil des Angiospermes. Ils ont également identifié des transcrits non représentés dans les banques d'ADNc de stigmates disponibles, qui pourraient être impliqués dans le mécanisme d'AIS propre aux Asteraceae.

1.4 Divergences et similitudes des systèmes d'AI homomorphique

Au final, l'autoincompatibilité se présente sous divers mécanismes dont le déterminisme génétique est différent. Ainsi, des systèmes ligand-récepteur ont été identifiés chez les Papaveraceae et les Brassicaceae. Dans la première famille le ligand est sécrété à la surface du stigmate. Dans la seconde, il est sécrété par les cellules du tapis de l'anthere et déposé à la surface du pollen. D'autre part, un mécanisme différent, faisant intervenir une protéine ayant une activité ribonucléase et une protéine de type F-box, a été identifié chez les Solanaceae, Rosaceae et Plantaginaceae.

De manière générale, on a constaté que les espèces autoincompatibles homomorphiques d'une même famille possèdent le même mécanisme, soit AIS, soit AIG. A ce jour, une seule famille regroupant ces 2 types d'AI a été identifiée : la famille des Polemoniaceae (AIS pour *Linanthus parviflorus* et AIG pour *Phlox drummondii*, Hiscock et Tabah,

2003). Enfin, chez certaines espèces ayant un mécanisme d'AIS, on observe également un système gamétophytique dit 'cryptique' (Allen et Hiscock, 2008 ; Hiscock, 2000b). Malgré les différences qui existent entre les mécanismes dont le déterminisme a été identifié, un certain nombre de caractéristiques génétiques et génomiques communes ont été mises en évidence :

- une importante diversité haplotypique,
- une faible identité des séquences d'acides aminés des déterminants mâles et femelles entre les différents haplotypes,
- une localisation dans une région chromosomique fortement divergente entre haplotypes (diversité de séquence et de structure),
- une localisation dans une région faiblement recombinante du génome,
- une expression préférentielle du déterminant mâle dans le pollen et du déterminant femelle dans le stigmate.

Chez les espèces AI appartenant à une famille pour laquelle le mécanisme n'a pas encore été identifié, on s'attend à trouver les déterminants dans une région génomique répondant à ces différentes caractéristiques. On constate également que les différents déterminants des mécanismes peuvent : 1) appartenir à des familles de gènes codant des protéines présentes chez différentes espèces, de fonctions connues (par exemple SRK ou S-RNase) ou inconnues (par exemple PrsS) ou, 2) n'avoir aucun équivalent structurel chez d'autres espèces (par exemple PrpS).

Plusieurs types d'approches (biochimique, gène candidat ou clonage positionnel) ont été utilisées jusqu'à présent afin d'identifier les déterminants de l'autoincompatibilité. D'un point de vue historique, l'approche biochimique est la première à avoir été utilisée. Chez les espèces pour lesquelles on ne dispose d'aucune information génétique, c'est effectivement une stratégie envisageable, mais cela présuppose que la protéine soit abondante dans le pistil ou au niveau du pollen, ce qui n'est pas toujours le cas. Une autre option est l'approche gène candidat, mais cette voie n'apporte pas nécessairement de réponse étant donné la diversité de mécanismes déjà observés. Enfin, si on dispose des outils nécessaires (cartes génétiques, descendances, ressources génomiques), on peut envisager une approche par clonage positionnel.

2 L'autoincompatibilité sporophytique chez les Asteraceae : la chicorée comme modèle ?

Avec 23 000 espèces décrites dans 1 300 genres, la famille des Asteraceae est la plus importante des Angiospermes (Funk *et al.*, 2009). En effet, ce groupe monophylétique, caractérisé par la présence de capitules (Figure 11), représente environ 10% des plantes à fleurs (Funk *et al.*, 2005). Cette famille de dicotylédones se caractérise également par une importance économique notable puisque de nombreuses espèces cultivées y sont recensées. Dans le domaine de l'agro-alimentaire, on trouve par exemple *Helianthus annuus* (tournesol), *Lactuca sativa* (laitue), *Cichorium intybus* L. (chicorée industrielle, endive²), *Cichorium endivia* (chicorées frisée et scarole), *Carthamus tinctorius* (carthame), *Cynara cardunculus* (artichaut), *Artemisia dracunculus* (estragon) ou bien encore *Tragopogon porrifolius* (salsifis). Dans le domaine de l'horticulture, on peut notamment citer les espèces du genre *Dahlia sp.* ou du genre *Aster sp.*, ou encore les chrysanthèmes qui regroupent des espèces appartenant à plusieurs genres. Un certain nombre d'espèces sont également classées dans la catégorie des 'mauvaises herbes', comme par exemple les pissenlits (*Taraxacum sp.*), les chardons (*Carduus sp.*) et les séneçons (*Senecio sp.*).

Selon Ferrer et Good-Avila (2007), plus de 60% des Asteraceae seraient autoincompatibles. Toutefois, malgré l'importance de cette famille, l'AI reste peu documentée au niveau moléculaire et le ou les mécanismes responsables n'ont toujours pas été identifiés. Pourtant, la notion d'incompatibilité a été présentée dès 1917 chez *C. intybus* (Stout) et, en 1950, l'AI sporophytique a été identifiée chez *Crepis foetida* (Hughes et Babcock, 1950) et *Parthenium argentatum* (Gerstel, 1950). Des études récentes ont été réalisées chez *S. squalidus* (*cf.* paragraphe p.33), mais les approches biochimiques et gènes candidats n'ont pas permis l'identification de déterminants de l'AI.

La chicorée appartient à la tribu des Lactuceae (comme les laitues ou les pissenlits) et au genre *Cichorium* ($2n = 18$) qui est composé de de 6 espèces distinctes : *C. intybus*, *C. endivia*, *C. pumilum*, *C. calvum*, *C. spinosum* et *C. bottae* (Kiers *et al.*, 2000). Selon Kiers (2000), 3 de ces espèces sont pérennes et autoincompatibles (*C. intybus*, *C. spinosum* et *C. bottae*), les autres étant annuelles et autogames. Le système de re-

2. Pour éviter toute confusion entre les chicorées endives appartenant à l'espèce *C. intybus* et les chicorées à feuilles de l'espèce *C. endivia*, le terme chicorée sera utilisé par la suite pour désigner *C. intybus*. L'appellation *C. endivia* sera réservée à cette seconde espèce.

2. L'autoincompatibilité sporophytique chez les Asteraceae : la chicorée comme modèle ?

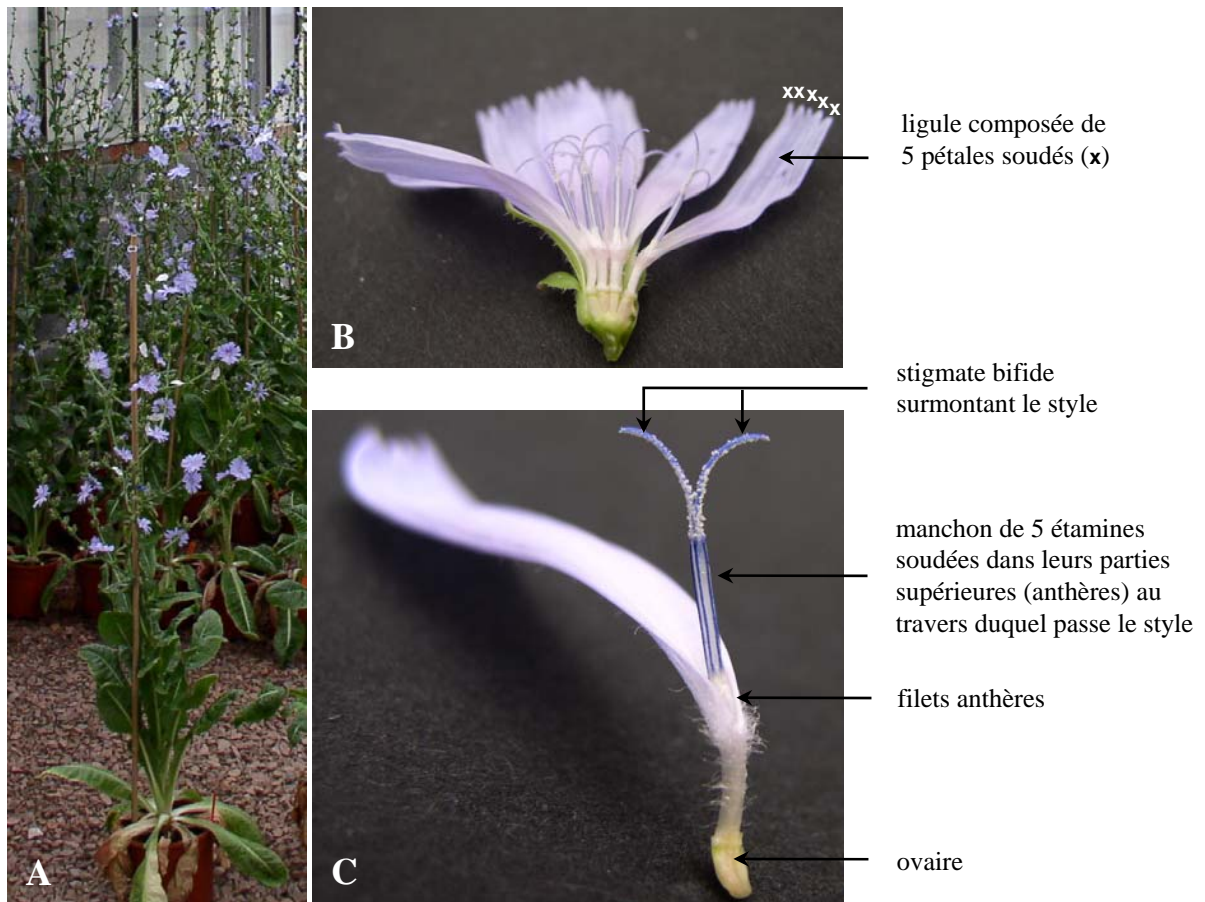


FIGURE 11 – Chicorée (*Cichorium intybus*). A) plante entière (échelle : 1/13), B) coupe longitudinale d'un capitule (échelle : 1/1). C) Détail d'une fleur isolée (échelle : 4). Photos : Theo Hendriks.

production de la chicorée (allogame préférentielle) et de *C. endivia* (autogame), seules espèces cultivées du genre, a été décrit à plusieurs reprises (Castaño *et al.*, 1997 ; Coppens d'Eeckenbrugge, 1990 ; Coppens d'Eeckenbrugge et Evrard, 1988 ; Eenink, 1981a,b, 1982, 1984 ; Kiers, 2000 ; Lucchin *et al.*, 2008 ; Varotto *et al.*, 1995). Celui des 4 autres espèces a, quant à lui, été décrit à une seule occasion, à partir des résultats obtenus pour 15 autofécondations contrôlées pour chaque espèce (Kiers, 2000).

Alors qu'à l'état sauvage la chicorée est pérenne, les variétés utilisées en agronomie sont cultivées de façon bisannuelle. Durant la première année après la germination, la plante se présente au stade végétatif. Après une période de vernalisation, on observe l'apparition d'une hampe florale pouvant atteindre jusqu'à 2 mètres de haut la seconde année. Les capitules bleu-violet sont constitués de 15 à 25 fleurs. La floraison de l'ensemble des

Introduction

fleurs d'un capitule est synchrone et se déroule en quelques heures (Figure 11). Cette synchronisation est rare dans la famille des Asteraceae, car pour la grande majorité des espèces, la floraison de l'ensemble des fleurs d'un capitule s'étend sur plusieurs jours (6-8 jours pour un capitule de tournesol cultivé, communication personnelle M-C Quillet). Une plante produit plusieurs centaines de capitules, durant la période de floraison qui s'étend sur 4 à 6 semaines. Ces caractéristiques propres à la chicorée permettent donc 1) de définir précisément le rapport nombre de graines/nombre d'ovules fécondés (seed-set) pour chaque capitule pollinisé, et 2) de réaliser de nombreuses répétitions de croisements sur une même plante.

La chicorée cultivée comprend de nombreuses variétés ou cultigrupes différents pouvant être groupés en 3 catégories. La chicorée industrielle est cultivée pour ses racines. Celles-ci sont utilisées pour la production de substitut au café ou d'inuline et de fructanes, molécules d'intérêt pour l'industrie alimentaire et médicale. La chicorée est également un produit de l'alimentation, consommée en tant que salade ou légume (chicorée endive aussi appelée witloof ou chicon, pain de sucre, radicchio...). Elle est aussi utilisée en tant que plante fourragère dans certaines parties du monde (Cadalen *et al.*, 2010).

Chez la chicorée, la germination des grains de pollen peut être observée 3 minutes seulement après la pollinisation, lorsque le croisement est compatible. Dans le cas des croisements incompatibles, la réaction de l'AI a lieu au niveau du stigmate. La majorité des grains de pollen ne germe pas et, dans les rares cas où cela se produit, la croissance du tube pollinique s'arrête rapidement (Coppens d'Eeckenbrugge, 1990). Les résultats de croisements pour les variétés witloof (Eenink, 1981a) et radicchio (Varotto *et al.*, 1995) ont montré que l'autofécondation est supprimée du fait d'un système d'autoincompatibilité sporophytique, contrôlé par un seul locus *S* multiallèlique. Le degré d'autoincompatibilité peut être variable et des individus pseudocompatibles³ ou autofertiles ont été identifiés (Castaño *et al.*, 1997 ; Coppens d'Eeckenbrugge *et al.*, 1987). Dans le cultigrupe witloof, des génotypes autofertiles ont été sélectionnés afin de produire des lignées et générer des variétés hybrides F1 homogènes (Doré et Varoquaux, 2006). Dans les autres cultigrupes, la pression de sélection artificielle concernant ce caractère est moins importante et l'AI est conservée (Lucchin *et al.*, 2008).

La chicorée étant une plante d'intérêt agronomique, un certain nombre d'études ont

3. Les individus pseudocompatibles produisent plus de graines par autofécondation que individus autoincompatibles, mais le nombre de graines reste inférieur à celui obtenu pour des croisements compatibles (de Nettancourt, 1977 ; Levin, 1996).

2. L'autoincompatibilité sporophytique chez les Asteraceae : la chicorée comme modèle ?

permis la genèse de ressources génétiques et génomiques significatives. Ainsi, 2 cartes génétiques ont été construites à partir de marqueurs moléculaires dominants (RAPD, AFLP) pour les variétés raddichio (De Simone *et al.*, 1997) et witloof (Van Stallen *et al.*, 2003). Plus récemment, la première carte génétique consensus de chicorée a été réalisée au sein de notre laboratoire, à partir de marqueurs codominants, pour des chicorées industrielle et witloof (Cadalen *et al.*, 2010). De plus, des banques d'EST ont été construites dans le cadre de l'étude de l'embryogénèse somatique chez la chicorée (Legrand *et al.*, 2007). Elles sont venues compléter les ressources accessibles sur le site NCBI dbEST (Dauchot *et al.*, 2009 ; Legrand *et al.*, 2007) et sur le site du CGPDB, avec au total, plus de 50 000 EST de chicorée. Depuis quelques mois, nous avons également accès à deux nouvelles collection d'EST, générées par pyroséquençage 454. La première est issue d'ARN de stigmates et a été produite par le laboratoire. La seconde provient d'ARN de racines et a été gracieusement mise à notre disposition par Jules Beekwilder de l'Université de Wageningen (Cankar *et al.*, 2011). A cela s'ajoute les ressources en EST obtenues à partir d'espèces appartenant à la même tribu que la chicorée, notamment *C. endivia*, mais également les pissenlits et les laitues, disponibles sur le site du CGPD.

Un frein à l'étude de l'autoincompatibilité chez les Asteraceae est sans aucun doute le faible intérêt de ce caractère pour l'amélioration génétique des 2 espèces les plus cultivées que compte la famille, à savoir la laitue et le tournesol. En effet, la laitue cultivée, ainsi que la quasi-totalité des espèces apparentées, utilisées comme ressources génétiques, sont des espèces autogames. Chez le tournesol, les génotypes, employés dans les programmes actuels d'amélioration des plantes, sont autofertiles. A cela s'ajoute le manque de ressources génomiques qui existent pour cette famille, en comparaison de son importance. Le séquençage du génome du tournesol a récemment débuté (Plant & Animal Genomes XVIII Conference⁴), mais aucun génome d'Asteraceae n'est séquencé à ce jour. Ainsi, si l'étude de l'AI chez les Asteraceae a été explorée chez *S. squalidus*, plus ponctuellement chez le tournesol et quelques autres espèces, aucune approche génomique globale ne semble avoir été développée afin d'identifier les déterminants de l'AI dans cette famille.

Dans ce contexte, la chicorée est apparue comme un modèle pertinent pour entreprendre l'identification des déterminants de l'autoincompatibilité chez les Asteraceae,

4. http://www.intl-pag.org/18/abstracts/W22_PAGXVIII_170.html

Introduction

et ce pour plusieurs raisons.

Cette culture est importante dans différents pays européens et la France est l'un des trois premiers pays producteurs de cette espèce avec la Belgique et les Pays-Bas. La connaissance des déterminants de l'AI permettrait sans aucun doute l'optimisation des schémas de sélection, dans les cultigroupes pour lesquels l'AI n'a pas été contre sélectionnée.

Il existe une importante variabilité du système de reproduction, présente dans les compartiments cultivés et sauvages de l'espèce *C. intybus* et au-delà au sein du genre *Cichorium*. Ce groupe d'espèces représente aussi un modèle d'intérêt pour l'étude des processus associés à la sélection naturelle/artificielle aboutissant à l'autofertilité.

D'un point de vue pratique, le grand nombre de capitules produits par plante et la synchronisation de floraison des fleurs d'un capitule permettent une grande précision de l'estimation du profil de compatibilité de chaque individu à analyser. De plus, pour la construction de la carte génétique consensus, une importante population en ségrégation a été générée à partir de 2 individus présentant un système d'autoincompatibilité strict (*cf.* chapitre 2). Cette population a donc été propice à la mise en place d'un système de phénotypage du locus *S*. Celui-ci a ainsi pu être cartographié à l'extrémité du groupe de liaison 2.

L'objectif de cette thèse était d'initier le clonage positionnel du locus *S* chez la chicorée. Cette stratégie consiste à cloner les gènes majeurs d'intérêt de par leur localisation génétique et physique sur le génome. Cette approche est sans aucun doute plus longue à mettre en œuvre que les approches biochimique ou gènes candidats, mais permettra d'aboutir à l'identification du ou des gènes recherchés. La vitesse de progression de cette démarche dépend toutefois de la distance physique à couvrir et donc de la relation distance physique/distance génétique associée à la région d'intérêt. Il était donc nécessaire, dans un premier temps, de construire une carte génétique fine du locus *S*, puis une carte physique de cette région. La construction d'une carte moyenne densité avait été initiée préalablement, permettant de positionner le locus *S* sur la carte consensus et d'établir en partie les relations de dominance entre haplotypes *S*. Ces résultats ont été complétés au début de la thèse et font l'objet du chapitre 2. Ils constituent les bases sur lesquels les travaux présentés ensuite ont été réalisés.

Dans le chapitre suivant cette introduction, nous présenterons le matériel végétale et les différentes techniques utilisées au cours de la thèse (chapitre 1). Puis, dans la partie I,

2. *L'autoincompatibilité sporophytique chez les Asteraceae : la chicorée comme modèle ?*

les différentes étapes menant à l'élaboration de la carte génétique fine du locus *S* seront abordées (chapitres 2, 3 et 4). La construction des banques BAC et l'établissement de la carte physique seront présentés dans la partie II (chapitres 6 et 7). Chacune de ces parties fera l'objet d'une discussion indépendante. Enfin, la conclusion aura pour but de faire la synthèse des résultats et de présenter les perspectives s'ouvrant à l'issue de ce travail.

Introduction

Chapitre 1

Matériels et méthodes

Chapitre 1

Matériel et méthodes

1.1 Matériel végétal

1.1.1 Présentation des populations utilisées

Huit populations différentes ont été utilisées dans le cadre du clonage positionnel du locus S chez la chicorée (Figure 1.2, Tableau 1.1). Chacune intervient dans une ou plusieurs des étapes successives de la démarche entreprise (Figure 1.1).

La population B a été obtenue à partir du croisement K28 x K59. K28 et K59 sont deux individus issus d'une population hongroise de *Cichorium intybus* var *sativum* appelée Koospol. K59 est un individu hermaphrodite dont le génotype au locus S est S_1S_2 . K28 est un individu mâle-stérile dont le génotype au locus S est S_3S_4 . La population B se divise en deux sous-populations dont les productions de graines ont eu lieu à des dates différentes. La population B_a est constituée de 389 individus et correspond aux lots de graines produits en 2001 et 2004. En 2006 et 2007, deux nouveaux lots de graines ont été produits et constituent la population B_b qui comprend un total de 253 individus. La population B_a a été utilisée lors de la construction de la carte génétique de la chicorée (Cadalen et al., 2010) et des cartes moyenne densité (*cf.* chapitre 2) et haute densité (*cf.* chapitre 3). L'ensemble de la population B a été utilisée lors de la construction de la carte haute résolution (*cf.* chapitre 4) du locus S .

La population C est constituée de 158 individus. Le lot C_a a été obtenu par autotécondation de K59 et le lot C_b par autotécondation d'individus du lot C_a . Cette population a été utilisée à la fois pour la carte haute densité du locus S (*cf.* chapitre 4) et la construction des banques BAC de chicorée (*cf.* chapitre 6).

Les populations D, E, F et G sont issues de croisements entre individus B. La population D (317 individus) a été obtenue à partir du croisement entre B251 (S_2S_4) et B266 (S_1S_3). Les populations E (213 individus), F (291 individus) et G (283 individus) sont issues du croisement de B272 (S_2S_4) avec les individus B240 (S_1S_3), B253 (S_1S_3) et B337 (S_1S_3). La population H (506 individus) est issue du croisement de 17 individus B dont le génotype au locus S est S_2S_3 , K59 ayant été utilisé comme parent mâle. Ces 17 individus ont en commun un génotype identique pour les marqueurs encadrant le locus S et constituent un groupe appelé B_H .

La population M (572 individus) a été obtenue à partir des croisements de K59 avec 7 individus B, utilisés comme parent mâle, et dont le génotype au locus S est S_2S_3 . Ces individus B ont un génotype identique pour les marqueurs encadrant le locus S et constituent un groupe appelé B_M .

Les populations D, E, F, G, H et M ont été utilisées afin de construire la carte génétique haute résolution (*cf.* chapitre 4).

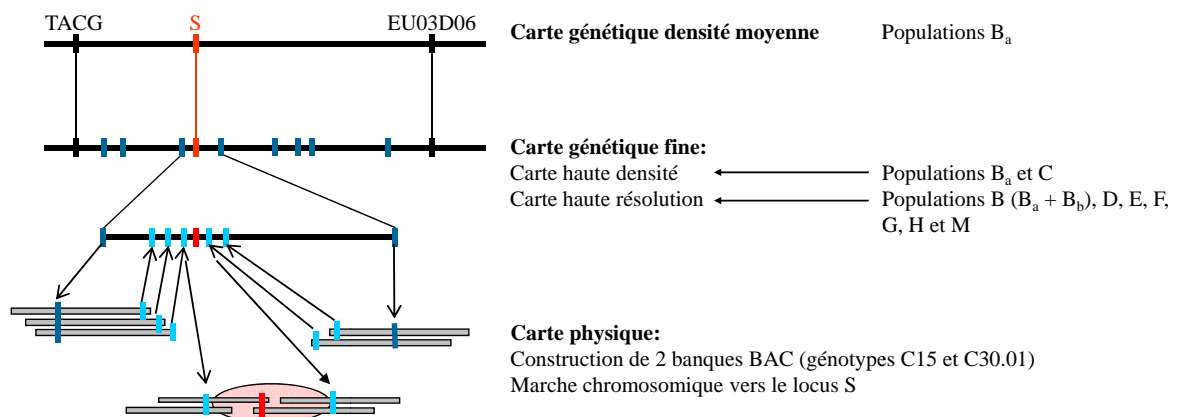


FIGURE 1.1 – Différentes étapes du clonage positionnel et matériel végétal utilisé pour chacune d’entre elles.

1.1.2 Culture *in vitro*

La culture *in vitro* permet de maintenir du matériel grâce à un repiquage régulier et de le multiplier par embryogénèse somatique. Ce mode de culture a été utilisé pour cloner les individus K28 et K59 ce qui a permis de réaliser les croisements sur plusieurs années pour produire les différents lots de la population B. La population C a également été maintenue et multipliée en culture *in vitro*. Toutes les manipulations nécessaires à la

TABLEAU 1.1 – Différentes populations utilisées dans le cadre du clonage positionnel du locus *S* chez la chicorée. La population B a été obtenue à partir du croisement K28 x K59, la population C correspond a des individus issus d'autofécondations de K59, les populations D, E, F, G a des croisements entre individus de la population B et les populations H et M a des croisements entre K59 et des individus de la population B.

Nom de la population	Parent femelle	Parent mâle	Effectif	Détails	
B (B _a + B _b)	K28	K59	642	B _a : n = 389 (lot de 2001 : n = 174 ; lot de 2004 : n = 215) B _b : n = 253 (lot de 2006 : n = 135 ; lot de 2007 : n = 118)	
C	C _a	K59	K59	158	C _a : n = 36
	C _b	C _a	C _a		C _b : n = 122
D	B251	B266	317		
E	B272	B240	213		
F	B272	B253	291		
G	B272	B337	283		
H	Groupe B _H	K59	506	Groupe B _H : B431, B602, B608, B610, B617, B623, B631, B644, B653, B657, B661, B663, B665, B667, B668, B679, B680	
M	K59	Groupe B _M	572	Groupe B _M : B602, B623, B631, B644, B667, B668, B679	

culture *in vitro* ont été réalisées en milieu stérile.

1.1.2.1 Germination et maintien du matériel végétal

Pour faire germer du matériel végétal *in vitro*, plusieurs étapes ont été nécessaires (Figure 1.3). Tout d'abord, les graines ont été aseptisées. Pour cela, elles ont été placées 10 minutes dans une solution de Domestos à 12% puis rincées deux fois 5 minutes et une fois 10 minutes. Les graines ont ensuite été disposées sur du milieu H5 solide (Annexe A.1) dans des boîtes de pétri et placées à l'obscurité à température ambiante. Cinq jours après la mise en culture, les graines germées ont été disposées individuellement dans des tubes à essais contenant du milieu H5 solide puis placées en chambre de culture avec une photopériode de 16h de lumière/8h d'obscurité. Les plantules ont ensuite été repiquées toutes les 6 à 8 semaines pour être maintenues en culture *in vitro* : la majorité des feuilles et les racines sous le collet des plantules ont été coupées et celles-ci ont été

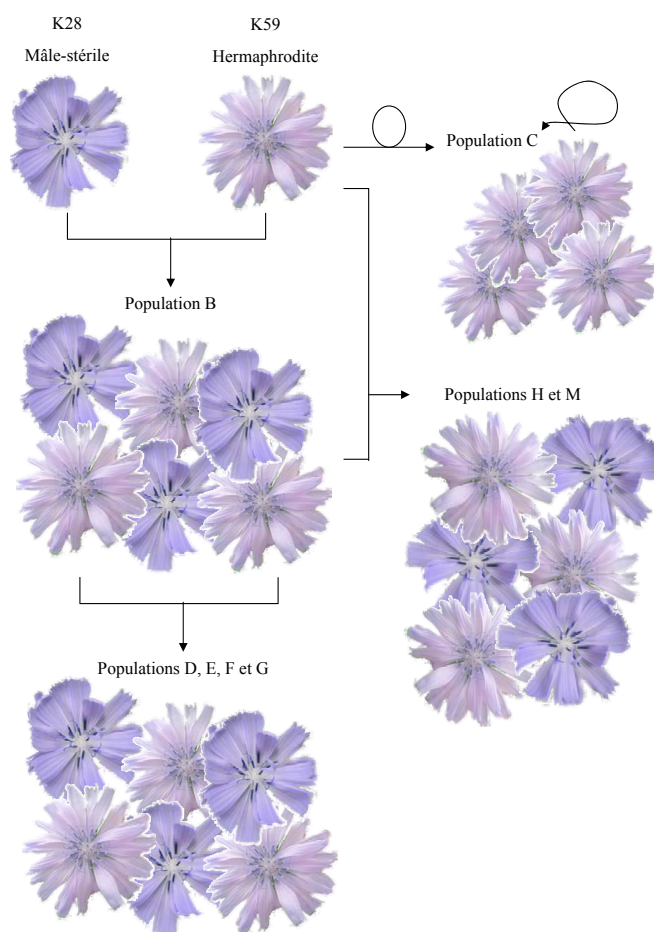


FIGURE 1.2 – Généalogie des différentes populations de chicorée utilisées.

repiquées dans un nouveau tube contenant du milieu H5 solide. Les plantules nécessaires aux croisements ou au phénotypage ont également été acclimatées en serre (Figure 1.3).

1.1.2.2 Multiplication du matériel végétal par embryogénèse somatique

La totalité des racines coupées d'une plante a été découpée en petits fragments d'1 cm environ. Ces fragments ont été placés dans un milieu d'induction M17S20 (Annexe A.1) à 35°C, à l'obscurité, sous agitation constante (80 rpm) pendant 7 jours (Figure 1.3). Les explants ont ensuite été placés dans un milieu H5 liquide (Annexe A.1), toujours sous agitation et ont été soumis à une alternance jour/nuit de 16h/8h. Après 30 jours, les embryons et les bourgeons ont été transférés sur milieu H5 solide en boîte de pétri. Lorsque les embryons ont été suffisamment développés, ils ont été repiqués dans des

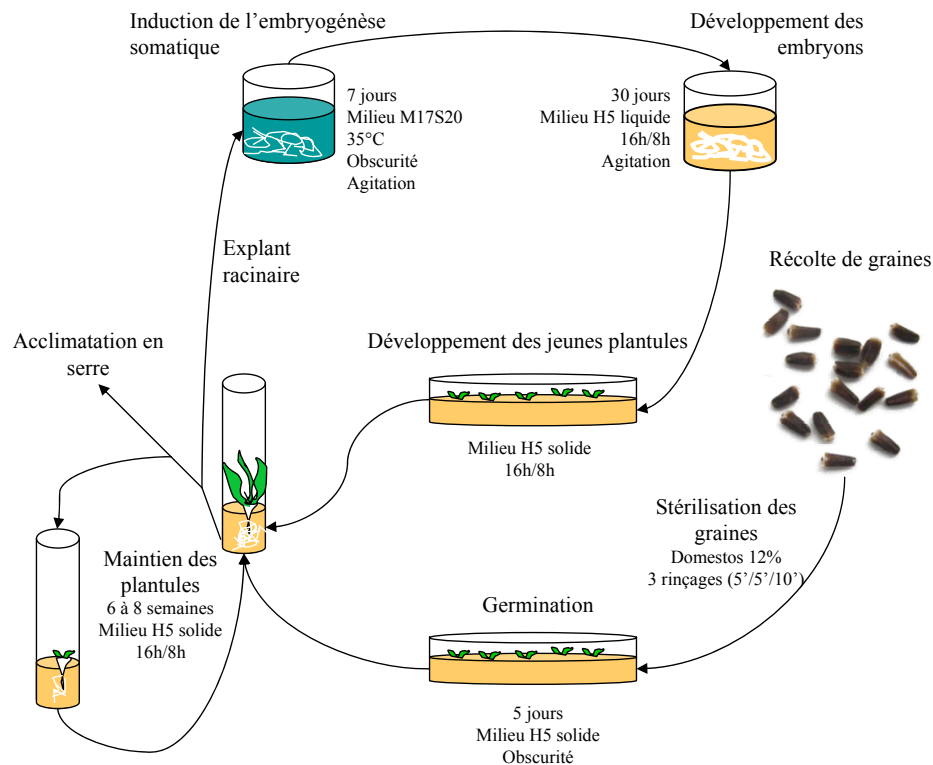


FIGURE 1.3 – Culture *in vitro*. Représentation des différentes étapes nécessaires au maintien des plantules et à leur multiplication par embryogénèse somatique (d'après Legrand, 2006).

tubes à essais contenant du milieu H5 solide.

1.1.3 Culture en serre

Le matériel végétal cultivé en serre peut avoir deux origines, soit il est issu de germinations qui ont eu lieu en milieu non stérile, soit il s'agit de plantules cultivées *in vitro* qui ont ensuite été transférées en serre. La culture en serre a permis d'obtenir des plantes atteignant le stade végétatif (Figure 1.4) puis le stade de montaison et de floraison après vernalisation.

1.1.3.1 Germination et maintien du matériel végétal

La germination des graines a été réalisée en chambre de culture (24°C, 16h lumière/8h obscurité). Les graines ont été placées sur du papier Wattman humide dans des boîtes de pétri, chaque boîte de pétri contenant les graines d'un seul capitule. Une fois les cotylédons et la racine bien développés (après une semaine environ), les plantules ont été



FIGURE 1.4 – Culture en serre. Plantules repiquées en plaques.

repiquées en terre dans des plaques de 20 pots de 7 cm de coté chacun. Dans le cas de plantules cultivées *in vitro*, celles-ci ont été extraites du milieu H5, leur racines nettoyées puis elles ont été repiquées en terre.

Ces plantules ont ensuite été placées en serre (environ 20°C avec un cycle jour/nuit de 16h/8h). Durant les 10 premiers jours, elles ont été conservées dans un milieu humide, soit dans des mini-serres pour les petits effectifs, soit sous un film plastique transparent. Elles ont ensuite été laissées à l'air libre (Figure 1.4). Le rempotage en pots individuels a eu lieu au bout de 1 à 3 mois . Au delà de 1 mois, une élimination régulière des feuilles sénescentes des plantes cultivées en plaque a eu lieu, avec ajout d'engrais si nécessaire.

1.1.3.2 Vernalisation

Pour obtenir la montaison et la floraison des individus, une étape de vernalisation est nécessaire. Pour cela, les plantes ont été placées pendant trois mois dans une fosse maintenue à une température de 4 à 6°C avant d'être de nouveau installées en serre.

1.2 Outils de biologie moléculaire

1.2.1 Echantillons d'ADN

1.2.1.1 Extraction d'ADN

L'extraction d'ADN a été réalisée soit sur du matériel frais, soit sur du matériel sec. Pour obtenir du matériel sec, les feuilles ont été placées 24h en étuve à 55°C. Entre

10 et 15 mg de matériel sec ou 100 mg de matériel frais ont été utilisés par individu. Les extractions ont été réalisées soit sur colonne avec le kit NucleoSpin plant II, soit en plaque avec le kit NucleoSpin 96 plant II (Macherey-Nagel). Dans les deux cas, le tampon de lyse PL1 a été utilisée et l'ADN a été élué dans 2 fois 100 μL de tampon d'élué.

1.2.1.2 Utilisation des échantillons individuels

Les échantillons d'ADN ont été utilisés afin de réaliser des réactions PCR (*cf.* paragraphe 1.2.2). Dans le cas des extractions en colonne, la concentration d'ADN a été mesurée au spectrophotomètre à 260 nm et les échantillons dilués à 1,5 ng/ μL . Les échantillons obtenus par extraction en plaque ont été dilués au 1/20^{ème}.

1.2.1.3 Utilisation des échantillons en 'bulks'

Les mélanges d'ADN (appelés bulks ou pools) utilisés pour la construction de la carte haute densité (*cf.* chapitre 3) ont été constitués de l'ADN de 8 à 15 individus présents en concentrations égales au sein de chaque pool. Ces bulks ayant été utilisés pour réaliser des analyses AFLP, chacun contenait au total 500 ng d'ADN dans un volume final de 20 μL .

1.2.2 Marqueurs moléculaires spécifiques de séquence

Les marqueurs spécifiques de séquence ou STS (Sequence-Tagged Site) sont des marqueurs moléculaires codominants définis à partir de copies uniques de séquences d'ADN. Parmi les différents types de marqueurs que désigne l'appellation STS, trois ont été utilisés dans le cadre du clonage positionnel du locus *S*, que ce soit pour la construction de la carte haute densité, haute résolution ou la marche chromosomique : des microsatellites (SSR, Simple Sequence Repeat), des marqueurs dérivés de marqueurs AFLP (SCAR, Sequence Characterized Amplified Region, *cf.* paragraphe 1.2.4) et des marqueurs BES (BAC End Sequences) définis à partir des séquences des extrémités des inserts des clones BAC.

1.2.2.1 Définition des couples d'amorce

Les couples d'amorces correspondants aux marqueurs SSR ont préalablement été définis dans le cadre de la carte consensus de la chicorée (Cadalen et al. 2010). Les amorces des marqueurs SCAR et BES ont été définies grâce au logiciel Primer3 avec une taille comprise entre 18 et 27 bases et une température d'hybridation (T_m) comprise entre 57 et 60°C (*cf.* annexe A.2).

1.2.2.2 Les réactions PCR

L'amplification des marqueurs STS par PCR a été réalisée dans un volume de 15 μL (Tableau 1.2). Un programme de type 'touchdown' (TD) a été utilisé, les conditions d'amplification étant les suivantes :

- 1ère étape :
Dénaturation initiale : 4 minutes à 94°C
- 2ème étape, 5 cycles :
Dénaturation : 30 secondes à 94°C
Hybridation : 40 secondes à T_m (-1°C à chaque cycle)
Elongation : 45 secondes à 72°C
- 3ème étape, 25 à 35 cycles selon le couple d'amorce utilisé :
Dénaturation : 30 secondes à 94°C
Hybridation : 40 secondes à T_m - 5°C
Elongation : 45 secondes à 72°C
- 4ème étape :
Elongation : 7 minutes à 72°C

TABLEAU 1.2 – Mélange PCR pour l'amplification de marqueurs STS.

	Fournisseur	Concentration stock	Concentration finale	Quantité (μL)
Tampon	Applied Biosystems	10 X	1 X	1,5
MgCl ₂	Applied Biosystems	25 mM	2 mM	1,2
dNTP	Promega	2,5 mM	100 μM	0,6
BSA	Sigma-Aldrich	10 mg/mL		0,3
Amorce R	Sigma-Aldrich	5 μM	0,13 μM	0,4
Amorce L	Sigma-Aldrich	5 μM	0,13 μM	0,4
Taq polymérase	Applied Biosystems	5 U/ μL	0,3 U/ μL	0,075
ADN		1,5 ng/ μL ou 1/20		2
H ₂ O				QSP 15

1.2.2.3 Analyse des produits d'amplification

Validation de la spécificité des produits d'amplification Les amorces nouvellement définies (marqueurs SCAR et BES) ont d'abord été commandées non-marquées et testées par PCR sur un panel d'individus de génotypes différents pour le locus *S*. Un volume de 8 μL de produit de PCR a été mélangé à 2 μL de bleu de charge puis a été déposé sur gel d'agarose (1% d'agarose, tampon TBE 0,5 X, BET 0,5 $\mu\text{g}/\text{mL}$). La migration s'est effectuée à 120 V pendant 30 à 45 minutes. Chaque gel a ensuite été placé sous lampe UV afin de visualiser les produits d'amplification. Un polymorphisme de taille (bandes de tailles différentes d'un génotype à l'autre) ou de séquence au niveau des amorces (présence/absence du fragment) a éventuellement été mis en évidence grâce à l'utilisation de génotypes différents au locus *S*. Cette étape de migration a également permis l'identification des couples d'amorces générant un profil multibande, c'est à dire aspécifique ou/et amplifiant des séquences répétées dans le génome.

Une fois les couples d'amorces validés comme amplifiant une séquence copie unique, plusieurs types d'analyses ont été réalisés.

Recherche de polymorphisme de taille - InDel L'analyse des produits PCR a été réalisée sur un séquenceur capillaire (*3130xl* Genetic Analyzer, Applied Biosystems). Pour chaque couple utilisé, une des deux amorces a été marquée avec un fluorochrome (6-FAM ou Hex). Les mélanges de migration réalisés à partir des produits PCR (Tableau 1.3) ont été dénaturés 4 minutes à 95°C, placés sur glace quelques minutes, puis les produits d'amplification ont migré en conditions dénaturantes (polymère POP7, tampon 10X + EDTA, Applied Biosystems). Les migrations ont été réalisées à 60°C avec un temps d'injection compris entre 15 et 18 secondes, un temps de migration de 900 secondes et un voltage d'injection de 1 à 3 kV. Les données génotypiques des différents échantillons amplifiés ont été collectées et analysées au moyen du logiciel Genemapper (v4.0, Applied Biosystems).

Recherche de polymorphisme de séquence - SNP

- Technique SSCP (Single Strand Conformation Polymorphism), étude du polymorphisme de conformation :

Les deux amorces utilisées pour les amplifications PCR ont été marquées avec un fluorochrome (6-FAM et Hex) afin de marquer les 2 brins d'ADN amplifiés puis les produits PCR ont été analysés sur un séquenceur capillaire (*3130xl* Genetic

Analyzer, Applied Biosystems). Les mélanges obtenus à partir des produits PCR (Tableau 1.3) ont été dénaturés 4 minutes à 95°C, placés sur glace quelques minutes puis les produits d'amplification ont migré en conditions dénaturantes (polymère CAP, tampon GAB 1 X, Applied Biosystems). Les migrations ont été réalisées à 18, 25 et 30°C avec un temps d'injection compris entre 15 et 18 secondes, un temps de migration allant de 1 400 à 1 800 secondes et un voltage d'injection de 1 à 3 kV. Les données génotypiques des différents échantillons amplifiés ont été collectées et analysées au moyen du logiciel Genemapper (v4.0, Applied Biosystems).

- Technique HRM (High-Resolution Melting), étude du polymorphisme des courbes de fusions :

L'amplification PCR a été réalisée avec des amorces non marquées et le kit d'amplification LightCycler®480 High Resolution Melting Master (Roche) dans un volume de 15 μL (Tableau 1.4). Les réactions PCR et la génération des courbes de fusion ont été réalisées à l'aide d'un thermocycleur Rotorgene (Corbett). Les résultats ont été analysés au moyen du logiciel Rotor-Gene Q Series Software 1.7.

TABLEAU 1.3 – Mélange de migration sur séquenceur ABI en condition dénaturante (POP7) et non-dénaturante (CAP).

	Fournisseur	POP7	CAP
Produit PCR		1 μL	1 μL
Formamide	Applied biosystems	10 μL	10 μL
Marqueur de taille 500 Lys	Applied biosystems	0,1 μL	0,3 μL
NaOH		-	0,3 μL

TABLEAU 1.4 – Mélange PCR utilisé dans le cadre de la technique HRM.

	Fournisseur	Concentration initiale	Concentration finale	Quantité
Master mix	Roche	2 X	1 X	7,5
MgCl ₂	Roche	25 mM	2 mM	1,2
Amorce R	Sigma-Aldrich	5 μM	0,125 μM	0,375
Amorce L	Sigma-Aldrich	5 μM	0,125 μM	0,375
ADN		1,5 ng/ μL		2,25
H ₂ O				QSP 15 μL

1.2.3 Marqueurs AFLP

La technique AFLP (Amplified Fragment-Length Polymorphism, Vos *et al.*, 1995) correspond à l'amplification sélective de fragments de restriction à partir d'ADN digéré (Figure 1.5). Cela permet d'obtenir un grand nombre de marqueurs moléculaires sans avoir d'information sur la séquence. Cette approche a été utilisée pour densifier la carte génétique de la région du locus *S*.

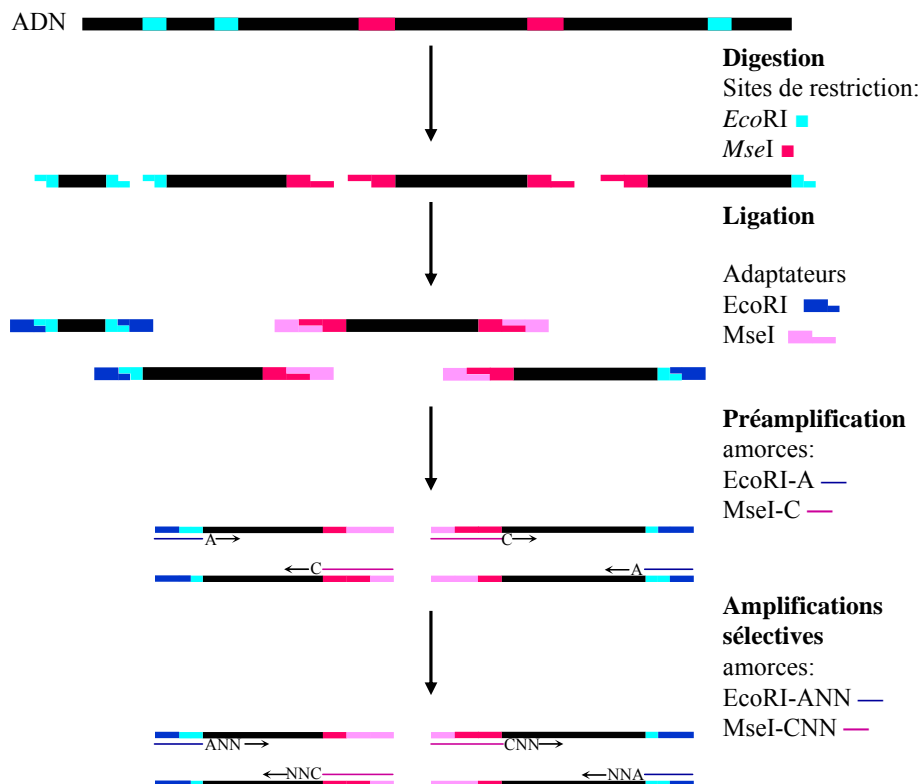


FIGURE 1.5 – Schéma des différentes étapes de la production de marqueurs AFLP.

1.2.3.1 Restriction enzymatique et ligation des adaptateurs

Dans un premier temps, 500 ng d'ADN obtenu par extraction sur matériel frais ont été digérés par les enzymes de restriction *EcoRI* et *MseI*, dans un volume total de 30 μL à 37°C (Tableau 1.5) au bain marie.

Après 3h, un mélange de ligation de 5 μL (Tableau 1.6) a été ajouté aux 30 μL de

Chapitre 1. Matériel et méthodes

restriction. Ce mélange de restriction-ligation (RL) a été maintenu à 37°C pendant une nuit afin de permettre à deux types d'adaptateurs (séquences ci-dessous), d'environ 20 pb, de se lier aux extrémités cohésives des fragments de restriction.

- adaptateur *EcoRI* :

5'-CTC GTA GAC TGC GTA CC
 CAT CTG ACG CAT GGT TAA-5'

- adaptateur *MseI* :

5'-GAC GAT GAG TCC TGA G
 TA CTC AGG ACT CAT-5'

Pour produire ces adaptateurs, 1 500 pmol de chacun des brins d'ADN qui constituent l'adaptateur *EcoRI* et 15 000 pmol de ceux qui constituent l'adaptateur *MseI* ont respectivement été mélangés dans un volume final de 300 μL . Chaque solution a été chauffée à 90°C puis ramenée progressivement à température ambiante.

Pour valider l'étape de restriction-ligation, 5 μL de ligation ont été déposés sur gel d'agarose (1% d'agarose, tampon TBE 0,5 X, BET 0,5 $\mu\text{g}/\text{mL}$) et ont migré pendant 15 minutes à 100 V, un 'smear' compris entre 100 et 1 200 bp devant être visible sous lumière UV.

TABLEAU 1.5 – Mélange de digestion enzymatique.

	Fournisseur	Concentration stock	Concentration finale	Quantité
ADN			25 ng/ μL	20 μL
Tampon Ligase	Biolabs	10 X	1 X	3 μL
BSA	Biolabs	100 X	5 X	1,5 μL
NaCl		0,5 M	50 mM	3 μL
<i>EcoRI</i>	Biolabs	20 U/ μL	10 U	0,5 μL
<i>MseI</i>	Biolabs	4 U/ μL	10 U	2,5 μL
ADN				500 ng
H ₂ O				QSP 30 μL

TABLEAU 1.6 – Mélange de ligation.

	Fournisseur	Concentration stock	Concentration finale	Quantité (μL)
Adaptateur <i>EcoRI</i>	Sigma-Aldrich	5 pmol/ μL	1 pmol/ μL	1
Adaptateur <i>MseI</i>	Sigma-Aldrich	50 pmol/ μL	10 pmol/ μL	1
Tampon Ligase	Biolabs	10 X	1 X	0,5
ATP	MBI Fermentas	10 mM	1 mM	0,5
T4 DNA ligase	Biolabs	400 coh.U/ μL	64 coh.U/ μL	0,16
H ₂ O				QSP 5

1.2.3.2 Préamplification

Cette seconde étape se caractérise par l'amplification PCR (Tableau 1.7) des fragments de restriction à partir d'amorces dont la séquence est complémentaire de celle des adaptateurs et à laquelle une base sélective a été rajoutée en 3' :

- EcoRI-A : 5'-GACTGCGTACCAATTCA-3'
- MseI-C : 5'-GATGAGTCCTGAGTAAC-3'

La préamplification (PA) a été réalisée dans les conditions PCR suivantes :

- 1ère étape, 20 cycles :
 - Dénaturation : 30 secondes à 94°C
 - Hybridation : 30 secondes à 56°C
 - Elongation : 1 minute à 72°C
- 2ème étape
 - Elongation finale 5 minutes à 72°C

Pour valider l'étape de PA, 8 μL de produit PCR ont été déposés sur gel d'agarose (1% d'agarose, tampon TBE 0,5 X, BET 0,5 $\mu\text{g}/\text{mL}$) et ont migré pendant 45 minutes à 100 V, un 'smear' compris entre 50 et 500 bp devant être visible sous lumière UV.

TABLEAU 1.7 – Mélange de préamplification.

	Fournisseur	Concentration stock	Concentration finale	Quantité (μL)
Tampon	Applied biosystems	10 X	1 X	3
Mgcl2	Applied biosystems	25 mM	2,5 mM	3
dNTP	Promega	2,5 mM	200 μM	2,4
Amorce EcoRI-A	Sigma-Aldrich	8 pmol/ μL	0,266 pmol/ μL	1
Amorce MseI-C	Sigma-Aldrich	8 pmol/ μL	0,266 pmol/ μL	1
Taq DNA polymérase	Applied biosystems	5 U/ μL	0,05 U/ μL	0,3
RL diluée 10 fois		1 X	0,1 X	3
H ₂ O				QSP 30

1.2.3.3 Amplification sélective

La dernière étape consiste en une amplification sélective (AS) à partir du produit de la préamplification (Tableau 1.8). Deux bases sélectives supplémentaires ont été rajoutées aux amorces utilisées. Cela correspond à 16 amorces EcoRI-A+2 et 16 amorces MseI-C+2 différentes, ce qui représente au final 256 combinaisons possibles de couples d'amorces (Tableau 1.9). Les amorces EcoRI-A+2 ont été marquées avec un fluorochrome (6-Fam ou Hex). L'AS a été réalisée dans les conditions PCR suivantes :

Chapitre 1. Matériel et méthodes

- 1ère étape, 13 cycles :
 - Dénaturation : 30 secondes à 94°C
 - Hybridation : 30 secondes à 65°C (-0,7°C par cycle)
 - Elongation : 1 minute à 72°C
- 2ème étape, 23 cycles :
 - Dénaturation : 30 secondes à 94°C
 - Hybridation : 30 secondes à 56°C
 - Elongation : 1 minute à 72°C
- 3ème étape :
 - Elongation finale 5 minutes à 72°C

TABLEAU 1.8 – Mélange d’amplification sélective.

	Fournisseur	Concentration stock	Concentration finale	Quantité (μL)
Tampon	Applied biosystems	10 X	1 X	1 μL
MgCl ₂	Applied biosystems	25 mM	2 mM	0,8 μL
dNTP	Promega	2,5 mM	200 μM	0,8 μL
EcoRI-A+2	Sigma-Aldrich	8 pmol/ μL	0,24 pmol/ μL	0,3 μL
MseI-C+2	Sigma-Aldrich	8 pmol/ μL	0,24 pmol/ μL	0,3 μL
Taq DNA polymérase	Applied biosystems	5 U/ μL	0,025 U/ μL	0,05 μL
PA diluée 50 fois				3 μL
H ₂ O				QSP 10 μL

TABLEAU 1.9 – Amorces EcoRI-A+2 et MseI-C+2 utilisées pour l’amplification sélective.

Amorces EcoRI-A+2	Amorces MseI-C+2
EcoRI-AAC	MseI-CAA
EcoRI-AAG	MseI-CAC
EcoRI-AGC	MseI-CAG
EcoRI-AGG	MseI-CAT
EcoRI-ACA	MseI-CCA
EcoRI-ACT	MseI-CCC
EcoRI-ACC	MseI-CCG
EcoRI-ACG	MseI-CCT
EcoRI-AAA	MseI-CGA
EcoRI-AAT	MseI-CGC
EcoRI-AGA	MseI-CGG
EcoRI-AGT	MseI-CGT
EcoRI-ATA	MseI-CTA
EcoRI-ATC	MseI-CTC
EcoRI-ATG	MseI-CTG
EcoRI-ATT	MseI-CTT

1.2.3.4 Analyse des produits d'amplification

La migration du produit des amplifications sélectives a été réalisée en conditions dénaturantes (polymère POP4, tampon 10 X, Applied Biosystems) sur séquenceur capillaire (*3130xl* Genetic Analyzer, Applied Biosystems). La séparation s'est effectuée à 60°C avec un temps d'injection de 15 secondes, un temps de migration de 1 400 secondes et un voltage d'injection de 3 kV. Les profils AFLP ont été collectés par Datacollector (Applied Biosystems) et analysés avec le logiciel Genotyper. Les fragments d'intérêt ont été nommés en fonction des deux bases sélectives des amorces EcoRI-A+2 et MseI-C+2 et de leur taille : par exemple, un marqueur de 134 pb obtenu à partir des amorces EcoRI-AAA et MseI-CCC a été nommé aacc134.

1.2.4 Transformation de marqueurs AFLP en marqueurs SCAR

L'objectif de cette transformation est de disposer de marqueurs spécifiques de locus (marqueurs SCAR) à partir de marqueurs AFLP.

1.2.4.1 Simplification d'un profil AFLP

La première étape pour transformer un marqueur AFLP en marqueur SCAR est de simplifier le profil AFLP afin de pouvoir extraire le fragment d'intérêt d'un gel de polyacrylamide et le cloner. Pour cela, nous avons choisi de suivre la procédure qui consiste à ajouter trois bases sélectives supplémentaires à l'amorce MseI-C+2 (Brugmans *et al.*, 2003 ; Paris *et al.*, 2011). Pour identifier ces trois bases, 12 amorces dégénérées (Tableau 1.10) ont été utilisées pour réaliser des amplifications dans des conditions identiques à l'AS. Le produit de la préamplification a été remplacé par celui de l'AS dilué 100 fois et une amorce EcoRI-A+2 marqué au fluorochrome IRD800 a été utilisée. Les produits PCR obtenus ont ensuite été déposés sur gel de polyacrylamide 8% (Tableau 1.11) et la migration s'est faite sur séquenceur automatique (LI-COR 4200, Sciencetec). Afin de faciliter l'identification du fragment d'intérêt, les amplifications ont été réalisées sur quatre individus S_1S_1 et quatre individus S_2S_2 déposés côte à côte sur le gel.

Une fois les 3 nucléotides supplémentaires identifiés, une amorce MseI-A+5 a alors été définie. Le produit de l'amplification sélective des individus homozygotes S_1S_1 et S_2S_2

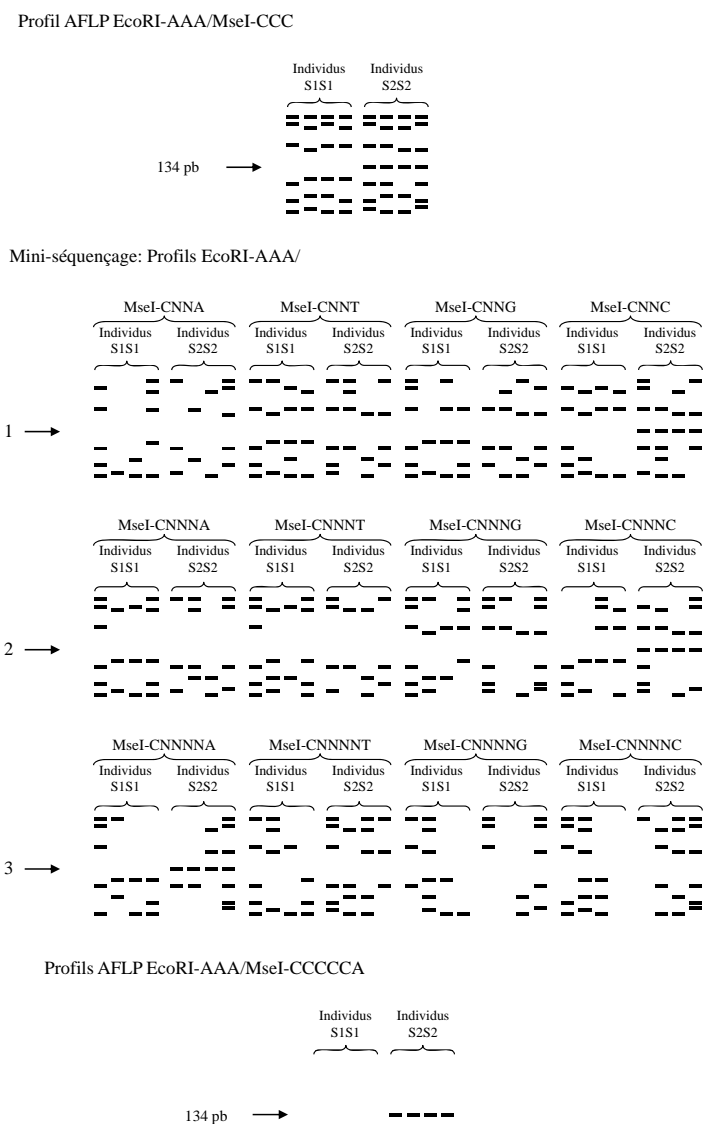


FIGURE 1.6 – Schéma des différentes étapes de la simplification du profil AFLP associé au marqueur aacc134. 1 : la 4ème base sélective est C. 2 : la 5ème base sélective est C. 3 : la 6ème base sélective est A.

a été dilué 100 fois et amplifié avec cette nouvelle amorce et l'amorce EcoRI-A+2. Le profil AFLP obtenu par migration sur séquenceur LI-COR a été simplifié par rapport à l'amplification sélective, ce qui a rendu les fragments d'intérêt facilement repérables. Ils ont été excisés du gel et placés dans 300 μ L de TE. Après 2h d'incubation à température ambiante, les tubes ont été placés au moins 12h à -20° C.

Une nouvelle amplification sélective a été réalisée à partir de 1 μ L de fragments élués et

les amorces EcoRI-A+2 et MseI-C+5 correspondantes. Une étape de co-migration avec le profil AFLP de l'AS a ensuite été réalisée sur gel polyacrylamide : un seul fragment devait être amplifié et correspondre au fragment du profil AFLP d'origine (Figure 1.6).

TABLEAU 1.10 – Amorces dégénérées utilisées pour identifier 3 bases sélectives supplémentaires (d'après Brugmans *et al.*, 2003).

Amorce MseI	Séquence
Détermination de la 4ème base sélective	
3N+A	GATGAGTCCTGAGTAA NNNA
3N+C	GATGAGTCCTGAGTAA NNNC
3N+G	GATGAGTCCTGAGTAA NNNG
3N+T	GATGAGTCCTGAGTAA NNNT
Détermination de la 5ème base sélective	
4N+A	GATGAGTCCTGAGTAA NNNNA
4N+C	GATGAGTCCTGAGTAA NNNNC
4N+G	GATGAGTCCTGAGTAA NNNNG
4N+T	GATGAGTCCTGAGTAA NNNNT
Détermination de la 6ème base sélective	
5N+A	GATGAGTCCTGAGTAA NNNNNA
5N+C	GATGAGTCCTGAGTAA NNNNNC
5N+G	GATGAGTCCTGAGTAA NNNNNG
5N+T	GATGAGTCCTGAGTAA NNNNNT

TABLEAU 1.11 – Composition d'un gel polyacrylamide.

	Quantités pour un gel de 25 mL
Urée	10 g
Long Ranger 50%	4 g
H ₂ O	compléter à 23,5 g
TBE classique 5 X	4 mL
APS 10%	170 μ L
TEMED	17 μ L

1.2.4.2 Clonage

Ligation Les produits PCR obtenus à partir des fragments élués ont été purifiés avec le kit NucleoSpin extract II (Macherey-Nagel) puis dosés sur gel (6 μ L de produit purifié, agarose 1%, TBE 0,5 X, BET 0,5 μ g/mL, 100 V, 50 min). Ils ont ensuite été insérés dans un vecteur pGEM®-T selon les conditions techniques du kit pGEM®-T Vector system I (Promega) données par le fabricant.

Transformation Le kit One Shot® TOP10 Chemically Competent E.coli (Invitrogen) a été utilisé pour la transformation des cellules compétentes avec le vecteur pGEM®-T (Promega). Trois microlitres de ligation ont été ajoutés à 40 μL de cellules Top10, puis le mélange a été placé 20 minutes dans la glace. Le protocole suivi ensuite correspond aux instructions du manuel technique. Les bactéries ont été étalées sur un milieu LB 1 X (1% de tryptone, 0,5% d'extrait de levure, 10% de NaCl, 15% d'agar, pH 7) préalablement traité avec un mélange contenant 50 ng/mL d'ampicilline, 50 mg/mL de X-Gal et 100 mM d'IPTG. Après une incubation à 37°C pendant une nuit, les colonies blanches qui correspondaient aux bactéries transformées ont été prélevées et lysées 10 minutes à 95°C dans du TTE (Triton 1 X, Tris-HCl 20 mM, pH 8,5, EDTA 0,5 M pH 8) puis conservées à -20°C.

Validation et séquençage Des amplifications PCR ont été réalisées à partir des préparations TTE avec les amorces M13 universelles qui encadrent le site de clonage du vecteur. La taille des produits PCR a été estimée par migration sur gel (agarose 1%, TBE 0,5X, BET 0,5 $\mu\text{g}/\text{mL}$, 120 V, 45 min) et les clones ont été sélectionnés en fonction de la taille des inserts par rapport à la taille attendue. Les produits PCR (40 μL) des clones sélectionnés ont été envoyés à la plateforme Genoscreen⁵ qui a réalisé le séquençage des inserts à l'aide de l'amorce M13-20. Une fois la séquence d'un fragment AFLP d'intérêt obtenue, de nouvelles amorces ont été définies sur cette séquence, hors des adaptateurs afin d'obtenir un marqueur SCAR.

1.2.5 Banques BAC

1.2.5.1 Construction de deux banques BAC

Deux banques BAC (Bacterial Artificial Chromosome), $\text{Cin}S_2S_2$ et $\text{Cin}S_1S_4$, ont été construites en collaboration avec l'équipe du CNRGV⁶ (Centre National de Ressources Génomiques Végétales, INRA de Toulouse), à partir des géotypes C15 (S_2S_2) et C30.01 (S_1S_4). Ces individus ont été clonés et une dizaine de plantes représentant chaque géotype ont été acclimatés en serre (*cf.* paragraphe 1.1.3). Après environ 3 mois, les plantes ont été placées 5 à 7 jour à l'obscurité puis les jeunes feuilles ont été collectées, congelées dans de l'azote liquide et stockées à -80 °C jusqu'au moment de l'extrac-

5. <http://www.genoscreen.fr/>

6. <http://cnrgv.toulouse.inra.fr/>

tion de l'ADN de haut poids moléculaire. L'enzyme de restriction *Hind*III et le vecteur pIndigoBAC-5 ont été utilisés pour la construction des banques. Ces deux banques ont ensuite été déposées sur des filtres (macroarrays), chaque clone étant présent en double sur le filtre. Les protocoles suivis sont décrits en détails dans l'annexe C (Gonthier *et al.*, 2010).

1.2.5.2 Criblage des banques BAC

Les sondes utilisées pour cribler les banques BAC dans le cadre du clonage positionnel du locus *S* ont été définies, soit à partir de marqueurs moléculaires cartographiés sur la carte génétique, soit à partir de marqueurs définis sur des BES (*cf.* paragraphe 1.2.2). Dans les deux cas, les sondes ont été obtenues par amplification PCR à partir d'ADN de clones stockés dans du TTE (*cf.* paragraphe 1.2.4.2 et ci-après). Les produits PCR obtenus ont ensuite été purifiés avec le kit nucleospin extract II (Macherey-Nagel) et dosés sur gel d'agarose (1% d'agarose, tampon TBE 0,5 X, BET 0,5 μ g/mL), après une migration de 30 minutes à 120 V. Une quantité de 200 ng de sonde purifiée par filtre à hybrider a été produite. Les sondes ont été marquées radioactivement et hybridées sur les macroarrays. Les clones positifs ont été sélectionnés comme indiqué dans l'annexe C. L'ADN des clones validés par PCR a été extrait dans du TTE, 5 minutes à 95°C, puis conservé à -20°C.

Remarques :

- Les sondes correspondant aux marqueurs microsatellites ont été définies sur les séquences adjacentes au motif répété. Pour chaque marqueur microsatellite, deux sondes ont donc été définies.
- Le nom des marqueurs et sondes obtenus à partir des BES a été défini en fonction du nom du clone et de l'extrémité sur lesquels il a été défini. Ainsi, le marqueur défini sur l'extrémité T7 du clone Si1 a été nommé Si1t7.

1.2.5.3 Séquençage des clones BAC

Méthodes de séquençage des BES Le séquençage des BES des clones BAC, positifs en hybridation et validés par PCR, a été réalisé selon la méthode de Sanger, soit par la plateforme Genoscreen, soit par le CNRGV et la plateforme génomique⁷ de l'INRA de Toulouse. Au CNRGV, l'extraction de l'ADN des clones a été réalisée en plaques

7. <http://genomique.genotoul.fr/>

96 puits, avec le kit d'extraction Nucleospin 96 Flash (Macherey-Nagel). L'utilisation du vecteur pIndigoBAC-5 pour la construction des banques BAC a permis d'amorcer le séquençage des BES avec les amorces universelles T7prom et M13rev de part et d'autre des inserts. Les réactions de séquences ont été réalisées avec le kit ABI PRISM Big Dye terminator v31 (Applied Biosystems) et les échantillons ont été analysés sur séquenceur capillaire ABI (3730, Applied Biosystems).

Méthodes de séquençage des clones BAC et traitement des données brutes

Le séquençage complet des clones BAC a été réalisé par le CNRGV en collaboration avec la Plateforme Génomique et la plateforme Bioinformatique⁸ de l'INRA de Toulouse. La technologie 454 (GS FLX Titanium, Roche) a été utilisé pour le séquençage. L'extraction d'ADN a été réalisée avec le kit Nucleobond Xtra Midi (Macherey-Nagel). Au minimum 500ng d'ADN ont été nécessaires pour chaque échantillon. Une banque de fragments à séquencer a ensuite été préparée pour chaque clone avec les kits Titanium (Roche) dédiés.

Les données ont été nettoyées avec le logiciel PyroCleaner qui supprime les séquences trop petites et de faible qualité (contenu en N supérieur à 4%). Les séquences ont ensuite été décontaminées par élimination des séquences contenant au minimum 95% d'homologie avec le génome de la bactérie hôte *E. coli*. L'assemblage a été réalisé avec Newbler (v2.3, De novo Assembler, Roche), en utilisant les paramètres par défaut du logiciel.

1.2.5.4 Analyse des séquences

Pour chaque séquence obtenue (BES ou insert total), un Blastn a été réalisé sur les banques d'EST de plusieurs Asteraceae (Tableau 1.12) avec un seuil de e-value inférieur à 10^{-30} et un pourcentage d'identité supérieur à 95% pour les EST de *C. intybus*, supérieur à 90% pour les EST de *C. endivia* et supérieur à 85% pour *Lactuca sp*, *H. annuus* et *Taraxacum sp*. La banque de *C. intybus* résulte de l'assemblage d'EST présents dans les bases de données publiques, d'EST produits par le laboratoire (organe source : stigmat) et d'EST de l'université de Wageningen (organe source : racine, Cankar *et al.*, 2011). Pour les BES, des Blastn et Blastx ont également été réalisés sur les bases de données nr (non-redondant sequences) de NCBI dans les conditions standards définies par le site. Des Blastx ont été réalisés avec un seuil de e-value inférieur à 10^{-5} sur les bases

8. <http://bioinfo.genotoul.fr/>

de donné de peptides d'*Arabidopsis thaliana*, *Vitis vinifera*, *Mimulus guttatus*, *Solanum lycopersicum*, *Medicago truncatula* et *Populus trichocarpa*. La source des bases de données utilisées est indiquée dans le tableau 1.13.

Le programme RepeatMasker⁹ a été utilisé en utilisant le mode 'cross_match' et l'espèce *A. thaliana* comme référence. Des prédictions de gènes ont été générées à partir des logiciels GenScan¹⁰ (organisme de référence : *A. thaliana*) et Fgenesh¹¹ (organismes de référence : *A. thaliana* et *V. vinifera*).

TABLEAU 1.12 – Bases de données EST des différentes espèces d'Asteraceae utilisées.

EST : nombre de séquences utilisées pour l'assemblage. SU : nombre de séquences uniques obtenues après assemblage (singletons et contigs). Contig : nombre de contigs parmi les SU. Annoté : pourcentage de SU présentant une homologie avec la base de peptides d'*A. thaliana*.

Espèce	EST	SU	Contig	Annoté (%)	Source
<i>Cichorium intybus</i>	198657	110155	25322	50,8	CGP Davis ¹² , NCBI ¹³ , SADV, WUR, Wageningen University
<i>Cichorium endivia</i>	30170	18951	5295	76,7	CGP Davis
<i>Helianthus annuus</i>	93428	31605	10796	57	CGP Davis
<i>Lactuca sativa</i>	92560	22870	13129	-	CGP Davis
<i>Taraxacum sp.</i>	41294	16661	6707	-	CGP Davis

TABLEAU 1.13 – Bases de données peptidiques utilisées.

Espèce	Source	Version
<i>A. thaliana</i>	TAIR ¹⁴	TAIR10
<i>V. vinifera</i>	Genoscope-CNS ¹⁵	12 X, 2010
<i>M. guttatus</i>	Phytozome ¹⁶	2010
<i>S. lycopersicum</i>	Sol genomics network ¹⁷	2011
<i>M. truncatula</i>	Phytozome	2010
<i>P. trichocarpa</i>	Phytozome	2010

9. A.F.A. Smit, R. Hubley et P. Green RepeatMasker at <http://repeatmasker.org>

10. <http://genes.mit.edu/GENSCAN.html>

11. <http://linux1.softberry.com/berry.phtml>

12. http://compgenomics.ucdavis.edu/compositae_index.php

13. <http://www.ncbi.nlm.nih.gov/dbEST/>

14. <http://www.arabidopsis.org/>

15. Centre National de Séquençage, <http://www.cns.fr/vitis>

16. <http://www.phytozome.net/>

17. http://solgenomics.net/genomes/Solanum_lycopersicum/index.pl

1.3 Cartographie génétique de la région du locus S

1.3.1 Calcul des distances génétiques

La fréquence de recombinaison (r) entre 2 locus est définie telle que :

$$r = \frac{\text{nombre de chromosomes recombinants}}{\text{nombre total de chromosomes}}.$$

Dans le cas où la fréquence de recombinaison entre locus est inférieure à 0,1, l'utilisation de fonctions de cartographie n'est pas nécessaire. En effet, la distance génétique entre marqueurs est directement proportionnelle à la fréquence de recombinaison, du fait de la faible probabilité d'observer des doubles crossing-overs dans l'intervalle considéré (Lorieux, 1994). Elle est exprimée en centimorgans (cM) : $D \text{ (cM)} = 100 r$

1.3.2 Etude de la ségrégation des marqueurs

La conformité de ségrégation des marqueurs a été évaluée par un test du χ^2 :

$$\chi^2 = \sum_{i=1}^t \frac{(n_i - n_{th})^2}{n_{th}}$$

n_i correspond aux effectifs des t génotypes observés et n_{th} aux effectifs attendus dans le cas d'une ségrégation de type mendélien. Si une différence significative est observée, alors la ségrégation est distordue. Pour déterminer si la sélection à l'origine de la distorsion est gamétique (A) ou zygotique (B), 2 test successifs ont été réalisés (Lorieux *et al.*, 1995) :

$$\text{(A)} \quad \chi^2 = \sum_{i=1}^t \frac{(2np_i - n)^2}{n}$$

t étant le nombre d'allèles au sein de la population, p_i la fréquence associée à chacun de ces allèles et n l'effectif total de la population ;

$$\text{(B)} \quad \chi^2 = \sum_{i,j}^t \frac{(n_{ij} - np_i p_j)^2}{np_i p_j}$$

i et j correspondant aux t allèles présents au sein de la population, n_{ij} étant l'effectif associé à chacune des combinaisons alléliques, p_i et p_j les fréquences associées à chacun de ces allèles et n l'effectif total de la population.

Selon Lorieux *et al.* (1995), si (A) est significatif et (B) non-significatif, la sélection est gamétique. Au contraire, si (B) est significatif, la sélection est zygotique, quel que soit le résultat obtenu pour (A).

1.3.3 Précision des distances génétiques en fonction de la taille de l'échantillon à cartographier

Lorsque l'on calcule un taux de recombinaison r entre deux marqueurs, la précision de l'estimation dépend de l'écart type exprimé en fonction de la quantité d'information porté par un individu (i_r) et de la taille de la population (n), tel que :

$$S_r = \sqrt{\frac{1}{ni_r}} \quad (\text{Allard, 1956 ; Lorieux, 1994}).$$

1.3.3.1 Quantité d'information portée par individu (i_r)

Dans le cadre de l'estimation des distances génétiques, la quantité d'information (i_r) porté par chaque individu est définie telle que :

$$i_r = \sum_{i=1}^t \left(\frac{1}{m} \left(\frac{\delta m}{\delta r} \right)^2 \right)$$

où m_1, m_2, \dots, m_t sont les proportions théoriques des t classes de ségrégation exprimées en fonction de la fréquence de recombinaison r (Allard, 1956 ; Lorieux, 1994). L'expression de i_r est donc dépendante du type de marqueurs utilisés (Tableau 1.14). Des exemples d'équations permettant d'obtenir i_r sont présentés ci-après.

Expression de i_r dans le cas de marqueurs codominants Dans le cas de marqueurs codominants hétérozygotes chez deux parents de génotypes distincts, on obtient une descendance complètement classée (CC), c'est-à-dire que toutes les classes de génotype sont discernables. La fréquence de chacun des génotypes possibles (m) et la dérivée associée ($\frac{\delta m}{\delta r}$) sont exprimées en fonction de r (Tableau 1.15a). A partir de ces équations, on obtient l'expression de i_r apportée par chaque individu :

$$\begin{aligned} i_r &= 4 \left[\frac{4}{(1-r)^2} \times \left(\frac{2r-2}{4} \right)^2 \right] + 8 \left[\frac{4}{r(1-r)} \times \left(\frac{1-2r}{4} \right)^2 \right] + 4 \left[\frac{4}{r^2} \times \left(\frac{2r}{4} \right)^2 \right] \\ &= 4 + 2 \left[\frac{(1-2r)^2}{r(1-r)} \right] + 4 \\ &= \frac{8(r-r^2) + 2(1-4r+4r^2)}{r-r^2} \\ &= \frac{2}{r-r^2} \end{aligned}$$

Expression de i_r dans le cas de marqueurs dominants Dans le cas d'un couple de marqueurs dominants chez les deux parents, on peut observer deux cas de figure qui dépendent des phases des allèles chez les parents. Soit les allèles dominants sont en couplage (descendance DD couplage), soit en répulsion (descendance DD répulsion). Du fait de la dominance, seuls quatre phénotypes différents sont discernables pour 16 génotypes possibles. La fréquence de chacun de ces phénotypes et la dérivée associée sont exprimées en fonction de r (Tableau 1.15b). A partir de ces équations, on obtient l'expression de i_r apportée par chaque individu, ici dans un cas DD en répulsion :

$$\begin{aligned} i_r &= \left[\frac{4}{2+r^2} \times \left(\frac{2r}{4}\right)^2 \right] + 2 \left[\frac{4}{1-r^2} \times \left(\frac{-2r}{4}\right)^2 \right] + \left[\frac{4}{r^2} \times \left(\frac{2r}{4}\right)^2 \right] \\ &= \frac{r^2}{2+r^2} + 2 \left(\frac{r^2}{1-r^2} \right) + 1 \\ &= \frac{r^2(1-r^2) + 2r^2(2+r^2) + (1-r^2)(2+r^2)}{(2+r^2)(1-r^2)} \\ &= \frac{2(1+2r^2)}{(2+r^2)(1-r^2)} \end{aligned}$$

TABLEAU 1.14 – Expression de i_r en fonction du type de marqueurs considéré. He : hétérozygote, Ho : homozygote, CC : descendance complètement classée, BC : backcross, DD : dominant-dominant. *Dans le cas de marqueurs dominants, l'individu est Ho pour les allèles récessifs.

Types de marqueurs	Génotype des parents	i_r	Configuration
Codominants	2 He (génotypes distincts)	$\frac{2}{r-r^2}$	CC
Codominant/Dominant	1 He et 1 Ho*	$\frac{1}{r-r^2}$	BC
Codominant/Dominant	2 He	$2 + \frac{(1-r)^2}{r(2-r)} + \frac{(1-2r)^2}{2(1-r+r^2)}$	Ségrégation 6.3.3.2.1.1
Dominants en répulsion	2 He	$\frac{2(1+2r^2)}{(2+r^2)(1-r^2)}$	DD répulsion
Dominants en couplage	2 He	$\frac{2(1+2(1-r)^2)}{(2+(1-r)^2)(1-(1-r)^2)} + \frac{r^2}{1-r^2} + \frac{(1-2r)}{2r(1-r)}$	DD couplage

1.3. Cartographie génétique de la région du locus S

TABLEAU 1.15 – Exemples de fréquences attendues (m) et dérivées correspondantes ($\frac{\delta m}{\delta r}$) pour les différents génotypes/phénotypes observés au sein d'une descendance. (a) Cas de deux marqueurs CC (génotype des parents : A1A2 B1B2 et A1A3 B1B3). (b) Cas de deux marqueurs DD en répulsion (génotype des parents : AabB). Pour obtenir les expressions correspondant aux marqueurs DD en couplage, $1 - r$ est substitué à r .

(a) Configuration CC			(b) Configuration marqueurs DD en répulsion		
Génotype	m	$\frac{\delta m}{\delta r}$	Phénotype	m	$\frac{\delta m}{\delta r}$
$\frac{A_1 B_1}{A_1 B_1}$			A_B_	$\frac{2 + r^2}{4}$	$\frac{2r}{4}$
$\frac{A_2 B_2}{A_1 B_1}$	$(1 - r)^2$	$\frac{\delta m}{\delta r} = \frac{(2r - 2)}{4}$	A_bb	$\frac{1 - r^2}{4}$	$\frac{4}{2r}$
$\frac{A_3 B_3}{A_1 B_1}$	$\frac{4}{4}$		aaB_	$\frac{4}{1 - r^2}$	$\frac{4}{2r}$
$\frac{A_3 B_3}{A_2 B_2}$			aabb	$\frac{r^2}{4}$	$\frac{4}{2r}$
$\frac{A_1 B_1}{A_1 B_1}$				$\frac{4}{4}$	
$\frac{A_1 B_2}{A_1 B_1}$					
$\frac{A_1 B_1}{A_1 B_2}$					
$\frac{A_3 B_3}{A_2 B_1}$					
$\frac{A_1 B_1}{A_2 B_1}$	$m = \frac{r(1 - r)}{4}$	$\frac{\delta m}{\delta r} = \frac{(1 - 2r)}{4}$			
$\frac{A_3 B_3}{A_1 B_1}$					
$\frac{A_1 B_3}{A_1 B_1}$					
$\frac{A_3 B_1}{A_2 B_2}$					
$\frac{A_1 B_3}{A_2 B_2}$					
$\frac{A_3 B_1}{A_3 B_1}$					
$\frac{A_2 B_1}{A_2 B_1}$					
$\frac{A_1 B_3}{A_2 B_1}$	$m = \frac{r^2}{4}$	$\frac{\delta m}{\delta r} = \frac{2r}{4}$			
$\frac{A_3 B_1}{A_1 B_2}$					
$\frac{A_3 B_1}{A_1 B_2}$					
$\frac{A_1 B_3}{A_1 B_3}$					

1.3.3.2 Calcul de la taille de l'échantillon à cartographier

Si on considère que la distribution des marqueurs sur le chromosome suit une loi normale, alors l'intervalle de confiance correspondant à r est :

$$c = r \pm z_{\frac{\alpha}{2}} S_r,$$

$z_{\frac{\alpha}{2}}$ correspondant à la déviation standard de la loi normale centrée réduite. Pour pouvoir déterminer avec certitude l'ordre de plus de deux marqueurs les uns par rapport au autre, il faut que les valeurs de c correspondant aux différents intervalles ne se chevauchent pas. Pour cela, $z_{\frac{\alpha}{2}} S_r$ doit être inférieur ou égale à $\frac{r}{2}$ (Liu, 1998 ; Lorieux, 1994). A partir de ce postulat, la valeur de n minimale peut être estimée :

$$\begin{aligned} \frac{r}{2} &\geq z_{\frac{\alpha}{2}} S_r \\ \Rightarrow \frac{r}{2} &\geq z_{\frac{\alpha}{2}} \sqrt{\frac{1}{ni_r}} \\ \Rightarrow r^2 &\geq (2z_{\frac{\alpha}{2}} \sqrt{\frac{1}{ni_r}})^2 \\ \Rightarrow n &\geq \frac{(2z_{\frac{\alpha}{2}})^2}{i_r r^2} \end{aligned}$$

La taille de l'échantillon à cartographier dépend donc de la précision que l'on souhaite atteindre (r), du type de marqueur dont on dispose (i_r) et du seuil α fixé.

1.3.4 Détermination du génotype au locus S

Dans le cadre de la carte haute résolution (*cf.* chapitre 4), le locus S est un des locus qui a du être cartographié. Deux méthodes ont permis de déterminer le génotype des plantes étudiées au locus S .

1.3.4.1 Inférence du génotype au locus S

Cette méthode est indirecte mais rapide à mettre en œuvre. Les individus ont été génotypés pour les marqueurs moléculaires encadrant le locus S obtenus à partir de la carte génétique de l'ensemble du génome (*cf.* chapitre 2) et de la carte haute densité

au voisinage du locus *S* (*cf.* chapitre 3). L'intervalle entre ces marqueurs étant très petit, aucun double recombinant n'a été observé (*cf.* chapitre 2). De ce fait, lorsque les combinaisons d'allèles se sont avérées identiques à celles observées sur les chromosomes parentaux, le génotype au locus *S* a pu être inféré.

1.3.4.2 Phénotypage du locus *S*

Lorsque le génotype au locus *S* ne peut être déterminé par inférence en raison d'un événement de recombinaison, il peut l'être par le phénotypage, comme cela a été fait pour la population B_a (*cf.* chapitre 2), grâce à des croisements avec des individus testeurs. Pour cela, des individus testeurs ont été sélectionnés en fonction des génotypes potentiels des individus à phénotyper (*cf.* chapitre 2). À partir du génotypage des marqueurs flanquants, le locus *S* est déjà partiellement génotypé : l'allèle *S* porté sur le chromosome parental est inféré et il existe deux génotypes *S* possibles sur le chromosome recombinant.

Deux phénotypes testeurs ont été définis pour chacune des combinaisons alléliques présentes au sein des populations utilisées pour la carte haute résolution (*cf.* chapitre 4). Les relations de dominance qui existent entre allèles *S* ont également été prises en compte (*cf.* chapitre 2, figure 2.3 p.78). Les combinaisons de testeurs utilisés comme donneurs de pollen ont été définies pour qu'au moins un des deux types de croisement soit compatible, quel que soit le génotype de l'individu à phénotyper, afin de disposer d'un témoin positif (Tableau 1.16). Quatre répétitions de chaque croisement ont été réalisées. Le groupe de compatibilité auquel appartient chaque individu a été déterminé sur la base du taux de mise à graine moyen (seed-set moyen, nombre de graines/nombre d'ovules par capitule) obtenu pour chaque croisement, ce qui a permis de déterminer le génotype au locus *S* correspondant (*cf.* chapitre 2, figure 2.2 p.78).

TABLEAU 1.16 – Schémas de croisements définis pour phénotyper les individus recombinants dans la région du locus S (*cf.* chapitre4). + indique que le croisement est compatible et - qu'il est incompatible.

Génotype de l'individu à phénotyper		Phénotype du testeur (pollen)			
Allèle S inféré	Allèles S à déterminer	$[S_1]$	$[S_2]$	$[S_3]$	$[S_4]$
S_1	S_3			+	+
	S_4			+	-
S_1	S_2			+	+
	S_4			+	-
S_1	S_2 S_3	Phénotypage impossible car $S_1 > S_2$ et $S_1 > S_3$			
S_2	S_3			-	+
	S_4			+	-
S_2	S_1	-		+	
	S_3	+		-	
S_2	S_1	-	+		
	S_2	+	-		
S_2	S_2		-	+	
	S_3		+	-	
S_3	S_1	-		+	
	S_2	+		-	
S_3	S_2	+			+
	S_4	+			-
S_4	S_1	-		+	
	S_2	+		+	
S_4	S_1	-		+	
	S_3	+		-	

Première partie
Carte fine du locus S

Chapitre 2

Résultats préliminaires : Carte moyenne densité du locus *S*

Afin d'étudier l'autoincompatibilité sporophytique chez la chicorée, une population d'individus nommée B_a a été produite à partir du croisement des individus K28 et K59. Ces deux individus sont issus de pollinisations libres de la variété population de chicorée industrielle Koospol. K28 présente un phénotype mâle-stérile (déterminisme nucléaire et monogénique, locus NMS_1) et K59 est un hermaphrodite très autoincompatible. K28 a donc été utilisé comme parent femelle et pollinisé par K59. Une première série de croisements a permis d'obtenir 174 individus en 2001, puis une seconde série a permis d'obtenir 215 individus en 2004. La population B_a est donc constituée d'un total de 389 individus.

Les travaux présentés dans ce chapitre ont été réalisés sur cette population et constituent la base à partir de laquelle a été entreprise la démarche de clonage positionnel du locus *S* (Figure 2.1). Ces travaux ont été initiés avant le début de la thèse et une partie des résultats présentés étaient acquis. Ils ont été complétés durant la première année de thèse par d'autres données de phénotypage, génotypage et l'interprétation complète des résultats de croisements. L'ensemble de ces résultats a permis d'aboutir à la carte moyenne densité du locus *S*.

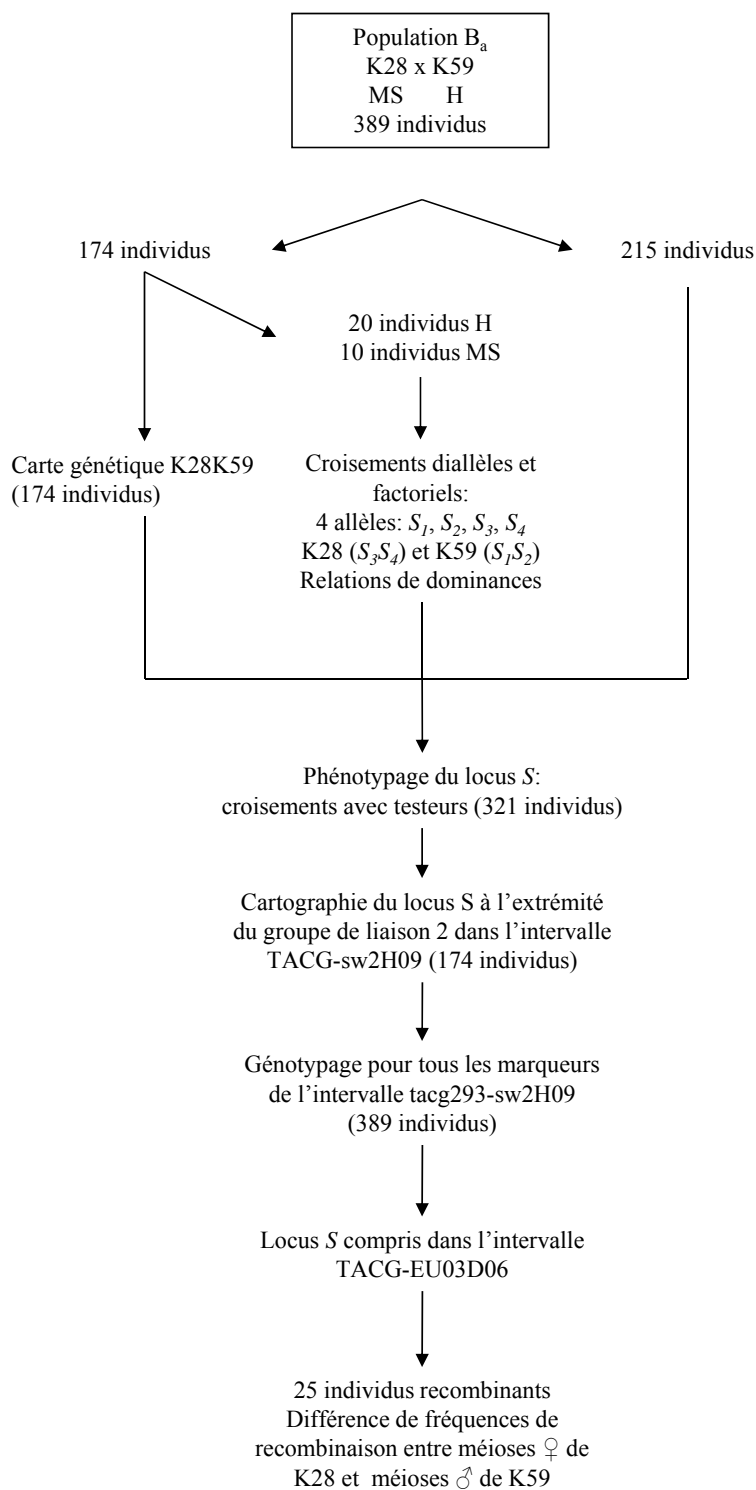


FIGURE 2.1 – Bases de la démarche de clonage positionnel du locus *S* : récapitulatif des différents travaux réalisés à partir de la population B_a. H : individu hermaphrodite, MS : individu mâle-stérile

2.1 Carte génétique

Le groupement d'intérêt scientifique CARTOCHIC a été constitué afin de construire une carte génétique de référence de la chicorée. Cette carte consensus a été obtenue à partir de 3 descendance distinctes : 2 de chicorée industrielle et 1 de chicorée witloof (Cadalen *et al.*, 2010). La population B_a de 2001 est l'une des 2 populations de chicorée industrielle utilisées pour le développement de cette carte consensus. Pour cette population, 274 marqueurs SSR et STS ont été cartographiés sur 9 groupes de liaison pour une longueur totale de 749,8 cM. Cette carte, appelée carte K28K59, présente une couverture du génome estimée à 93,6%. De plus, les 9 groupes de liaison correspondent vraisemblablement aux 9 chromosomes du génome haploïde de la chicorée. Cette carte génétique en voie de saturation peut donc être considérée comme une carte de référence.

Dans le cadre de l'étude de la stérilité mâle nucléaire, le polymorphisme généré par 36 couples d'amorces AFLP a été exploité et a permis de cartographier 350 marqueurs supplémentaires sur la carte K28K59 (Quillet *et al.*, en préparation). L'ajout de ces marqueurs n'a pas significativement augmenté la distance totale de la carte. Cela a confirmé que la carte génétique K28K59 est pratiquement saturée en marqueurs et donc représentative de l'ensemble du génome.

2.2 Cartographie du locus *S*

2.2.1 Identification des allèles *S* présents chez K28 et K59 et de leurs relations de dominance

A partir de la population de 2001, 20 individus hermaphrodites et 10 individus mâle-stériles ont été sélectionnés. Des croisements diallèles et factoriels ont été réalisés entre ces individus ainsi qu'avec K28 et K59. Pour chaque croisement, le seed-set a permis de distinguer clairement deux types de croisements (Figure 2.2) : les croisements compatibles (seed-set moyen > 0,4) et les croisements incompatibles (seed-set moyen < 0,1). Cinq groupes de compatibilité ont été identifiés dont un, le groupe I, correspondant uniquement à l'individu K28 (Tableau 2.1a). Les individus appartenant à la descendance B_a se sont répartis dans les quatre groupes de compatibilité II à V indépendamment de leur phénotype sexuel, hermaphrodite ou mâle-stérile. Pour compléter ces informations, des croisements ont également été réalisés lors de la thèse, avec des individus de même

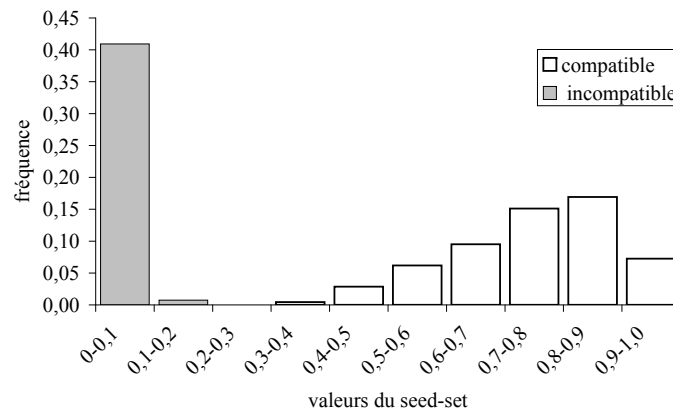


FIGURE 2.2 – Fréquences des valeurs de seed-set. Ces fréquences ont été calculées à partir de 662 capitules obtenus lors des croisements diallèles et factoriels. Les valeurs de seed-set ont été mise en relation avec le caractère compatible ou incompatible de ces croisements.

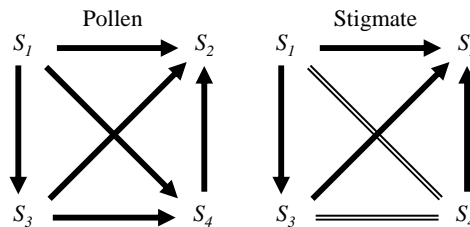


FIGURE 2.3 – Relations de dominance entre les différents allèles S au niveau du pollen et du stigmate. Les flèches indiquent qu'un allèle est dominant sur un autre, les doubles traits indiquent que les deux allèles sont codominants.

génotype que K28 (S_3S_4) mais hermaphrodites et issus de la population D (*cf.* Matériels et méthodes p. 45).

Ces résultats ont permis de définir 1 seul modèle possible à 4 allèles S différents (S_1, S_2, S_3, S_4) avec K59, S_1S_2 , et K28, S_3S_4 . L'ensemble des relations de dominance qui existent entre les allèles S dans le pollen et le stigmate ont également pu être déterminées (Figure 2.3). Les croisements réciproques entre individus hermaphrodites des groupes I et V et les groupes III et V ont montré des comportements différents indiquant qu'il existe des relations de dominance variables dans le pollen et le stigmate pour les groupes I (S_3S_4) et III (S_1S_4). De plus, les résultats tranchés obtenus entre croisements compatibles et incompatibles ont validés l'utilisation de la descendance de K28 et K59 pour réaliser la cartographie du locus S .

TABLEAU 2.1 – Groupes de compatibilité et relations de dominance entre allèles du locus *S*, obtenus à partir des croisements diallèles et factoriels. (a) Les croisements ont été réalisés entre les individus de la population B_a et avec K28 et K59. K28 étant mâle-stérile, le groupe I pollen correspond à des individus S_3S_4 hermaphrodites de la population D, population obtenue à partir du croisement de deux individus de la population B_a . Les cases grisées correspondent aux croisements compatibles, les cases rayées correspondent aux croisements réciproques dont les résultats diffèrent. (b) Choix des génotypes testeurs et stratégie de croisements mis en place pour phénotyper le locus *S* des individus de la population B_a .

(a) Groupes de compatibilité

		Pollen							
		Groupes	I (D)	II (K59)	II	III	IV		
Stigmate	I (K28)	-	+	+	+	-	-	[S3S4]	S_3S_4
	II (K59)	+	-	-	-	+	+	[S1]	S_1S_2
	II	+	-	-	-	+	+	[S1]	S_1S_3
	III	+	-	-	-	+	-	[S1S4]	S_1S_4
	IV	-	+	+	+	-	+	[S3]	S_2S_3
	V	+	+	+	+	+	-	[S4]	S_2S_4
		[S3]	[S1]	[S1]	[S1]	[S3]	[S4]	Phénotype	
		S_3S_4	S_1S_2	S_1S_3	S_1S_4	S_2S_3	S_2S_4		Génotype

(b) Choix des génotypes testeurs et stratégie de croisement

		Testeurs pollen		Génotype		
		Groupes	II			V
			S_1-			S_2S_4
Descendance B_a Stigmate	II	-	+	[S1]	S_1S_3	
	III	-	-	[S1S4]	S_1S_4	
	IV	+	+	[S3]	S_2S_3	
	V	+	-	[S4]	S_2S_4	

2.2.2 Carte moyenne densité du locus *S*

Une fois identifiés les 4 allèles *S* présents dans la population B_a , le génotype au locus *S* a été déterminé à partir de croisements avec des individus testeurs pour 321 des 389 individus disponibles. Du fait de la présence d'individus mâle-stériles dans la population B_a , les individus testeurs ont été utilisés exclusivement comme pollinisateurs. Deux génotypes ont été choisis comme testeurs car permettant d'identifier les 4 groupes de compatibilité avec un minimum de croisements : les génotypes des groupes II et V (Tableau 2.1b).

A partir de ces informations, le locus *S* a pu être cartographié sur la carte K28K59 à une extrémité du groupe de liaison 2 dans un intervalle de près de 10 cM, compris entre

les marqueurs AFLP tacg293 et tgct160 qui coségrègent et le marqueur SSR sw2H09 (Figure 2.4).

En parallèle, le locus *NMS*₁ a été cartographié sur le groupe de liaison 5 (Quillet *et al.*, en préparation). Cela a permis de confirmer l'indépendance du locus *S* et du locus *NMS*₁, le premier étant associé à l'autoincompatibilité et le second à la stérilité mâle nucléaire. De plus, le phénotypage au locus *S* n'a pas été limité par la présence d'individus mâle stériles dans la descendance. En effet, les individus testeurs, utilisés comme donneurs de pollen, ont permis d'identifier le génotype au locus *S* de tous les individus, quel que soit leur phénotype sexuel.

Les 389 individus de la population B_a ont ensuite été génotypés pour tous les marqueurs compris dans l'intervalle tacg293-sw2H09. Cela a permis de restreindre la localisation du locus *S* à l'intervalle tacg293-EU03D06, soit 3,34 cM (Figure 2.4). Dans le même temps, le marqueur AFLP tacg293 a été cloné, séquencé et transformé en marqueur codominant, renommé TACG (*cf.* Matériels et méthodes p. 59).

2.3 Différence des fréquences de recombinaison entre les méioses femelles de K28 et les méioses mâles de K59

Pour les 389 individus de la population B_a, un génotype graphique (Encart 1) a été construit pour l'intervalle TACG-sw2H09. Cette représentation a permis de mettre en évidence un nombre de recombinants significativement différent entre les méioses femelles de K28 et les méioses mâles de K59 ($\chi^2 = 15,92$; $p < 0,0001$). Dans l'intervalle TACG-EU03D06, 25 individus recombinants ont été obtenus avec 3 événements de recombinaison issus de la méiose de K28 et 23 issus de la méiose de K59, ceci correspondant à un rapport M/F (fréquence de recombinaison des méioses mâles/ fréquence de recombinaison des méioses femelles) de 7,7. Plusieurs hypothèses ont été émises pour expliquer cette observation. Cette différence pouvait être due à un fond génétique différent pour les individus parentaux, être spécifique de la région du locus *S* et liée aux différents haplotypes, ou encore être liée au type de méiose considéré, mâle ou femelle.

A partir de cette observation, une comparaison des méioses K28 et K59 a été réalisée sur l'ensemble du génome pour les 174 individus de la population de 2001 à partir de la carte K28K59 (Figure 2.5). Pour chaque groupe de liaison, des marqueurs polymorphes

2.3. Différence des fréquences de recombinaison...

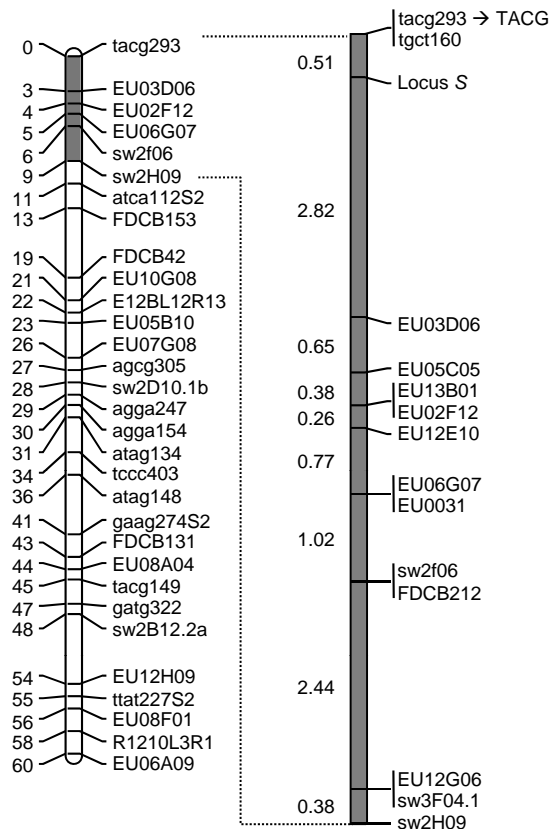
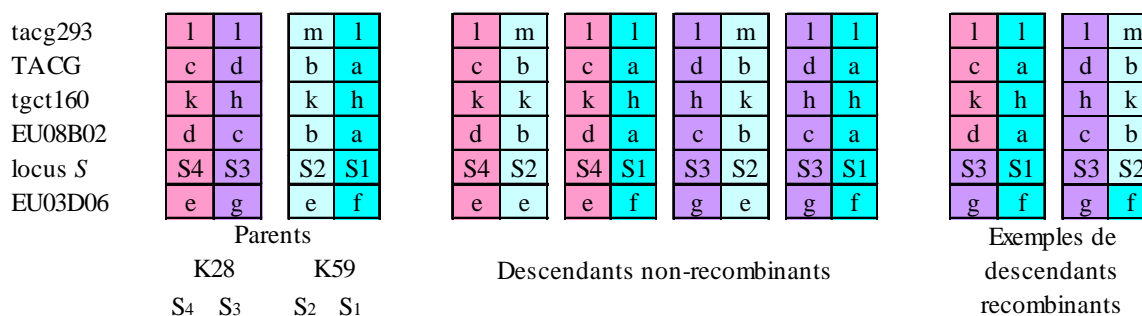


FIGURE 2.4 – Carte génétique moyenne densité, du groupe de liaison 2 à gauche (n = 174) et détail de l'intervalle tacg293-sw2H09 à droite (population B_a, n = 389).

codominants chez les 2 parents et, dans la mesure du possible, distants d'environ 10 cM sur la carte génétique ont été sélectionnés. Une carte K28 et une carte K59 ont été réalisées. Le nombre de recombinants par intervalle a été comparé entre les méioses de K28 (F) et celles de K59 (M) et lorsqu'une différence significative a été observée (test de χ^2), le rapport M/F a été calculé (Figure 2.5). Dans le cas des groupes de liaison 5 et 9, un grand nombre de marqueurs de la carte génétique se sont avérés polymorphes uniquement chez l'un des 2 parents. Ces 2 groupes de liaison n'ont donc pas été pris en compte.

L'étude des groupes de liaison a montré que les rapports M/F n'étaient pas homogènes avec des valeurs allant de 0,26 à 9,33. Pour chaque groupe de liaison, la taille totale de la région commune entre la carte K28 et la carte K59 a été comparée : pour les groupes de liaison 1, 3, 6, 7 et 8, elle était significativement plus grande pour K59, tandis qu'il n'y avait pas de différence significative entre K28 et K59 pour les groupes de liaison 2 et

Un génotype graphique est une représentation schématique d'une région chromosomique donnée : une ligne correspond à un locus et une colonne à un chromosome. Un code couleur est attribué à chacun des chromosomes parentaux et est utilisé pour représenter le génotype de la descendance. Il s'agit d'une représentation de l'ordre des marqueurs les uns par rapport aux autres, indépendamment des distances génétiques. Cette représentation permet de visualiser de façon précise les événements de recombinaison et de les associer à la méiose de l'un ou l'autre des parents. Elle permet également de détecter les erreurs de génotypage : celles-ci génèrent artificiellement des doubles crossing-overs ce qui est très peu probable pour des intervalles de l'ordre du cM.



Pour chaque marqueur utilisé lors de la construction de la carte génétique, la dénomination des allèles a été faite selon la nomenclature du logiciel de cartographie JoinMap, utilisé pour la construction des cartes génétiques.

Quand les 2 parents (K28 et K59) étaient hétérozygotes :

- en présence de quatre allèles, les allèles ont été appelés a, b, c et d ;
- en présence de trois allèles, les allèles ont été appelés e, f et g ;
- en présence de deux allèles, les allèles ont été appelés h et k.

Quand un seul des parents était hétérozygote, ici K59, les allèles ont été appelés l et m.

Encart 1 – Le génotype graphique : description et exemples.

2.3. Différence des fréquences de recombinaison...

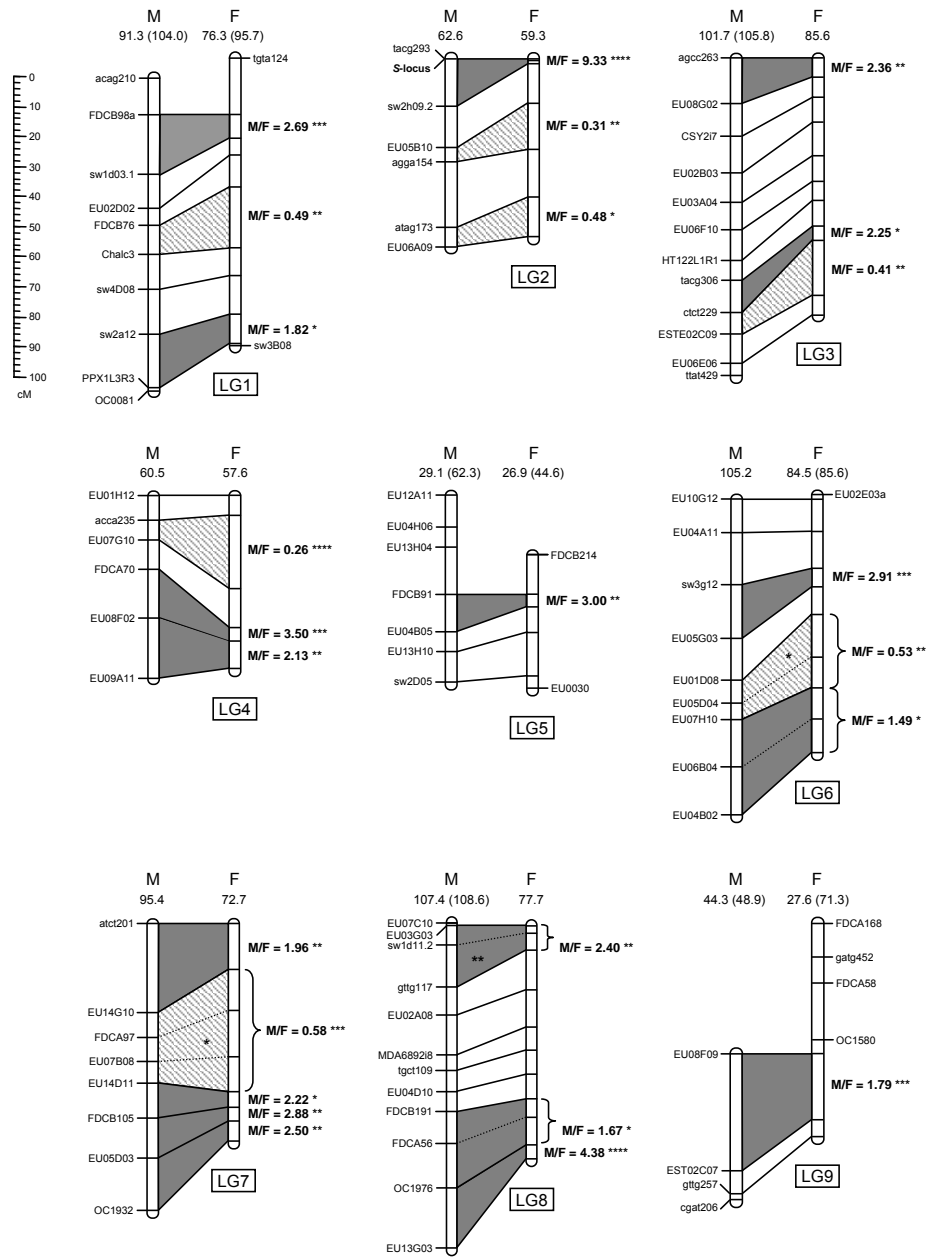


FIGURE 2.5 – Comparaison des fréquences de recombinaison des méioses de K28 et K59 sur les 9 groupes de liaison (LG1 à LG9, population B_a, n = 174). Pour chaque groupe de liaison, des marqueurs polymorphes codominants chez K59 et K28 et distants d'environ 10 cM sur la carte K28K59 ont été sélectionnés. Le nombre de recombinants par intervalle a été comparé entre les méioses de K28 (F) et celles de K59 (M) et lorsqu'une différence significative a été observée (test de χ^2), le rapport M/F a été calculé (M/F > 1, intervalle gris foncé; M/F < 1, intervalle gris clair). La longueur totale estimée à partir des marqueurs communs et de la totalité des marqueurs génotypés (entre parenthèses) ont été indiqués en cM au dessus de chaque groupe de liaison.

4. C'est pourtant sur les groupes de liaison 2 et 4 que les rapports M/F respectivement le plus élevé (9,33) et le plus faible (0,26) ont été observés. Ces résultats suggèrent que, pour ces 2 groupes de liaison, un phénomène de compensation aurait lieu à l'échelle du groupe de liaison entier.

Le rapport M/F le plus élevé ($M/F = 9,33$) de cette étude a été observé dans la région du locus S (intervalle TACG-sw2H09, Figure 2.5), mais d'autres régions du génome ont également montré des différences de fréquence de recombinaison. Si la région du locus S n'est pas la seule à présenter un rapport $M/F > 1$, il s'agit toutefois de la région où la différence de fréquence de recombinaison est la plus importante. Cette observation a donc été prise en compte lors de la démarche de clonage positionnel du locus S , et plus particulièrement lors de la construction des cartes haute densité et haute résolution (*cf.* chapitres 3 et 4).

Chapitre 3

Carte haute densité

La démarche de clonage positionnel nécessite de disposer de marqueurs moléculaires qui encadrent le locus d'intérêt dans une fenêtre qui soit la plus petite possible, en relation avec l'ordre de grandeur des inserts des banques BAC qui est d'une centaine de kb (*cf.* chapitre 6, Gonthier *et al.*, 2010). Si on considère la carte génétique obtenue à partir du croisement K28 x K59, dont la longueur totale estimée est de 930 cM (Cadalen *et al.*, 2010), et la taille approximative du génome qui est d'environ 1 400 Mb (Koopman, 2002), on obtient une distance physique de près de 1,5 Mb/cM pour l'ensemble du génome. Le locus *S* a été cartographié à une extrémité du groupe de liaison 2 dans un intervalle moyen de 3,34 cM (*cf.* figure 2.4 p.81). Cette distance a été précisée pour les méioses femelles (0,77 cM) et les méioses mâles (5,91 cM) (*cf.* figure 3.1). Afin de rendre le projet plus réaliste, il était donc nécessaire de réduire au maximum la fenêtre autour du locus *S*, idéalement à quelques dixièmes de cM. Pour cela, la densité de marqueurs a été augmentée dans l'intervalle TACG-EU03D06. L'obtention de nouveaux marqueurs et leur cartographie à partir de la population B_a (n = 389) ont permis de construire une carte génétique haute densité.

Pour obtenir une carte haute densité, nous avons défini une stratégie permettant de produire plusieurs milliers de marqueurs polymorphes et de sélectionner uniquement ceux compris dans l'intervalle TACG-EU03D06. Cette stratégie a été complétée par les résultats obtenus, en parallèle, lors de l'enrichissement de la carte génétique de la chicorée.

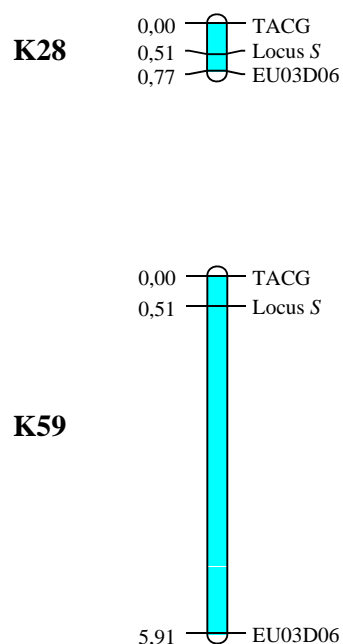


FIGURE 3.1 – Cartes génétiques moyenne densité de la région du locus S (population B_a , $n = 389$), pour les méioses femelles (K28) et les méioses mâles (K59). Les distances génétiques cumulées (en cM) sont indiquées à gauche des groupes de liaison et les marqueurs à droite. La fenêtre d’encadrement du locus S est indiquée en bleu.

3.1 Mise en place de la stratégie d’identification de marqueurs moléculaires liés au locus S :

BSA assistée par marqueurs

3.1.1 Intérêt de la BSA assistée par marqueurs

L’approche BSA (Bulked Segregant Analysis) définie par Michelmore *et al.* (1991) est une méthode permettant d’identifier rapidement des marqueurs dans une région spécifique du génome. Dans une population ségrégeante pour un caractère donné, des mélanges (bulks ou pools) d’ADN sont réalisés (*cf.* Matériels et méthodes p. 51). Chaque bulk est composé de l’ADN d’individus ayant le même phénotype. Par exemple, pour une population F2 obtenue à partir d’un parent résistant et d’un parent sensible à un pathogène, un bulk d’individus résistants et un bulk d’individus sensibles sont réalisés. Les individus qui constituent un pool sont identiques pour le trait choisi mais différents pour tous les autres traits. Le fond génétique des 2 bulks est donc homogène sauf pour la

ou les régions génomiques impliquées dans le trait étudié. L'étude du polymorphisme de marqueurs moléculaires entre ces bulks permet d'identifier les marqueurs génétiquement liés aux locus déterminant le trait choisi pour leur construction.

Cependant, les marqueurs détectés par cette méthode peuvent être localisés à plusieurs cM du locus d'intérêt. Notre objectif était de trouver de nouveaux marqueurs liés au locus S et compris dans la fenêtre encadrée par TACG et EU03D06. Les pools utilisés dans notre approche ont donc été construits à partir d'individus sélectionnés en fonction de leur phénotype et de leur génotype dans la région du locus S , selon une stratégie inspirée de Giovannoni *et al.* (1991) et Jordan *et al.* (2006). Deux populations étaient disponibles au début de ce travail pour la construction des bulks : la population B_a , décrite dans le chapitre 2, et la population C (*cf.* Matériels et méthodes p. 45). Les individus C sont issus d'autofécondation de K59 et de descendants de K59. Bien que ce génotype soit très autoincompatible, la récolte de l'ensemble des capitules de 10 clones de K59 a permis d'obtenir 36 individus (C_a), qui à leur tour ont produit un total de 122 descendants (C_b).

3.1.2 Construction des bulks permettant l'identification de marqueurs liés au locus S

3.1.2.1 Identification de marqueurs liés aux haplotypes S_1 ou S_2

Les individus de la population C ($n = 158$) obtenus à partir d'autofécondations d'individus S_1S_2 , S_1S_1 et S_2S_2 (*cf.* Matériels et méthodes p. 45) ont tous été génotypés pour les marqueurs TACG et EU03D06. Cela a permis d'inférer le génotype de ces individus au locus S (*cf.* Matériels et méthodes p. 70) et de distinguer les homozygotes S_1S_1 ou S_2S_2 des individus hétérozygotes S_1S_2 . Deux types de pools (*cf.* Matériels et méthodes p. 51) ont été construits à partir d'individus homozygotes S_1S_1 (P1 : $n = 8$; P2 : $n = 8$) et S_2S_2 (P3 : $n = 15$; P4 : $n = 15$). Les marqueurs spécifiquement liés à l'haplotype S_1 devaient être présents dans les bulks P1 et P2 et absents dans les bulks P3 et P4 et inversement pour ceux liés à S_2 (Figure 3.2). Cependant, dans la mesure où la région homozygote autour du locus S pouvait s'étendre au delà des marqueurs TACG et EU03D06, les marqueurs liés à S_1 ou S_2 et identifiés grâce au bulks P1 à P4 pouvaient se trouver hors de cet intervalle.

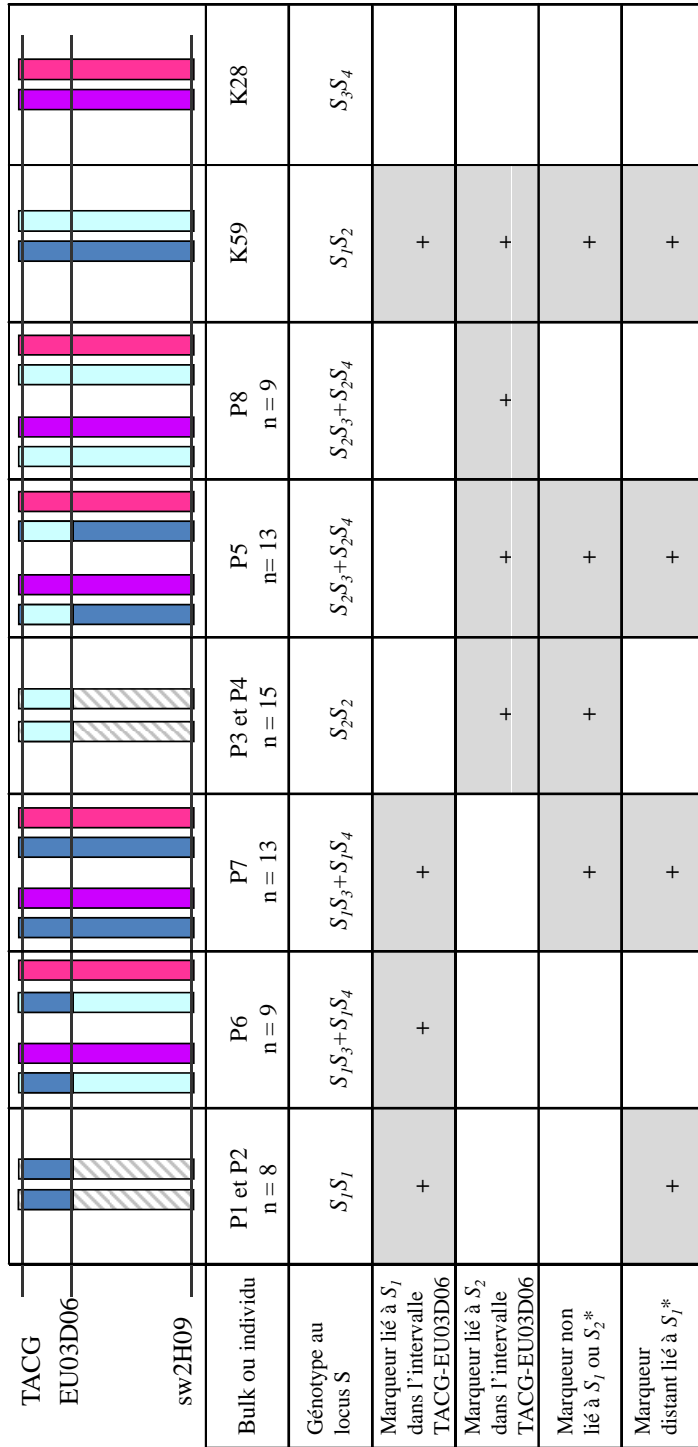


FIGURE 3.2 – Stratégie des bulks. Représentation du génotype graphique des individus constituant les différents bulks et des résultats attendus en fonction de la liaison des marqueurs aux haplotypes S_1 ou S_2 . + indique la présence d'un marqueur. Chaque couleur correspond à un chromosome parental. Pour les bulks d'individus S_1S_1 et S_2S_2 , les parties rayées correspondent aux régions non génotypées. Les situations où des marqueurs liés à l'haplotype S_1 ou S_2 ont été représentés ainsi que 2 exemples de résultats pouvant être obtenus (*).

3.1.2.2 Sélection des marqueurs liés aux haplotypes S_1 ou S_2 compris dans l'intervalle TACG-EU03D06

Pour distinguer les marqueurs spécifiques de S_1 ou S_2 , compris dans l'intervalle TACG-EU03D06 de ceux hors de cet intervalle, des bulks supplémentaires ont été construits. Pour ces pools d'ADN, le génotype des individus devait être connu pour un large intervalle autour du locus S . La population B_a , pour laquelle le génotype des individus est connu dans l'intervalle TACG-sw2H09 (*cf.* chapitre 2), a été utilisée et 2 catégories de bulks ont été construites (Figure 3.2). Dans la première, les individus utilisés devaient posséder un chromosome issu de K59 sur lequel un événement de recombinaison avait eu lieu entre EU03D06 et sw2H09. Il s'agissait :

- soit d'individus S_2S_3 et S_2S_4 (P5 : $n = 13$),
- soit d'individus S_1S_3 et S_1S_4 (P6 : $n = 9$).

Dans la seconde catégorie, les individus devaient posséder un chromosome issu de K59, sans aucun événement de recombinaison dans l'intervalle TACG-sw2h09. Il s'agissait alors :

- soit d'individus S_1S_3 et S_1S_4 (P7 : $n = 13$),
- soit d'individus S_2S_3 et S_2S_4 (P8 : $n = 9$).

Les 8 bulks construits ainsi que les individus K59 (S_1S_2) et K28 (S_3S_4) ont été utilisés afin d'identifier avec précision des marqueurs liés à l'haplotype S_1 ou S_2 dans l'intervalle TACG- EU03D06 (Figure 3.2). Les marqueurs liés spécifiquement à l'haplotype S_1 devaient être présents uniquement dans les bulks P1, P2, P6, P7 et chez K59. Les marqueurs liés spécifiquement à l'haplotype S_2 devaient être présents uniquement dans les bulks P3, P4, P5, P8 et chez K59.

3.2 Recherche de nouveaux marqueurs liés au locus S

3.2.1 Criblage AFLP des bulks

Pour produire de nouveaux marqueurs, nous avons choisi une méthode permettant de générer un grand nombre de marqueurs polymorphes sans avoir d'information *a priori* sur la séquence. Il s'agit de la technique AFLP (Vos *et al.*, 1995, *cf.* Matériels et méthodes p. 55.). Cette technique permet de produire plusieurs milliers de fragments avec un

nombre limité de couples d'amorces. Les 256 combinaisons possibles d'amorces AFLP EcoRI-A+2/MseI-C+2 ont été testées sur les 8 bulks et les individus K59 et K28. En moyenne, 120 fragments ont été obtenus par couple d'amorces, ce qui correspond à un total d'environ 31 000 fragments AFLP générés. Parmi ces fragments, 2 350 se sont avérés polymorphes entre K28 et K59, 136 fragments étaient spécifiques des bulks S_1S_1 (P1 et P2) et 255 des bulks S_2S_2 (P3 et P4). Sur les 391 fragments potentiellement liés aux haplotypes S_1 ou S_2 , environ 80 n'étaient pas présents chez K59. Soixante se sont avérés également présents dans les bulks P5 à P8, ainsi que chez K59 et K28.

Au total, 14 marqueurs AFLP ont été identifiés comme se trouvant potentiellement dans l'intervalle TACG-EU03D06 : 13 nouveaux marqueurs liés soit à S_1 soit à S_2 (Tableau 3.1, Figure 3.3), ainsi que tacg293 cartographié préalablement (*cf.* chapitre 2) et retrouvé lié à l'allèle S_1 .

TABLEAU 3.1 – Caractéristiques des marqueurs AFLP identifiés avec la stratégie des bulks et informations relatives à leur cartographie au voisinage du locus S .

Couple d'amorces EcoRI-A+2/MseI-C+2	Taille du fragment (pb)	Couplage avec l'haplotype	Nom du marqueur	Cartographie dans la population B_a
EcoRI-AGG/MseICAT	125	S_2	ggat125	Colocalisation avec TACG
EcoRI-ATC/MseICAA	300	S_2	tcaa300	Colocalisation avec le locus S
EcoRI-ATT/MseICAA	440	S_2	ttaa440	
EcoRI-AAA/MseICCA	83	S_1	aaca83	Entre le locus S et EU03D06
EcoRI-AAA/MseICCC	134	S_2	aacc134	
EcoRI-ACC/MseICGT	487	S_1	ccgt487	
EcoRI-AGC/MseICAG	172	S_2	gcag172	En ségrégation mais
EcoRI-ATA/MseICCT	98	S_2	tact98	en dehors de l'intervalle
EcoRI-ACC/MseICGG	304	S_1	ccgg304	TACG-EU03D06
EcoRI-ACC/MseICGA	312	S_2	ccga312	
EcoRI-AAT/MseICAA	295	S_1	ataa295	
EcoRI-AAT/MseICCA	82	S_2	atca82	Non amplifiés
EcoRI-ACC/MseICTA	448	S_1	ccta448	chez les recombinants

3.2.2 Cartographie et validation des nouveaux marqueurs AFLP

Une fois ces nouveaux marqueurs identifiés, l'étape suivante consistait à valider leur localisation à proximité du locus S . Les 13 marqueurs AFLP ont donc été génotypés chez les 25 individus recombinants entre TACG et EU03D06 de la population B_a ($n = 389$).

3.2. Recherche de nouveaux marqueurs liés au locus S

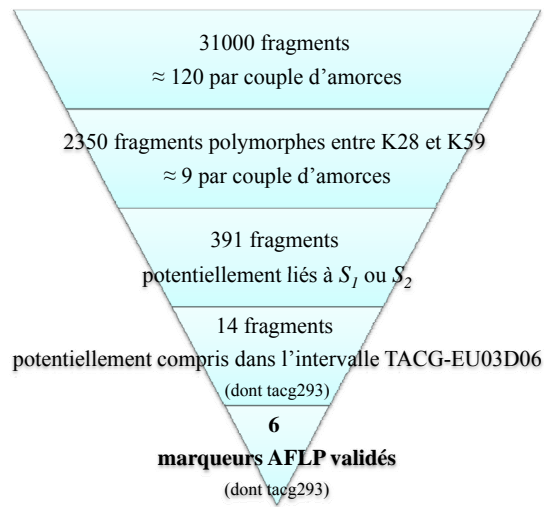


FIGURE 3.3 – Résultats successifs obtenus dans le cadre de la recherche de nouveaux marqueurs AFLP par BSA.

Le génotypage de ces seuls 25 individus a permis de cartographier les marqueurs compris dans l'intervalle TACG-EU03D06 pour toute la population B_a , tout en limitant le nombre de réactions PCR. Le génotype des parents K28 et K59 étant connu pour ces marqueurs, les génotypes obtenus ont été placés dans le génotype graphique des 25 recombinants (Figure 3.4). Parmi les 13 marqueurs identifiés, 5 ont été validés comme compris dans l'intervalle TACG-EU03D06 (Figure 3.4), soit une densité de 1,8 marqueur/cM. Deux marqueurs n'ont pas été amplifiés à partir de l'ADN des recombinants et 6 n'ont pu être cartographiés dans la région du locus S , c'est-à-dire que les génotypes obtenus pour ces marqueurs ne concordaient pas avec le génotype graphique (Tableau 3.1, Figure 3.4).

Les 5 marqueurs validés grâce au génotype graphique (Figure 3.4) ont ensuite été placés sur la carte génétique. Étant des marqueurs spécifiques de S_1 ou S_2 et à l'état homozygote chez K28, le calcul des distances génétiques s'est fait uniquement à partir des méioses de K59. Ces 5 marqueurs ont été positionnés sur la carte génétique (Figure 3.5) :

- tcaa300 colocalisant avec le locus S ,
- ggat125 colocalisant avec TACG,
- ttaa440 et aaca83 colocalisants à 0,77 cM du locus S ,
- aacc134 à 1,02 cM du locus S .

Ainsi, après avoir cartographié ces 5 marqueurs AFLP, la fenêtre autour du locus S a été réduite de 5,91 à 1,29 cM pour les méioses mâles de K59. La stratégie BSA assistée par marqueurs associée aux validations sur les individus recombinants s'est donc révélée

efficace et a permis l'identification de marqueurs réellement au voisinage du locus *S*.

3.2.3 Identification de nouveaux marqueurs microsatellites

Parallèlement à l'approche BSA assistée par marqueurs associée au criblage AFLP, l'enrichissement de la carte génétique K28K59 en marqueurs microsatellites s'est poursuivi dans le cadre du GIS CARTOCHIC (Cadalen *et al.*, 2010). Deux nouveaux marqueurs microsatellites ont été cartographiés dans l'intervalle TACG-EU03D06 (Figures 3.4 et 3.5) : EU08B02 (4 allèles) et EU12G02 (3 allèles). Ces 2 marqueurs étant hétérozygotes chez K28 et K59, ils ont été placés sur les cartes de chacun des parents. Dans le cas de la carte K28, EU08B02 et EU12G02 coségrégeaient respectivement avec TACG et le locus *S* et la fenêtre autour du locus *S* restait de 0,77 cM. Dans le cas de la carte K59, EU08B02 a été placé à 0,25 cM et EU12G02 à 2,83 cM du locus *S*. En prenant en compte les marqueurs AFLP, la fenêtre autour du locus *S* a donc été réduite de 1,29 à 1,02 cM pour la carte K59.

3.3 Transformation des marqueurs AFLP en marqueurs SCAR

3.3.1 Intérêt de la démarche

La technique AFLP permet de produire un grand nombre de marqueurs polymorphes mais sa mise en œuvre nécessite plusieurs étapes, ce qui n'est pas idéal dans le cas du génotypage d'un grand nombre d'individus, lors de la construction d'une carte haute résolution (*cf.* chapitre 4). De plus, les marqueurs obtenus sont des marqueurs dominants qui offrent donc une information limitée en comparaison de marqueurs codominants (*cf.* chapitre 4). Enfin, il n'est pas possible de produire les sondes nécessaires au crible d'une banque BAC (*cf.* chapitre 6) à partir de marqueurs AFLP, car l'utilisation d'un couple d'amorces donné produit un profil multibande. En raison de ces aspects méthodologiques liés aux marqueurs AFLP, nous avons voulu transformer les 5 nouveaux marqueurs (ggat125, tcaa300, ttaa440, aaca83 et aacc134) en marqueurs spécifiques de locus (marqueurs SCAR, Matériels et méthodes p. 59) .

3.3. Transformation des marqueurs AFLP en marqueurs SCAR

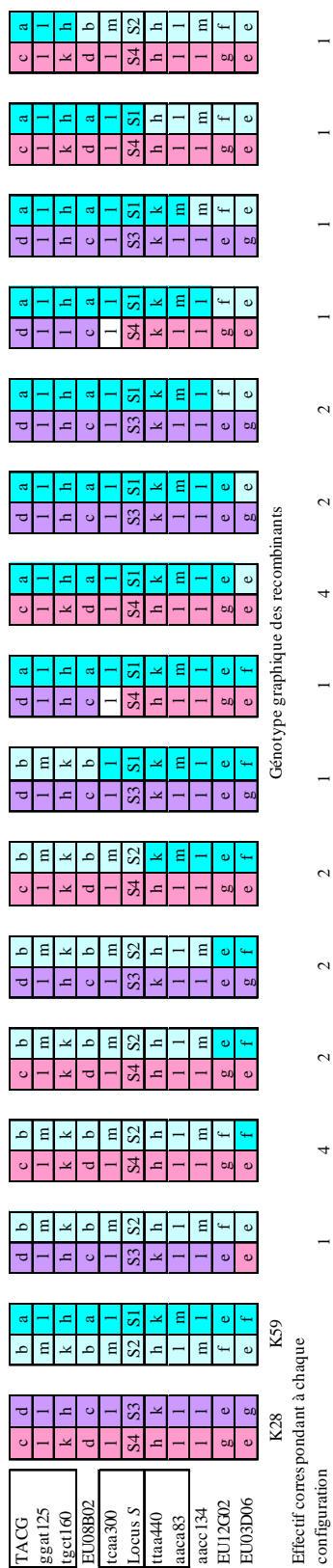


FIGURE 3.4 – Cartographie génétique des marqueurs obtenus par BSA dans l'intervalle TACG-EU03D06, à l'aide du génotype graphique des 25 individus recombinants de la population B_a. Le génotype graphique des parents (K28 et K59) et celui correspondant aux différentes configurations obtenues pour les individus recombinants sont représentées. Les marqueurs sont indiqués à gauche. Ceux qui coségrègent sont regroupés dans un même cadre. * La position du marqueur tcaa300, au dessus ou au dessous du locus S, n'a pas pu être déterminée, du fait du manque d'information apportée par ce marqueur.

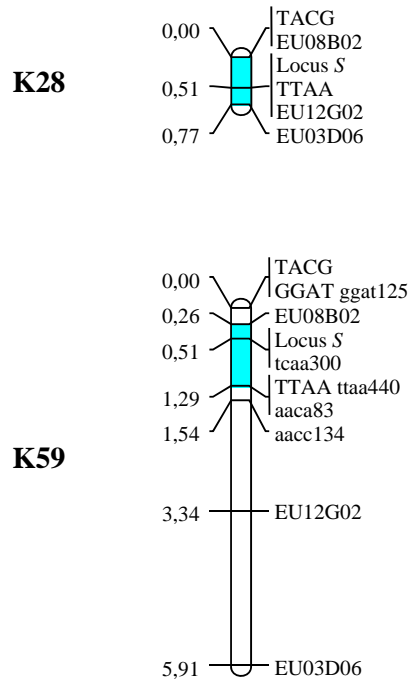


FIGURE 3.5 – Cartes génétiques haute densité de la région du locus *S* (population B_a, n = 389), pour K28 (méioses femelles) et K59 (méioses mâles). Tous les nouveaux marqueurs sont cartographiés. les marqueurs AFLP sont indiqués en minuscules et les marqueurs SCAR correspondants en majuscules. Les marqueurs microsatellites sont indiqués en majuscules. Les distances génétiques cumulées exprimées en cM sont indiquées à gauche des groupes de liaison. La fenêtre d’encadrement du locus *S* est indiquée en bleu.

3.3.2 Simplification des profils AFLP par miniséquençage

La première étape nécessaire à l’obtention de marqueurs SCAR consiste à simplifier les profils AFLP par ajout de 3 bases sélectives supplémentaires à l’amorce MseI-C+2 (*cf.* Matériels et méthodes p. 59). La migration des profils AFLP simplifiés se déroule sur gel de polyacrylamide afin de pouvoir ensuite exciser le fragment d’intérêt.

Cette démarche de simplification a été entreprise pour les profils AFLP associés aux marqueurs ggat125, tcaa300, ttaa440, et aacc134, mais pas pour celui associé au marqueur aaca83. En effet, la taille de ce marqueur (83 pb) correspond à une séquence d’une quarantaine de bases une fois les amorces AFLP soustraites, ce qui n’est pas suffisant pour définir de nouvelles amorces spécifiques (*cf.* 1.2.2).

Trois bases sélectives supplémentaires ont été identifiées sans ambiguïté pour les marqueurs ggat125, ttaa440 et aacc134 et une amorce MseI-C+5 a été définie pour chacun

3.3. Transformation des marqueurs AFLP en marqueurs SCAR

d'eux afin de simplifier les profils AFLP (Tableau 3.2). Pour le marqueur tcaa300, seules les première et troisième bases sélectives ont été identifiées avec certitude. Une amorce MseI-C+3 a donc été utilisée. Moins sélective qu'une amorce MseI-C+5, elle a toutefois permis de simplifier le profil associé à tcaa300. Les fragments correspondant à chacun de ces 4 marqueurs ont pu être repérés et excisés des gels de polyacrylamide.

Lors de la simplification du profil associé à ggat125, une bande de 128 pb a été observée chez les individus S_1 , en parallèle de la bande de 125 pb présente chez les individus S_2 (Figure 3.6). Les 25 individus recombinants de la population B_a ont été génotypés avec l'amorce MseI-C+5 correspondante, ce qui a révélé que cette bande était liée aux allèles S_1 , S_3 et S_4 . Les marqueurs AFLP ggat125 et ggat128 semblaient être en réalité 2 allèles d'un même marqueur présentant un polymorphisme dû à une insertion-délétion de 3 pb. Le fragment correspondant à ggat128 a donc également été excisé du gel de polyacrylamide.

3.3.3 Spécificité des séquences correspondant aux fragments AFLP

Les fragments AFLP excisés ont été clonés et séquencés (*cf.* Matériels et méthodes p. 59). Une séquence a été obtenue pour chacun des fragments (tcaa300, ttaa440, aacc134, ggat125 et ggat128). Les fragments ggat125 et ggat128 se sont effectivement différenciés par une insertion-délétion de 3 pb. Aucune homologie n'a été trouvée dans les bases de données pour les séquences obtenues à partir de tcaa300 et ggat125/128. Une homologie a été trouvée entre la séquence de aacc134 et un EST d'*Helianthus annuus* (HuEL482717, e-value 1,9e-9, 86% d'identité sur 72 pb), ainsi qu'entre ttaa440 et un EST de *Lactuca sativa* (DW167729, e-value 2e-12, 90% d'identité sur 59 pb). En revanche, aucune fonction de gène n'a pu être associée à ces EST.

TABLEAU 3.2 – Résultats du miniséquençage

Marqueur AFLP	Bases sélectives supplémentaires identifiées			Amorce MseI
	4ème	5ème	6ème	
ggat125	G	C	C	MseI-CATGCC
aacc134	C	C	A	MseI-CCCCCA
tcaa300	T	C ?	A	MseI-CAAT
ttaa440	C	A	G	MseI-CAACAG

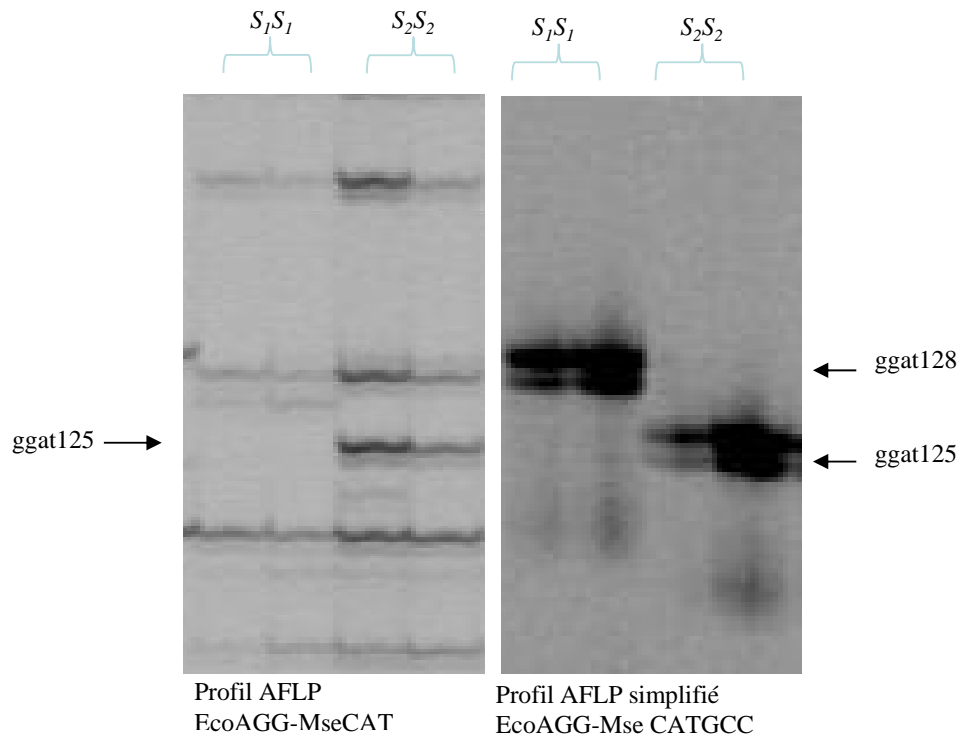


FIGURE 3.6 – Exemple de simplification de profil AFLP obtenue à partir du miniséquençage du marqueur *ggat125*. Les 2 premières pistes de migration correspondent aux profils d'individus homozygotes S_1S_1 et les deux secondes aux profils d'individus homozygotes S_2S_2 . La simplification a mis en évidence le marqueur *ggat128* spécifique du génotype S_1S_1 .

Aucune des séquences obtenues ne présentant d'homologie avec des séquences répétées répertoriées, elle ont toutes été utilisées pour la définition de nouvelles amorces.

3.3.4 Définition de nouvelles amorces et recherche de polymorphisme

Des couples d'amorces spécifiques ont été définis en fonction des séquences obtenues pour les 4 marqueurs AFLP *tcaa300*, *ttaa440*, *aacc134* et *ggat125/128*, et les marqueurs correspondants ont été respectivement nommés TCAA, TTAA, AACC et GGAT. Ces amorces ont ensuite été testées chez des individus au génotype connu dans la région du locus *S*, afin d'associer les éventuels allèles obtenus à un haplotype *S*. Chacun des produits PCR obtenus a été analysé en recherchant du polymorphisme de taille et/ou de conformation (*cf.* Matériels et méthodes p. 51).

Le marqueur AACC s'est avéré monomorphe. Un profil multibande a été obtenu pour le

3.3. Transformation des marqueurs AFLP en marqueurs SCAR

marqueur TCAA mais aucun polymorphisme associé aux haplotypes S n'a été détecté. Du polymorphisme de conformation (SSCP) a été observé pour le marqueur TTAA avec 4 allèles associés aux différents haplotypes S . Les 2 allèles correspondant au polymorphisme de taille précédemment observé pour le marqueur AFLP *ggat125/128* ont été conservés pour le marqueur SCAR GGAT, l'un associé à l'haplotype S_2 , l'autre aux haplotypes S_1 , S_3 et S_4 (Tableau 3.3).

Sur les 4 marqueurs SCAR définis, seuls TTAA et GGAT ont pu être utilisés en tant que marqueurs codominants dans le cadre de la carte haute résolution (*cf.* chapitre 4). Cependant, la transformation en marqueur SCAR n'a apporté un supplément d'information que dans le cas de TTAA. Il a donc également pu être cartographié sur la carte K28 (Figure 3.5). Les marqueurs AACC et TCAA se sont révélés, quant à eux, moins informatifs que les marqueurs AFLP. Le polymorphisme, dû à au moins un des sites de restriction (*EcoRI* ou *MseI*) ou aux bases sélectives, a été perdu après la définition d'amorces internes au fragment et aucun polymorphisme ni de taille, ni de séquence, n'a pu être identifié. De plus, le profil multibande obtenu pour TCAA indique que la séquence obtenue à partir de *tcaa300* doit être répétée dans le génome, malgré l'absence d'homologie dans les bases de données actuellement disponibles.

Compte tenu de ces résultats, ce sont les marqueurs AFLP (*aacc134* et *tcaa300*) qui ont été cartographiés lors de la construction de la carte haute résolution (*cf.* chapitre 4).

TABLEAU 3.3 – Comparaison du polymorphisme des marqueurs AFLP et de leurs équivalents transformés en marqueurs SCAR.

Marqueur AFLP	Polymorphisme AFLP	Marqueur SCAR	Nombre d'allèles	Polymorphisme SCAR
<i>ggat125/1128</i>	InDel	GGAT	2	InDel
<i>tcaa300</i>	présence/absence	TCAA	profil multibande	-
<i>ttaa440</i>	présence/absence	TTAA	4	SNP
<i>aacc134</i>	présence/absence	AACC	1	non

Chapitre 4

Carte haute résolution

Dans le cadre d'une démarche de clonage positionnel, une fois la région d'intérêt densifiée en nouveaux marqueurs, leur position doit être déterminée avec précision. Comme la fenêtre autour du locus S était déjà de l'ordre de quelques cM (entre 0,77 et 5,91 cM), la distance génétique attendue entre ces nouveaux marqueurs pouvait être de l'ordre du dixième de cM. Du fait de ces faibles distances génétiques, plusieurs marqueurs identifiés lors de la construction de la carte haute densité coségrégeaient avec d'autres marqueurs. Par conséquent, il était nécessaire de disposer d'une population de taille suffisante pour obtenir un nombre plus important de recombinants dans la région du locus S , ceci afin de positionner les nouveaux marqueurs et d'augmenter la résolution de la carte génétique.

4.1 Détermination du nombre d'individus à cartographier

Du fait du caractère autoincompatible de notre matériel végétal, les croisements ont été réalisés entre 2 individus hétérozygotes compatibles. La quantité d'information contenue par chaque individu (i_r , Allard, 1956) dépend de la fréquence de recombinaison (r) et du type de configuration des marqueurs dont on dispose (Figure 4.1, *cf.* Matériels et méthodes p. 67). Si les marqueurs encadrant le locus S sont à l'état hétérozygote chez les 2 parents, alors on dispose d'autant d'information que pour une F2. S'il s'agit de marqueurs codominants, i_r est égal à celui d'une F2 complètement classée (CC), c'est-à-dire que les marqueurs sont informatifs pour les méioses mâles et femelles. S'il s'agit de marqueurs dominants, i_r est égal à celui d'une F2 dont les marqueurs sont dominants

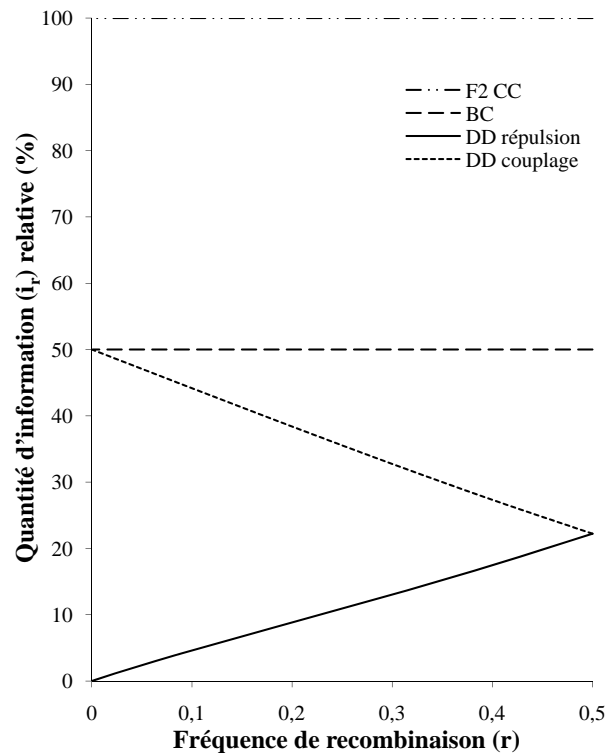


FIGURE 4.1 – Quantité d'information portée par individu (i_r) en fonction de r et exprimée proportionnellement à une F2 complètement classée (100%, F2 CC). BC correspond à un backcross, DD couplage correspond à une F2 pour 2 marqueurs dominants en couplage et DD répulsion à une F2 pour 2 marqueurs dominants en répulsion.

(DD), soit en couplage, soit en répulsion. Enfin, si les marqueurs sont hétérozygotes chez un seul des parents, on dispose alors d'autant d'information que pour un back-cross (BC). Une configuration BC apporte moitié moins d'information qu'une configuration CC. Une configuration DD en couplage apporte presque autant d'information qu'un back-cross pour les petites valeurs de r , mais devient moins informative à mesure que r augmente. Quant à la configuration DD en répulsion, c'est la moins informative pour les petites valeurs de r .

Le calcul de l'écart type associé à la fréquence de recombinaison tient compte de i_r et de l'effectif total (n) utilisé pour la cartographie (*cf.* Matériels et méthodes p. 70, Allard, 1956 ; Liu, 1998) :

$$S_r = \sqrt{\frac{1}{ni_r}}$$

4.1. Détermination du nombre d'individus à cartographier

En considérant que les estimations de r sont distribuées selon une loi normale, l'intervalle de confiance se calcule alors de la manière suivante (*cf.* Matériels et méthodes p. 70, Liu, 1998) :

$$c = r \pm z_{\frac{\alpha}{2}} S_r$$

Lorsque l'on souhaite cartographier plus de 2 marqueurs adjacents, l'ordre de ces marqueurs et la distance entre eux peuvent être considérés comme statistiquement fiables si les intervalles de confiance calculés pour chaque valeur de r ne se chevauchent pas ($z_{\frac{\alpha}{2}} S_r \leq \frac{r}{2}$, *cf.* Matériels et méthodes p. 70).

Pour estimer le nombre d'individus (n) à cartographier, il est donc nécessaire de prendre en compte différents paramètres : la fréquence de recombinaison (r) entre 2 marqueurs adjacents et l'intervalle de confiance associé, ainsi que l'information (i_r) apportée par chaque type de marqueurs. L'effectif n a donc été calculé pour différentes valeurs de i_r (CC et DD couplage) et de α (0,05 et 0,01) tel que (Figure 4.2, *cf.* 1.3.3.2) :

$$n \geq \frac{(2z_{\frac{\alpha}{2}})^2}{i_r r^2}$$

Par exemple, pour $\alpha = 0,05$ et $r = 0,01$ on obtient :

- $n \geq 761$ dans le cas d'une configuration CC,
- $n \geq 1\,539$ dans le cas de 2 marqueurs dominants en couplage,
- $n \geq 153\,626$ dans le cas de 2 marqueurs dominants en répulsion.

A l'inverse, si on a 1 500 individus, $\alpha = 0,05$ et une configuration CC, on peut obtenir un ordre précis des marqueurs jusqu'à une valeur de $r = 0,0051$. Ces résultats montrent l'importance de disposer de marqueurs codominants lorsqu'une carte haute résolution doit être construite.

A partir de ces résultats et n'ayant pas d'information *a priori* sur les marqueurs que nous allons identifier grâce à la carte haute densité, nous avons choisi de baser notre carte haute résolution sur une population d'au moins 2 800 individus, soit 5 600 méioses informatives dans le cas d'une configuration CC. Cela représentait à la fois un effectif réaliste du point de vue technique et méthodologique et la possibilité d'obtenir une fourchette allant de $r = 0,0095$ (configuration DD couplage, $\alpha = 0,01$) à $r = 0,0027$ (configuration CC, $\alpha = 0,05$) (Figure 4.2). Le cas de marqueurs dominants en répulsion n'a pas été pris en compte car trop peu informatif pour des valeurs de $r < 0,01$ et, de ce fait, nécessitant un effectif trop élevé pour être mis en œuvre.

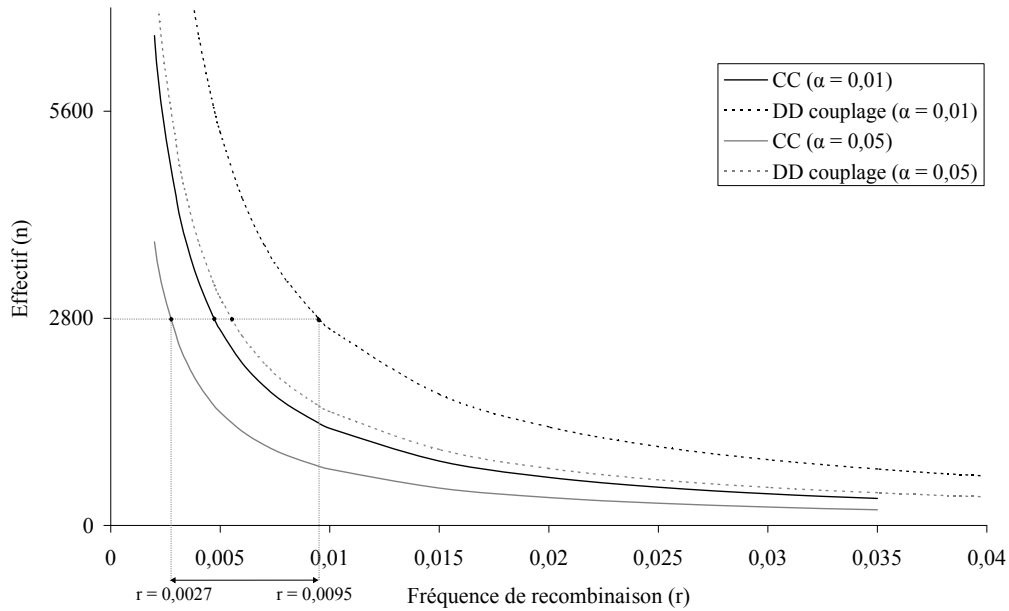


FIGURE 4.2 – Effectif minimum nécessaire pour estimer l'intervalle de confiance $r \pm z_{\frac{\alpha}{2}} S_r$ (avec $z_{\frac{\alpha}{2}} S_r < \frac{r}{2}$) dans le cas d'une classification complète (CC) et dans le cas de marqueurs dominants en couplage (DD couplage) pour $\alpha = 0,01$ et $\alpha = 0,05$. Les points (•) indiquent les valeurs de r pour $n = 2800$ en fonction de α et de la configuration des marqueurs.

4.2 Sélection des populations pour l'établissement de la carte haute résolution

4.2.1 Caractéristiques des populations étudiées

Nous disposons de 389 individus issus de croisements de K28 avec K59 (population B_a) utilisés pour réaliser la carte moyenne densité et de 318 graines supplémentaires (population B_b) également issues de croisements K28 x K59. Cette population B devait donc être complétée pour atteindre l'effectif ciblé (2800). Nous disposons d'une importante banque de graines issues de croisements entre individus B ou entre individus B et K59. De plus, au moment de compléter la population B, nous disposons de 2 marqueurs codominants dans la région du locus *S* : TACG et EU03D06 (*cf.* chapitre 2, Figure 4.3). Dans cette population, le polymorphisme de ces marqueurs correspondait à une configuration complètement classée (CC). Le choix des croisements supplémentaires a été basé sur la présence d'une configuration CC pour ces 2 marqueurs. Ces croisements ont également été sélectionnés afin de pouvoir tester les différentes hypothèses

4.2. Sélection des populations pour l'établissement de la carte haute résolution

pouvant expliquer la différence de fréquence de recombinaison entre les méioses de K28 et K59 (*cf.* paragraphe 4.4). En plus de la population B, les populations suivantes ont été sélectionnées (Figure 4.3) :

- la population D issue du croisement B251 x B266,
- les populations E, F et G issues du croisement de B272 avec 3 pollinisateurs de même génotype pour les marqueurs TACG et EU03D06, respectivement B240, B253 et B337,
- la population H issue du croisement de 17 individus B (groupe B_H) avec K59,
- la population M issue du croisement de K59 avec 7 individus B (groupe B_M).

Ainsi, nous avons atteint un total de 2 563 graines. Suite aux étapes de germination et de repiquage de ces populations, nous avons obtenu 2 496 individus en supplément des 389 dont nous disposions déjà, soit un total de 2 885 individus.

4.2.2 Informations apportées par les nouveaux marqueurs

Les nouveaux marqueurs tcaa300, TTAA, aaca83, aacc134 et EU12G02 ont été identifiés ultérieurement au choix des populations. Le génotypage des parents (Figure 4.3) a révélé que, excepté TTAA et EU08B02, les nouveaux marqueurs n'étaient pas hétérozygotes chez tous les parents. Parmi les marqueurs codominants, le marqueur EU12G02 était à l'état homozygote chez les individus B266, B240, B253 et B337. Parmi les marqueurs dominants, il a été constaté que :

- K28 était homozygote pour les marqueurs tcaa300, aaca83 et aacc134,
- B266, B240, B253 et B337 étaient homozygotes pour les marqueurs tcaa300 et aaca134,
- B251, B_H et B_M étaient homozygotes pour le marqueur aaca83,
- B272 était homozygote pour le marqueur aacc134.

Ces marqueurs à l'état homozygote n'étaient pas informatifs pour les individus considérés et n'ont donc pas été représentés sur les cartes génétiques correspondant à chacun des parents. De plus, pour les descendants dont les 2 parents étaient hétérozygotes pour un marqueur dominant donné, il n'a pas été possible de distinguer un génotype hétérozygote lm d'un génotype homozygote mm , m étant l'allèle dominant. Cela concernait les populations E, F et G pour le marqueur aaca83 et les populations H et M pour les marqueurs aacc134 et tcaa300. Dans certains cas, la réalisation du génotype graphique des descendants de ces populations a permis d'inférer leur génotype pour ces marqueurs à partir

Population	B	D	E, F, G	H	M
Parents (♀ x ♂)	K28 x K59	B251 x B266	B272 x B253 B337	B _H x K59	K59 x B _M
Nombre d'individus	642	317	787 (213/291/283)	506	572
Marqueurs:					
TACG	c d	b d	b d	b c	b a
GGAT	k k	h k	h k	h k	h k
tgct160	k h	k h	k h	k h	k h
EU08B02	d c	b d	b d	b c	b c
tcaa300*	l l	m l	m l	m l	m l
Locus S	S4 S3	S2 S4	S2 S4	S2 S3	S2 S1
TTAA	c d	b c	a c	b a	b a
aaca83*	l l	l l	m l	l m	l m
aacc134*	l l	l l	l l	m l	m l
EU12G02	g e	e g	e g	f e	f e
EU03D06	e g	f e	f e	e f	e f

FIGURE 4.3 – Génotypes graphiques des parents des populations utilisés pour la carte haute résolution pour l'intervalle TACG-EU03D06. Les 3 marqueurs moléculaires soulignés ainsi que le génotype au locus *S* sont ceux dont nous disposons pour la sélection des différentes populations, les autres marqueurs moléculaires étant ceux obtenus lors de l'établissement de la carte haute densité. Les cases blanches du génotype graphique indiquent que le marqueur n'est pas informatif pour l'individu concerné. Dans la descendance des croisements présentant des configurations lm x lm pour les marqueurs (*), les génotypes mm et lm n'ont pas pu être différenciés, m étant un allèle dominant.

des marqueurs adjacents. Cependant, en cas de recombinaison dans cet intervalle, le génotype pour ces marqueurs n'a pu être déterminé. Ils se sont donc avérés non informatifs pour ces populations et n'ont pas été représentés sur la carte génétique des parents de ces populations (*cf.* paragraphe 4.4).

4.3 Criblage séquentiel des recombinants

4.3.1 Stratégie de génotypage en 3 étapes

La réalisation des cartes génétiques a nécessité plusieurs étapes en amont. Tout d'abord, l'ADN des 2 496 nouveaux individus a été extrait (*cf.* Matériels et méthodes p. 50) puis le génotypage de la région du locus S s'est déroulé en 3 phases distinctes afin de limiter le nombre de réactions PCR et de réduire le nombre d'individus à phénotyper pour le locus S (Figure 4.4).

Dans un premier temps, l'ensemble des individus a été génotypé pour les 2 marqueurs qui encadrent la région du locus S : TACG et EU03D06. En considérant les associations alléliques des chromosomes parentaux, ce premier génotypage a permis de distinguer 2 291 individus non-recombinants et 144 individus recombinants dans cet intervalle. De plus, 61 individus hors types, c'est-à-dire des individus ayant un génotype incohérent par rapport à celui des parents (contamination pollinique lors des croisements ou contamination des échantillons), ont été éliminés de nos effectifs. Une fois connu le génotype pour les marqueurs TACG et EU03D06, le génotype au locus S a pu être inféré pour les individus non-recombinants.

Nous avons ensuite réalisé une deuxième série de génotypages avec les marqueurs moléculaires compris dans l'intervalle TACG-EU03D06, uniquement sur les 144 individus recombinants. L'ordre de ces marqueurs a été préalablement déterminé : 1) à partir du génotypage des 25 individus recombinants de la population B_a et 2) grâce à la réalisation du génotype graphique de ces recombinants (*cf.* chapitre 1). Pour la plupart des individus recombinants, les marqueurs encadrant le point de recombinaison ont été génotypés 2 fois. Quatre vingt douze individus se sont avérés non-recombinants dans l'intervalle EU08B02-TTAA qui encadre directement le locus S . Pour ces individus, le génotype au locus S a pu être inféré.

Les 52 individus pour lesquels l'événement de recombinaison a eu lieu entre les marqueurs EU08B02 et TTAA ont été retenus pour déterminer leur génotype au locus S .

Pour cela, les plantes ont été placées en vernalisation afin d'induire la floraison puis des croisements ont été réalisés avec des individus testeurs (*cf.* Matériels et méthodes p. 72). Le calcul des seed-set moyens a permis de distinguer les croisements compatibles (seed-set moyen $> 0,4$) des croisements incompatibles (seed-set moyen $< 0,1$) de façon aussi tranchée que pour la population B_a (*cf.* chapitre 2).

A l'issue de ces 3 étapes et en prenant en compte la population B_a , nous avons analysé un total de 2824 individus dont 169 recombinants (Figure 4.5). En se référant aux données théoriques utilisées pour estimer la taille optimale de la population à échantillonner (*cf.* paragraphe 4.1), en se plaçant au seuil $\alpha = 0,05$, cet effectif permet d'atteindre une valeur minimale de r de 0,0027 (Tableau 4.1). Si on considère l'estimation de 1,5 Mb/cM pour l'ensemble du génome de la chicorée, cette fréquence minimale permet d'ordonner avec confiance 2 marqueurs et d'estimer la distance entre eux dans un intervalle de 500 kb.

TABLEAU 4.1 – Valeurs minimales de r pouvant être calculées, à partir d'une population de 2824 individus, dans le cas où les intervalles de confiance des estimations de r ne peuvent pas se chevaucher.

Valeur de α	Configuration des marqueurs	$r \geq$
0,01	DD couplage	0,00942
	CC	0,00468
0,05	DD couplage	0,00545
	CC	0,00271

4.3.2 Gestion des données manquantes

Parmi les 52 individus recombinants dans l'intervalle EU08B02-TACG, 44 ont pu être phénotypés. Quatre individus n'ont pas fleuri malgré plusieurs périodes de vernalisation : les croisements nécessaires pour déterminer le génotype au locus S n'ont donc pas pu être réalisés. Trois autres étaient de génotype S_1 sur le chromosome parental et pouvaient être S_2 ou S_3 sur le chromosome recombinant. L'allèle S_1 étant dominant par rapport aux allèles S_2 et S_3 , aucun croisement n'aurait permis de phénotyper ces individus au locus S . Enfin, un individu de génotype potentiellement (phénotype S_2) ou S_1S_2 (phénotype S_1) s'est avéré compatible à la fois avec les testeurs de phénotype S_1 et ceux de phénotype S_2 : son génotype au locus S n'a donc pas pu être déterminé. Il s'agit du seul cas où il n'a pas été possible de déterminer le génotype au locus S par croisement. Les testeurs utilisés

4.3. Criblage séquentiel des recombinants

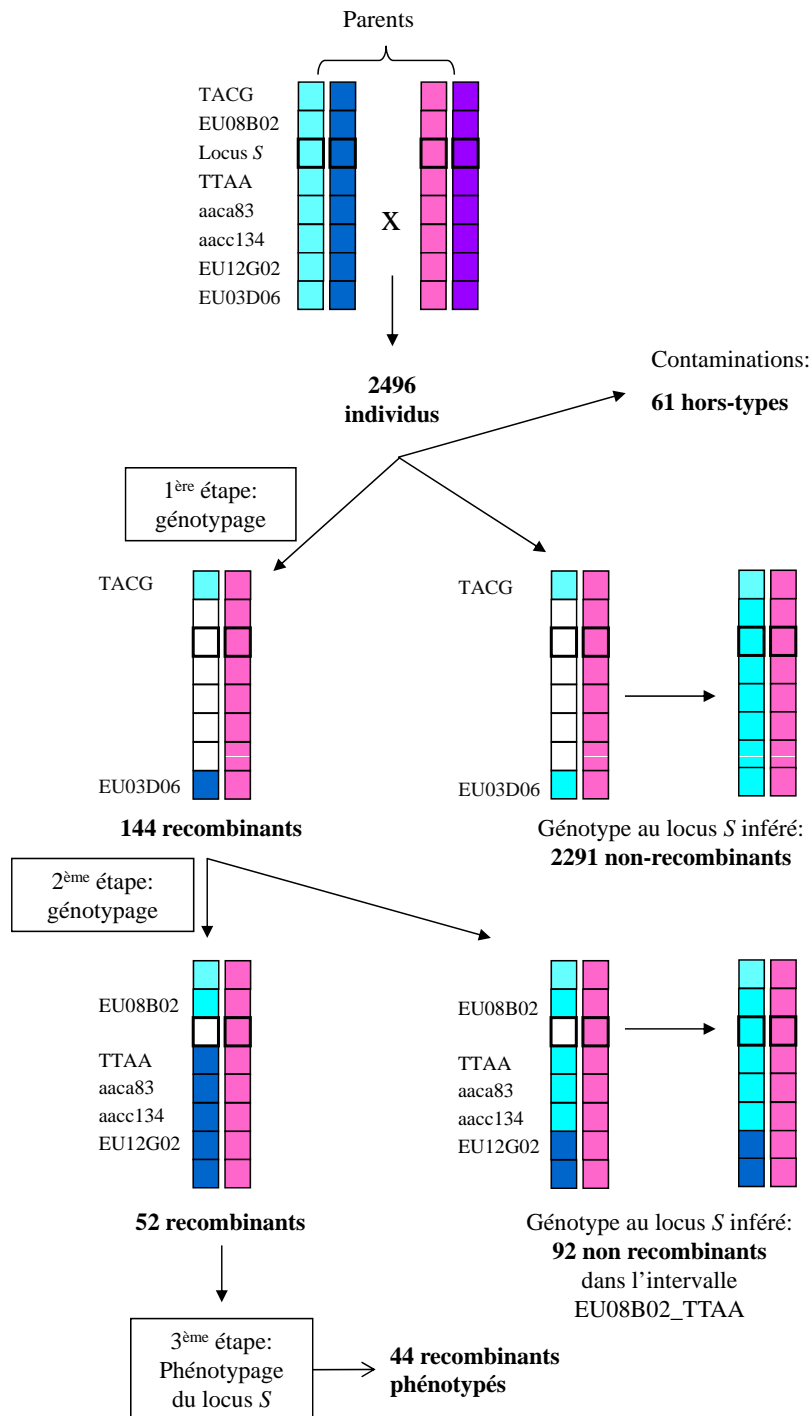


FIGURE 4.4 – Stratégie de génotypage et phénotypage en 3 étapes. Etape 1 : génotypage des 2496 individus pour les marqueurs TACG et EU03D06. Etape 2 : génotypage des 144 individus recombinants pour les marqueurs TTAA, aaca83, aacc134, EU12G02. Etape 3 : phénotypage du locus *S* par croisement avec des individus testeurs des 52 individus recombinants dans l'intervalle EU08B02-TTAA. Les chromosomes parentaux et ceux de la descendance sont représentés sous forme de génotype graphique, chaque case correspondant au locus d'un marqueur moléculaire, une couleur différente étant attribuée à chacun des chromosomes parentaux.

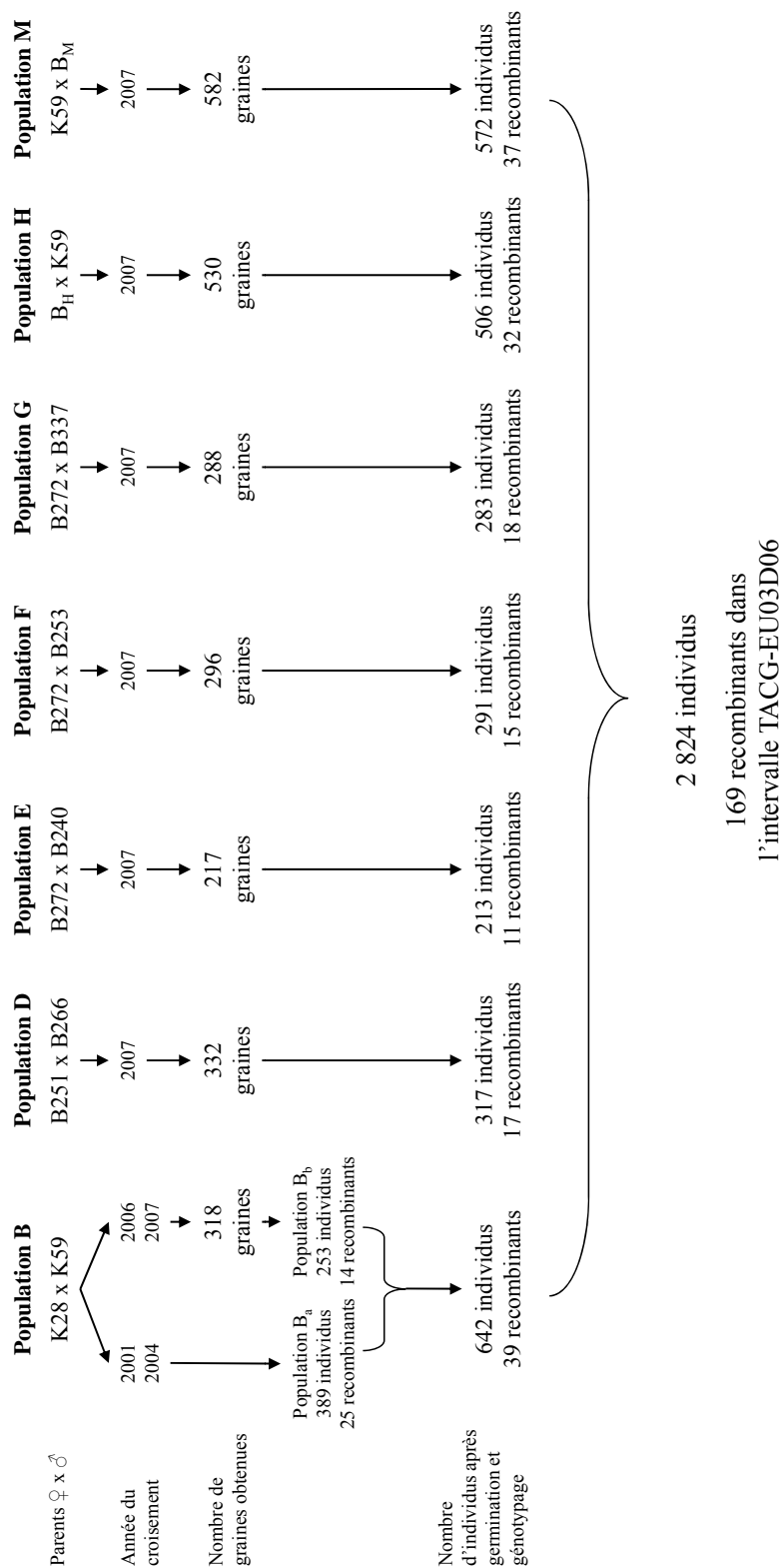


FIGURE 4.5 – Répartition des effectifs dans les différentes populations. Le nom de la population et le croisement correspondant (femelle x mâle), l'année de production des graines et le nombre obtenu, le nombre final d'individus analysés et le nombre d'individus recombinants dans l'intervalle TACG-EU03D06 sont représentés. Le nombre final d'individus correspond au nombre d'individus obtenus après germination et élimination des individus hors-types.

pour polliniser cet individu ont également été utilisés pour des croisements avec d'autres individus recombinants. Comme les résultats obtenus étaient cohérents, ces testeurs n'ont pas été mis en cause. Pour géotyper ce recombinant, il faudrait réaliser une nouvelle série de croisements à partir des graines récoltées.

Pour les 8 individus qui n'ont pas pu être phénotypés, le génotype au locus S était connu pour le chromosome non recombinant et manquant pour le chromosome recombinant. Afin de pouvoir placer le locus S sur toutes les cartes, un tirage aléatoire a été réalisé parmi les 2 allèles possibles pour ces chromosomes recombinants, permettant ainsi de leur attribuer un génotype pour le locus S . Ces tirages ont pris en compte la proportion des événements de recombinaison ayant eu lieu au dessus et en dessous du locus S . Cette attribution peut avoir un effet minime sur la précision des distances génétiques entre le locus S et les marqueurs qui l'encadrent (EU08B02 et TTAA). Cependant, elle n'a en aucun cas modifié l'ordre de ces marqueurs, celui-ci ayant été défini grâce au génotype graphique. De même, cette attribution n'a pas d'effet sur l'estimation de la distance entre EU08B02 et TTAA.

4.4 Construction de 2 cartes haute résolution

L'objectif était d'obtenir une carte génétique consensus à partir des différentes populations choisies. Lors de la construction de la carte moyenne densité (population B_a), nous avons constaté que la fréquence de recombinaison dans l'intervalle TACG-EU03D06 était beaucoup plus élevée pour les méioses de l'individu mâle K59 que pour les méioses de l'individu femelle K28 ($M/F = 7,7$, cf. chapitre 2). Ce phénomène pouvait avoir plusieurs explications : être dépendant de l'environnement ou bien du fond génétique de ces individus, être lié à leur génotype dans la région du locus S ou au locus S lui-même, ou encore être lié au sexe. Parmi les parents utilisés pour produire nos différentes populations, nous avons sélectionné des individus de même génotype ou de génotypes différents au locus S afin de distinguer un potentiel effet haplotype S ou génotype sur la fréquence de recombinaison. Des croisements réciproques entre individus B et K59 ont également été choisis afin de distinguer les méioses mâles et femelles du génotype K59. Nous avons donc testé l'homogénéité de ces différents paramètres afin de déterminer si il était possible de construire une ou bien plusieurs cartes consensus.

4.4.1 Effet de l'environnement : homogénéité des fréquences de recombinaison entre différents lots de graines pour les génotypes K28 et K59

La population B a été obtenue à partir de plusieurs lots de graines produits successivement en 2001, 2004, 2006 et 2007 (Figure 4.5). En 2007, K59 a également été utilisé comme pollinisateur pour le croisement H. Les croisements ont été réalisés à partir de clones des individus K28 et K59 produits par embryogénèse somatique (*cf.* Matériels et méthodes p. 46), ce qui pouvait constituer une source de variabilité des fréquences de recombinaison. De même, les conditions de culture dans la serre ont pu ne pas être strictement identiques d'une année à l'autre (lumière, température, hygrométrie...). Nos résultats montrent que pour les méioses mâles de K59 ($p = 0,98$, Tableau 4.2a) et pour les méioses femelles de K28 ($p = 0,87$, Tableau 4.2b), les fréquences de recombinaison dans l'intervalle TACG-EU03D06 étaient homogènes au cours du temps, ce qui suggère que, pour ces individus, l'environnement n'a pas eu d'effet sur la fréquence de recombinaison.

L'individu K59 a été utilisé comme pollinisateur pour les croisements B et H et comme plante mère pour les croisements M. La fréquence de recombinaison dans l'intervalle TACG-EU03D06 s'est révélée être homogène pour l'ensemble des méioses mâles de K59 (Tableau 4.2a), mais elle était significativement plus élevée que celle des méioses femelles de ce génotype ($p < 0,001$, $M/F = 3,6$, Tableau 4.2c). Les fréquences de recombinaison obtenues pour les méioses mâles et femelles de K59 ne peuvent donc pas être regroupées en un échantillon unique.

4.4.2 Effet du génotype : homogénéité des fréquences de recombinaison au sein des méioses mâles et femelles

L'homogénéité des taux de recombinaison, entre les méioses femelles d'une part et entre les méioses mâles d'autre part, a été testée pour les différents individus parentaux afin de mettre en évidence un éventuel effet du génotype dans l'intervalle TACG-EU06D06.

Dans le cas des populations H et M, les groupes B_H et B_M correspondaient à plusieurs individus de même génotype. Cependant, dans certains cas, les effectifs des descendants recombinants obtenus étaient trop faibles pour qu'ils soient traités indépendamment. De

4.4. Construction de 2 cartes haute résolution

TABLEAU 4.2 – Tests d’homogénéité (χ^2). Résultats obtenus pour la comparaison des proportions d’individus recombinants et non recombinants dans l’intervalle TACG-EU03D06 :

- en fonction de l’année de production des graines (effet de l’environnement) (a) pour les méioses mâles de K59 et (b) pour les méioses femelles de K28 ;
- en fonction du sexe (c) pour l’individu K59 ;
- en fonction du génotype (d) pour les méioses femelles et (e) pour les méioses mâles ;
- en fonction du sexe (f) pour la totalité des descendants.

* Lorsqu’au moins un des effectifs était inférieur à 5, la correction de Yates a été appliquée au test de χ^2 .

(a) Effet de l’environnement sur r pour K59 (méioses mâles)

Année de production des graines	2001	2004	2006	2007
Recombinants	11	12	7	35
Non recombinants	163	203	128	589

$$\chi^2 = 0,205 ; p = 0,98$$

(b) Effet de l’environnement sur r pour K28 (méioses femelles)*

Année de production des graines	2001	2004	2006	2007
Recombinants	2	1	0	2
Non recombinants	172	214	135	116

$$\chi^2 = 0,696 ; p = 0,87$$

(c) Effet du sexe sur r pour K59

Méiose	Mâle (croisements B et H)	Femelle (croisement M)
Recombinants	35	9
Non recombinants	589	563

$$\chi^2 = 13,715 ; p < 0,001$$

(d) Effet du génotype sur r pour les méioses femelles*

Individus	K28 (S_3S_4)	K59 (S_1S_2)	B251 (S_2S_4)	B272 (S_2S_4)	BH (S_2S_3)
Recombinants	5	9	6	17	2
Non recombinants	637	563	311	770	504

$$\chi^2 = 4,503 ; p = 0,21 \text{ (sans la population } B_H)$$

$$\chi^2 = 7,515 ; p = 0,11 \text{ (avec la population } B_H)$$

(e) Effet du génotype sur r pour les méioses mâles

Individus	K59 (S_1S_2)	B266 (S_1S_3)	B240 (S_1S_3)	B253 (S_1S_3)	B337 (S_1S_3)	BM (S_2S_3)
Recombinants	65	12	7	10	13	29
Non recombinants	1083	305	206	281	270	543

$$\chi^2 = 4,894 ; p = 0,30 \text{ (sans la population } B_M)$$

$$\chi^2 = 4,933 ; p = 0,42 \text{ (avec la population } B_M)$$

(f) Effet du sexe sur r

	Méioses femelles	Méioses mâles
Recombinants	39	136
Non recombinants	2785	2688

$$\chi^2 = 55,485 ; p < 0,00001$$

ce fait, les tests d'homogénéité ont été réalisés dans un premier temps sans les groupes B_H et B_M .

Dans l'intervalle TACG-EU03D06, les fréquences de recombinaison des méioses femelles étaient homogènes pour les génotypes K28, K59, B251 et B272 ($p = 0,21$, Tableau 4.2d) et celles des méioses mâles étaient homogènes pour les génotypes K59, B266, B240, B253 et B337 ($p = 0,30$, Tableau 4.2e).

A partir de ces résultats, nous avons choisi de considérer les groupes B_H et B_M comme des groupes homogènes. Ils ont été inclus dans de nouveaux tests d'homogénéité des fréquences de recombinaison réalisés pour les méioses femelles ($p = 0,11$, Tableau 4.2d) et pour les méioses mâles ($p = 0,42$, Tableau 4.2e). Les résultats indiquent que les fréquences de recombinaison étaient homogènes, d'une part pour les méioses mâles et d'autre part pour les méioses femelles, ceci indépendamment du génotype et de l'haplotype S .

4.4.3 Effet du sexe sur la fréquence de recombinaison

Après avoir établi l'homogénéité des fréquences de recombinaison entre génotypes différents pour les méioses mâles et les méioses femelles, nous avons testé l'homogénéité des fréquences de recombinaison entre méioses mâles et femelles dans l'intervalle TACG-EU03D06, en considérant la totalité de nos descendants. Avec un total de 39 recombinants pour les méioses femelles et 139 pour les méioses mâles, il existait une différence très significative ($p < 0,00001$, Tableau 4.2f) avec une fréquence de recombinaison 3,5 fois supérieure pour les méioses mâles.

Ce résultat confirme que la différence de taux de recombinaison observée entre K28 et K59 pour la population B_a était liée au sexe (*cf.* chapitre 2).

4.4.4 Distorsion de ségrégation dans la région du locus S

Pour chacune des populations, la ségrégation des marqueurs TACG, EU03D06 et du locus S a été évaluée par un test de χ^2 (Tableau 4.3). Une distorsion de ségrégation a été observée pour 1 (EU03D06), 2 (EU03D06 et le locus S) et les 3 locus respectivement dans les populations E ($p < 0,05$), M ($p < 0,05$) et F ($p < 0,0001$). Dans chacun des cas, la distorsion a pu être attribuée à de la sélection gamétique (*cf.* Matériels et méthodes p. 66, Tableau 4.3).

La faible distorsion de ségrégation observée dans la population E n'a pas affecté la ségrégations des allèles du locus S . Dans les populations F et M, cette distorsion s'est

4.4. Construction de 2 cartes haute résolution

traduite par une sous représentation de l'haplotype S_3 . Cependant, cette sélection n'a pas pu être attribuée à un allèle ou un haplotype donné dans la région du locus S . En effet, la distorsion de ségrégation n'a pas été observée de façon systématique dans les populations issues de parents partageant le même génotype au locus S (parent S_1S_3 pour les populations D, E, F et G, et parent S_2S_3 pour les populations H et M, Figure 4.3).

TABLEAU 4.3 – Etude de la conformité de ségrégation (tests de χ^2) des allèles de TACG, du locus S et de EU03D06. Les résultats significatifs ($p < 0,05$) sont indiqués en gras. * Dans ces populations, la distorsion de ségrégation s'est traduite par une sous représentation de l'haplotype S_3 et des allèles liés à cet haplotype.

Population	Marqueur	Distorsion		Sélection gamétique		Sélection zygotique	
		χ^2	p	χ^2	p	χ^2	p
B _b	TACG	1,150	0,765				
	locus S	0,771	0,856				
	EU03D06	0,486	0,922				
D	TACG	5,334	0,149				
	locus S	4,628	0,201				
	EU03D06	6,192	0,103				
E	TACG	5,986	0,112				
	locus S	6,399	0,094				
	EU03D06	8,690	0,034	9,484	0,024	3,759	0,289
F*	TACG	25,203	0,000	48,893	0,000	0,431	0,934
	locus S	25,649	0,000	49,333	0,000	0,554	0,907
	EU03D06	27,541	0,000	54,007	0,000	3,759	0,289
G	TACG	2,837	0,417				
	locus S	3,403	0,334				
	EU03D06	1,537	0,674				
H	TACG	2,221	0,528				
	locus S	1,874	0,599				
	EU03D06	2,253	0,522				
M*	TACG	7,385	0,061				
	locus S	8,014	0,046	15,804	0,001	0,232	0,972
	EU03D06	10,392	0,016	20,277	0,000	0,123	0,989

4.4.5 Construction des cartes génétiques des parents et de 2 cartes consensus

A partir des résultats précédents, nous avons pu construire une carte mâle et une carte femelle pour K59 et une carte pour chacun des autres individus, soit un total de 11 cartes (Figure 4.6). Cinq cartes correspondaient aux méioses femelles (K28, K59, B251, B272, B_H) et 6 aux méioses mâles (K59, B266, B240, B253, B337, B_M). La construction de ces cartes a permis de constater que l'ordre des marqueurs était identique, quand les événements de recombinaison permettaient de les positionner.

Tous les marqueurs n'ont pas été représentés sur la totalité des cartes. En effet, quelques individus n'ont pas pu être génotypés sur l'ensemble d'une population, en raison du caractère dominant ou du manque de polymorphisme associé à ces marqueurs. La construction du génotype graphique des individus recombinants a cependant permis de vérifier que l'ordre des marqueurs était conservé quelle que soit la population. De même, les génotypes graphiques ont permis de constater que le marqueur *tcaa300* coségrégait avec le locus *S*, tandis que le marqueur *aaca83* coségrégait avec le marqueur TTAA pour toutes les populations.

Le fait que l'ordre des marqueurs ait été conservé sur l'ensemble des cartes était l'une des conditions nécessaires à la création d'une carte consensus. Toutefois, en raison de la différence entre méioses mâles et femelles, 2 cartes consensus ont été construites. Ainsi, les distances génétiques représentées étaient le reflet des différences biologiques qui existaient entre les méioses mâles et femelles du matériel végétal étudié (Figure 4.7). Les marqueurs utilisés pour construire ces 2 cartes étaient ceux qui ont pu être cartographiés chez tous les parents (Figure 4.6).

Le locus *S* a été cartographié entre les marqueurs EU08B02 et TTAA, dans un intervalle de $0,60 \pm 0,20$ cM pour les méioses femelles, cet intervalle étant de $1,24 \pm 0,29$ pour les méioses mâles ($\alpha = 0,05$). Les marqueurs n'ayant été cartographiés que sur une partie des populations ont également été positionnés sur ces cartes à titre indicatif (Figure 4.7).

4.4.6 Implication du locus *S* dans l'effet du sexe sur la fréquence de recombinaison ?

Les 11 cartes des parents illustrent la différence existant entre les méioses mâles et femelles (Figure 4.6). Pour l'intervalle TACG-EU03D06, nous avons observé une distance

4.4. Construction de 2 cartes haute résolution

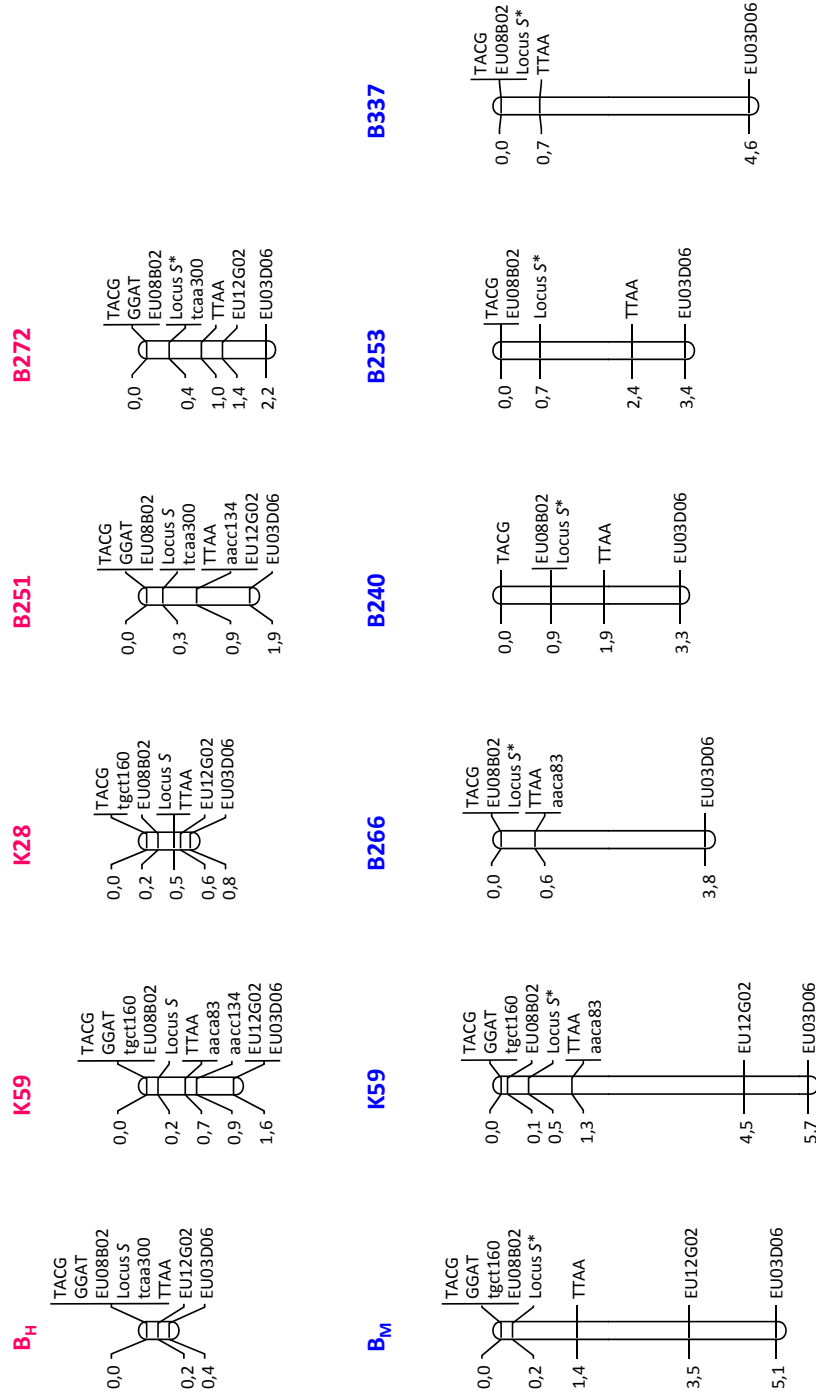


FIGURE 4.6 – Cartes génétiques de l'intervalle TACG-EU03D06 (extrémité du groupe de liaison 2) obtenues pour les différents génotypes utilisés comme parents. Les distances génétiques cumulées sont indiquées à gauche des groupes de liaison et sont exprimées en cM. Les cartes obtenues à partir des méioses mâles sont indiquées en rose. Les cartes obtenues à partir des méioses femelles sont indiquées en bleu. * : la distance du locus *S* par rapport aux marqueurs qui l'encadrent est le résultat d'une approximation en raison de données manquantes.

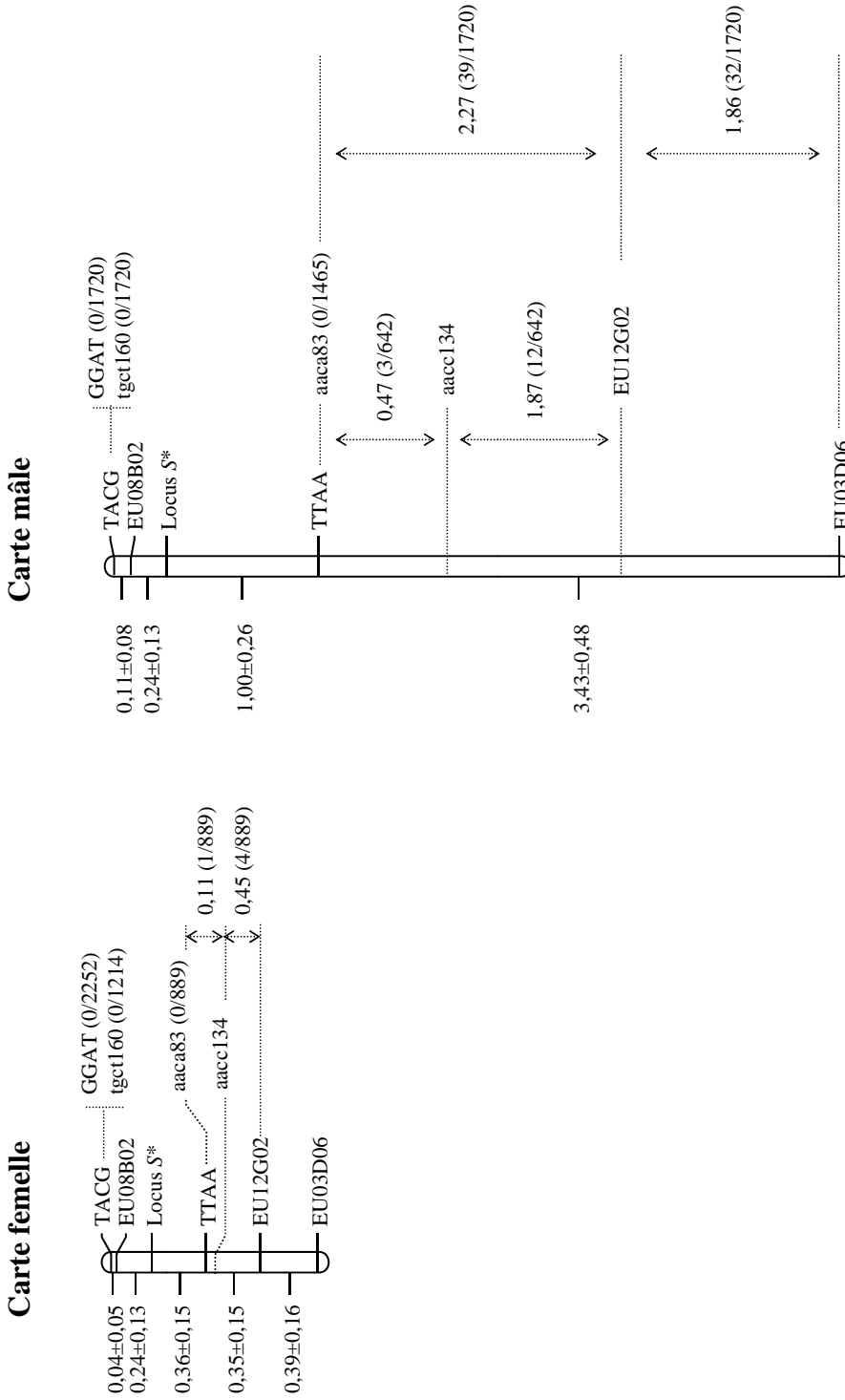


FIGURE 4.7 – Cartes génétiques consensus de l'intervalle TACG-EU03D06 (extrémité du groupe de liaison 2) des méiotes mâles et femelles. Les traits pleins sur les groupes de liaison correspondent aux marqueurs informatifs pour les 2 824 méiotes. Les distances (en cM ± intervalle de confiance, $\alpha = 0,05$) entre les marqueurs sont indiquées à gauche du groupe de liaison. * La distance du locus *S* par rapport aux marqueurs qui l'encadrent est le résultat d'une approximation en raison de données manquantes. Les lignes pointillées correspondent aux marqueurs informatifs pour une partie des méiotes, les distances calculées entre ces marqueurs et les marqueurs adjacents sont exprimées en cM avec, entre parenthèses, le nombre de recombinants rapporté au nombre de méiotes informatives.

4.4. Construction de 2 cartes haute résolution

de 2,2 cM maximum pour les méioses femelles et de 3,3 cM minimum pour les méioses mâles, avec une distance génétique variant de 0,4 à 5,7 cM selon l'individu considéré. L'intervalle encadrant le locus S variait, quant à lui, de 0,2 cM (B_H) à 2,4 cM (B253). Le rapport M/F final observé entre K28 et K59 (7,1) s'est révélé 2 fois plus élevé que celui obtenu tous génotypes confondus (M/F = 3,6). En comparant les fréquences de recombinaison des méioses mâles et femelles de K59, le rapport M/F était de 3,5. Nous avons pu constater que la fréquence de recombinaison de K28 (méiose femelle) était l'une des plus faibles alors que celle de K59 (méiose mâle) était la plus grande obtenue pour l'intervalle TACG-EU03D06. Cela explique que le rapport M/F estimé à partir de la descendance K28 x K59 ait été si élevé, comme cela avait déjà été observé à partir de la population B_a (*cf.* chapitre 2).

Afin de déterminer si le rapport M/F était constant sur l'intervalle TACG-EU03D06, les données obtenues ont été regroupées pour l'ensemble des méioses mâles et femelles et l'intervalle TACG-EU03D06 a été divisé en 3 segments (Tableau 4.4). Les résultats ont montré que le rapport M/F n'est pas significativement différent de 1 pour l'intervalle TACG-locus S ($p = 0,64$) et devient significatif pour les intervalles locus S -TTAA ($p = 0,0034$) et TTAA-EU03D06 ($p < 0,00001$).

La différence entre méioses mâles et femelles ne semble pas directement liée au locus S , le rapport M/F le plus élevé étant obtenu pour l'intervalle TTAA-EU03D06 (M/F = 4,6).

TABLEAU 4.4 – Effet du sexe sur les proportions d'individus recombinants et non recombinants pour différents intervalles de la région du locus S (Tests de χ^2).

Intervalle Méiose	TACG-locus S		locus S -TTAA		TTAA-EU03D06	
	femelle	mâle	femelle	mâle	femelle	mâle
Recombinants	8	10	10	28	21	98
Non recombinants	2816	2814	2814	2796	2803	2726
	$\chi^2 = 0,223$; $p = 0,64$		$\chi^2 = 8,584$; $p = 0,003$ M/F = 2,8		$\chi^2 = 50,896$; $p < 0,0001$ M/F = 4,6	

Chapitre 5

Discussion

Pour qu'une stratégie de clonage positionnel soit envisageable, le ou les gènes d'intérêt doivent être positionnés sur une carte génétique. A partir des marqueurs flanquants et de banques d'ADN génomique (banque BAC, YAC, . . .), une marche chromosomique peut alors être initiée. Le rapport distance physique/distance génétique est variable d'une espèce à l'autre et au sein d'un même génome, et quelques cM peuvent correspondre à quelques kb comme à plusieurs Mb (Tableau 5.1, Tagu et Moussard, 2003). Toutefois, ce rapport dépend surtout de la taille du génome de l'espèce considérée, la longueur des cartes génétiques, rapportée au nombre de chromosomes, étant relativement homogène entre espèces (De Vienne, 1998). Chez la chicorée, le rapport distance physique/distance génétique a été estimé à 1,5 Mb/cM en considérant la taille du génome (1,4 Mb) et de la longueur totale de la carte génétique (930 cM, Cadalen *et al.*, 2010). Il s'agit d'une valeur élevée, mais cohérente avec la taille du génome qui est plus de 10 fois plus importante que celle du génome d'*A. thaliana*.

De plus, pour que la complémentarité des spécificités du pollen et du stigmate, essentielle au maintien de l'autoincompatibilité, soit conservée, le locus *S* doit théoriquement être localisé dans une région où une suppression de la recombinaison est observée (Castric et Vekemans, 2004 ; Uyenoyama, 2005). Chez *P. inflata* et *I. trifida*, le rapport distance physique/distance génétique dans la région du locus *S* est entre 10 et 20 fois plus élevé que pour l'ensemble du génome (Tableau 5.2, Suzuki *et al.*, 2004 ; Wang *et al.*, 2003). Toutefois, dans le cas de *B. oleracea*, il n'y a pas de différence entre la région du locus *S* et le reste du génome, même si au niveau du locus *S* lui-même on observe de la suppression de recombinaison (Casselman *et al.*, 2000). Le rapport de 1,5 Mb/cM, estimé chez la chicorée, justifie à lui seul la décision de densifier la région

du locus S en marqueurs et de les cartographier avec précision, avant d'entreprendre la marche chromosomique. Cependant, on s'attend à ce que, au locus S , le rapport Mb/cM soit supérieur à l'estimation faite pour l'ensemble du génome, tout en espérant que la fréquence de recombinaison ne soit pas 20 fois inférieure à la moyenne du génome, comme c'est le cas chez *P. inflata*.

La stratégie BSA consiste à identifier des marqueurs moléculaires liés à un caractère donné, en utilisant des pools d'ADN d'individus partageant le même phénotype (Michelmore *et al.*, 1991). Cette stratégie est fréquemment couplée à la technique AFLP afin de trouver des marqueurs liés à un gène ou un QTL d'intérêt, que ce soit chez les végétaux ou les animaux (de Brito *et al.*, 2010 ; Wei *et al.*, 2009 ; Xiao *et al.*, 2007 ; Zhang *et al.*, 2009). Lorsque cette approche est utilisée, la fenêtre dans laquelle se trouvent les marqueurs AFLP d'intérêt est de l'ordre d'une dizaine de cM (Tableau 5.1, de Brito *et al.*, 2010 ; Pierre *et al.*, 2000 ; Wei *et al.*, 2009 ; Zhang *et al.*, 2009). Le fait de mettre en place une stratégie BSA assistée par marqueur est moins fréquente (Cnops *et al.*, 1996 ; Jordan *et al.*, 2006 ; Lahaye *et al.*, 1998 ; Meksem *et al.*, 1995). En effet, elle nécessite de disposer préalablement de marqueurs encadrant le locus d'intérêt, dans une fenêtre de quelques cM, et de connaître les points de recombinaison chez les individus utilisés pour construire les bulks. Elle permet alors de cibler très précisément la région d'intérêt.

Le dispositif utilisé pour densifier la région du locus S en marqueurs s'est avéré complexe (6 types de bulks différents plus les individus K28 et K59), mais devait permettre d'identifier rapidement les marqueurs d'intérêt parmi le nombre important de marqueurs polymorphes générés par la technique AFLP. Notre objectif était double : 1) identifier des marqueurs liés de façon spécifique à un haplotype S (S_1 ou S_2) dans un intervalle moyen de 3,34 cM (de 0,77 cM pour la carte K28 et 5,91 cM pour la carte K59) et 2) éliminer les marqueurs localisés hors de cet intervalle et les faux positifs, afin de limiter les étapes de validation.

Cette stratégie de bulks a permis de valider 6 des 14 marqueurs AFLP identifiés dans un intervalle moyen de 3,34 cM. Six des 14 marqueurs identifiés se sont avérés hors de l'intervalle ciblé. Dans une approche menée chez l'orge (Meksem *et al.*, 1995), des résultats similaires ont été obtenus : des pools constitués pour moitié d'individus recombinants sous le locus d'intérêt ont été construits et des marqueurs AFLP identifiés comme liés au locus d'intérêt ont été cartographiés dans l'intervalle où ont eu lieu les

TABLEAU 5.1 – Exemples de résultats obtenus pour des approches BSA, utilisant les AFLP, basées sur le phénotype (BSA) ou assistées par marqueurs (BSA+), pour différentes plantes et différents caractères. Sont indiqués : la taille de la carte génétique (1), la taille du génome (2), le nombre total de fragments générés (3), le pourcentage de marqueurs polymorphes entre les parents (4), le nombre de marqueurs identifiés comme liés au locus d'intérêt (5), l'intervalle dans lequel ont été identifiés les marqueurs liés au locus d'intérêt (6), la répartition moyenne des marqueurs polymorphes sur le génome (7) et dans la région d'intérêt (8), l'estimation du rapport distance physique/distance génétique pour l'ensemble du génome (9). Références (Réf.) : a) Lahaye *et al.* (1998) et Wicker *et al.* (2008), b) Meksem *et al.* (1995) et Zhu *et al.* (2008), c) Jordan *et al.* (2006), Pierre *et al.* (2000) et Yoo *et al.* (2003), d) Cnops *et al.* (1996), Hauge *et al.* (1993) et Tagu et Moussard (2003), e) Cette étude, f) Zhang *et al.* (2009), g) Noguera *et al.* (2005), h) Pierre *et al.* (2000), i) Wei *et al.* (2009), j) de Brito *et al.* (2010), k) Quillet *et al.* (en préparation).

Approche	Espèce	Caractère	cM ¹	Mb ²	Nb ³	% ⁴	Nb ⁵	cM ⁶	AFLP/cM ⁷	AFLP/cM ⁸	Mb/cM ⁹	Réf.
BSA+	Orge	résistance Rar1	1 800	5 500	40 000	7,0	2	2,2	1,6	0,9	3,056	a
BSA+	Pomme de terre	résistance R1	1 000	865	14 580	44,0	8	3,7	6,4	2,2	0,865	b
BSA+	Poivre	résistance à <i>X. campestris</i>	1 276	2 702	102 400	16,0	11	1,7	12,8	6,5	2,118	c
BSA+	<i>A. thaliana</i>	gène tornado1	675	150	27 280	30,0	17	3	12,1	5,7	0,222	d
BSA+	Chicorée	locus <i>S</i>	930	1 400	31 000	7,6	6	3,34	2,5	1,8	1,505	e
BSA	<i>Carthamus tinctorius</i>	Hydroxysafflor yellow A			769	16-43	4	14,2				f
BSA	Melon	andromonoécie			5 119	37,0	1	3,3				g
BSA	Poivre	résistance à <i>X. campestris</i>			90 000	14,0	16	15				h
BSA	<i>Brassica rapa</i>	stérilité mâle			5 280	18,0	4	13,9				i
BSA	<i>Coffea arabica</i>	résistance à <i>H. vastatrix</i>			11 000	57,0	3	54,3				j
BSA	Chicorée	stérilité mâle	930	1 400	31 000	7,6	17	12	2,5	1,4	1,505	k

TABLEAU 5.2 – Localisation chromosomique du locus S et relation distance physique/distance génétique au voisinage du locus S pour différentes espèces. (*) relation distance physique/distance génétique à proximité du locus S , (**) relation distance physique/distance génétique estimée sur l'ensemble du génome. Références : a) Wang *et al.* (2003), b) Yang *et al.* (2007), c) Casey *et al.* (2010), d) Casselman *et al.* (2000) et Imano *et al.* (1998), e) Suzuki *et al.* (2004), f) Wheeler *et al.* (2003).

Famille	Espèce	Localisation	Mb/cM*	Mb/cM**	Références
Solanaceae	<i>Petunia inflata</i>	sub-centromérique	17,6	0,9	a
Plantaginaceae	<i>Antirrhinum majus</i>	hors du centromère, région hétérochromatique			b
Fabaceae	<i>Trifolium repens L</i>	probablement non-centromérique			c
Brassicaceae	<i>Brassica oleracea</i>	extrémité d'un chromosome	0,41	0,5 à 0,7	d
Convolvulaceae	<i>Ipomoea trifida</i>	extrémité d'un chromosome	2,273	0,2	e
Papaveraceae	<i>Papaver rhoeas</i>	centromérique?			f

recombinaisons. Une explication possible est qu'il existe un effet quantitatif associé à la détection des marqueurs AFLP. Les événements de recombinaison ayant eu lieu dans l'intervalle EU03D06-sw2H09, pour les individus des bulks P5 et P6, se répartissaient en différents points de cet intervalle. De ce fait, la composition de ces bulks pour cibler la région du locus S aurait été moins spécifique. Les 6 marqueurs hors de l'intervalle TACG-EU03D06 sont vraisemblablement localisés plus loin sur le groupe de liaison 2. De plus, un nombre élevé de fragments faux positifs a été obtenu. Deux des 14 marqueurs identifiés à partir des bulks n'ont pas été amplifiés chez les recombinants de la population B_a et 80 pics ont été identifiés dans les bulks P1 à P4, mais se sont avérés absents chez K59 et dans les bulks P5 à P8. Tous les individus utilisés pour construire les bulks d'homozygotes étant issus de K59, il est intrigant de trouver autant de pics orphelins, d'autant plus que la technique AFLP est présentée comme une technique hautement reproductible (Jones *et al.*, 1997 ; Meudt et Clarke, 2007). Néanmoins, un certain nombre de causes peuvent entraîner des différences de profils AFLP générés pour un même individu : une variation de l'intensité des pics ou de leur mobilité sur le profil, l'apparition/disparition de pics causée par des anomalies pendant les étapes successives nécessaires à la production des marqueurs AFLP, le degré de méthylation de l'ADN qui peut être variable, et enfin la subjectivité de l'analyse d'un profil AFLP (Bonin *et al.*, 2004 ; Reyna-Lopez et Ruiz-Herrera, 2004 ; Schellenbaum *et al.*, 2008). Dans notre cas, il s'agit de la comparaison de profils AFLP d'individus différents. De plus, K59 est un individu qui a été multiplié pendant plusieurs années par embryogénèse somatique, alors que les individus utilisés pour les bulks P1 à P4 ont été semés quelques mois avant la construction des pools. Enfin, les extractions d'ADN ont été réalisées à différents moments. Cela pourrait expliquer en partie ces faux positifs, cependant, il est difficile de conclure sur l'origine de ces fragments. La stratégie mise en place a toutefois permis d'exclure ces pics alors même qu'ils semblaient dans un premier temps spécifiques des haplotypes S_1 ou S_2 et a ainsi limité le nombre de fragments à valider.

Si on considère que les marqueurs AFLP polymorphes sont répartis de façon uniforme sur le génome, le nombre de marqueurs AFLP validés et cartographiés dans la région du locus S (1,8 AFLP/cM) est cohérent avec la répartition des marqueurs estimée sur la totalité du génome (2,5 AFLP/cM, Tableau 5.1). L'approche mise en place a permis de densifier la région du locus S en marqueurs, de restreindre le nombre de marqueurs à valider, et surtout d'identifier un marqueur (tcaa300) qui coségrège avec le locus S . Cependant, à partir de cette stratégie, seuls les marqueurs spécifiquement liés à S_1 ou S_2

et associés en phase de couplage ont été détectés. Ainsi les marqueurs AFLP spécifiques d'un haplotype mais associés en phase de répulsion, comme par exemple le fragment *ggat128*, et les marqueurs partagés avec un autre haplotype (S_3 ou S_4), n'ont pas été décelés à partir des bulks. Cette stratégie n'a donc permis de détecter qu'une partie des AFLP polymorphes liés au locus S . La densité des marqueurs validés dans la région du locus S est similaire à la fréquence des AFLP estimée sur l'ensemble du génome. Dans les études citées dans le tableau 5.1, la fréquence des marqueurs identifiés dans la région d'intérêt est au moins 2 fois plus faible que celle estimée sur l'ensemble du génome. La stratégie mise en place chez la chicorée, pour densifier la région du locus S en marqueurs, s'est donc révélée particulièrement efficace.

Grâce à la technique d'isolement des fragments AFLP mise en place au laboratoire (Paris *et al.*, 2011, d'après Brugmans *et al.* 2003), la séquence de 3 des 4 marqueurs AFLP transformés en marqueurs SCAR a été obtenue sans ambiguïté, la séquence du 4^e marqueur étant trop courte pour être utilisée. Dans le cas du marqueur *ttaa440*, de nombreuses séquences de tailles variables ont été clonées, mais les étapes de ré-amplification et de co-migration avant séquençage ont permis de distinguer les clones contenant le fragment d'intérêt des autres clones. Une fois la séquence de ces marqueurs obtenue, leur transformation en marqueurs codominants n'a pas été possible pour tous. En effet, le polymorphisme AFLP observé étant lié au(x) site(s) de restriction ou aux bases sélectives, 2 marqueurs (TCAA et AACC) n'ont pas pu être utilisés en tant que marqueurs codominants. Le cas du marqueur TCAA qui correspond à une séquence répétée dans le génome n'est pas un résultat singulier. En effet, ce type de résultat a également été obtenu chez le poivre (Jordan *et al.*, 2006 ; Pierre *et al.*, 2000), la pomme de terre (Meksem *et al.*, 1995), l'asperge (Reamon-Buttner *et al.*, 1999) ou encore le bœuf (Xiao *et al.*, 2007).

La technique AFLP est donc une méthode efficace pour produire un grand nombre de marqueurs et enrichir une carte génétique. Cependant, le fait que la séquence d'un fragment AFLP corresponde à une séquence répétée n'est pas propre à la chicorée et est un des éléments à prendre en compte pour ce type d'approche. Cela indique que la transformation de marqueurs AFLP en marqueurs codominants n'est pas possible de façon systématique. Leur utilisation en tant que sonde, par exemple pour cribler des banques BAC, doit donc être sujette à caution.

Avec les marqueurs identifiés dans cette étude, nous avons construit une carte haute résolution afin de valider avec précision la position du locus S , ainsi que l'ordre des mar-

queurs autour de celui-ci.

Nous avons obtenu une carte consensus pour les méioses mâles et une carte consensus pour les méioses femelles des différentes populations. Pour ces 2 cartes, l'ordre des marqueurs est le même avec une fenêtre autour du locus *S* réduite par rapport aux cartes obtenues dans le chapitre 3 (0,6 cM pour les méioses femelles et 1,24 cM pour les méioses mâles). Le marqueur *tcaa300* reste lié de façon absolue au locus *S*. De plus, si le rapport de 1,5 Mb/cM est conservé au niveau du locus *S*, le nombre d'individus analysés (2824) devrait permettre de déterminer avec précision la distance génétique entre des marqueurs séparés de 500 kb.

La construction des cartes a également confirmé le caractère monocus de l'autoincompatibilité dans nos populations. Ainsi, la stratégie de phénotypage associée aux marqueurs moléculaires s'est avérée efficace pour déterminer le génotype *S* d'un maximum d'individus. Les seules difficultés rencontrées ont été générées par des problèmes de vernalisation des plantes, ou bien de dominance entre allèles *S* et ne concernent que 7 individus au total. Parmi tous les génotypes croisés avec des testeurs, que ce soit pour la construction de la carte moyenne densité (*cf.* chapitre 2) ou bien de la carte haute résolution (soit près de 400 individus), un seul individu (H423) n'a pu être génotypé au locus *S*, puisqu'il s'est avéré compatible avec les 2 testeurs. L'origine du caractère compatible de cet individu n'est pas connue mais pourrait être expliquée par un événement de recombinaison au niveau du locus *S*, ou par l'expression d'un ou plusieurs gènes modificateurs atténuant la réponse d'incompatibilité. La seconde hypothèse semble moins vraisemblable, puisqu'un seul individu a présenté ce phénotype et que l'ensemble des populations étudiées dérive du même fond génétique (K28 x K59). Cependant, pour déterminer la possibilité de l'une ou l'autre hypothèse, de nouveaux croisements tests et des autofécondations vont être réalisés à partir des descendants de l'individu H423. La ségrégation du caractère compatible dans la descendance de cet individu, si elle est retrouvée, permettra de déterminer si elle est liée au locus *S* ou non. Ainsi, une première explication pourrait être apportée quant à l'origine du caractère pseudoincompatible ou compatible/autofertile de la chicorée, observé dans certains cultigrupes (Castaño *et al.*, 1997 ; Coppens d'Eeckenbrugge *et al.*, 1987).

Dans la région du locus *S*, on s'attend à observer un phénomène de suppression de la recombinaison, tel que ce que l'on peut observer au niveau des chromosomes sexuels chez les animaux ou chez les végétaux, ou encore dans le cas de locus associés à la compati-

lité sexuelle chez les champignons (Hedrick, 2007 ; Kejnovsky *et al.*, 2009 ; Menkis *et al.*, 2008 ; Uyenoyama, 2005 ; Yu *et al.*, 2009). Chez la chicorée, le locus *S* n'étant pas encore atteint, l'importance de la suppression de recombinaison n'a pu être étudiée. Toutefois, une différence significative entre les fréquences de recombinaison des méioses mâles et femelles a attiré notre attention lors de la construction de la carte moyenne densité (*cf.* chapitre 2). L'étude de l'ensemble du génome à partir de la population B_a (n = 174) avait montré une hétérogénéité des fréquences de recombinaison avec des rapports M/F allant de 0,26 à 9,33 selon les intervalles considérés. Un rapport moyen de 1,3 a été calculé à partir des régions comparées sur l'ensemble de la carte génétique (*cf.* chapitre 2). Dans cette étude, l'intervalle TACG-sw2H09 était celui qui présentait le rapport M/F le plus élevé (9,33). L'observation de l'intervalle TACG-EU03D06 dans le cadre de la carte haute résolution a confirmé cette différence, mais avec un rapport presque 3 fois moins important (M/F = 3,5). L'importance des différences observées sur l'ensemble du génome est donc à relativiser en raison de l'utilisation de seulement 2 génotypes et d'un effectif relativement faible. En effet, au niveau de la région du locus *S*, nous avons pu constater que même si les fréquences de recombinaison observées pour les méioses mâles d'une part et les méioses femelles d'autre part étaient statistiquement homogènes, il existait de la variabilité au sein de chaque groupe. Ainsi, les méioses mâles de K59 présentaient la fréquence de recombinaison la plus élevée et les méioses femelles de K28 une des fréquences les plus faibles, parmi les différents génotypes analysés.

Chez les espèces autoincompatibles, ce type de distribution des fréquences de recombinaison à l'échelle du génome (M/F > 1) a été observé chez le pommier (Maliepaard *et al.*, 1998) ou *B. nigra* (Lagercrantz et Lydiate, 1995) alors que chez *B. oleracea*, *r* est plus élevée pour les méioses femelles (Kearsey *et al.*, 1996). Chez *Turnera subulata*, qui est une espèce distyle, les taux de recombinaison des méioses mâles et femelles ont été comparés au niveau de 3 locus, dont le locus *S*. Ces 3 locus placés sur des groupes de liaison différents ont tous montré un taux de recombinaison 4 à 6 fois plus important pour les méioses femelles (Labonne *et al.*, 2007). Dans le cas de *Coffea canephora*, à partir de l'étude d'intervalles répartis sur les différents groupes de liaison, aucune différence n'a été observée entre les fréquences de recombinaison des méioses mâles et femelles (Lashermes *et al.*, 2001). Enfin, une étude menée chez l'espèce autogame *A. thaliana* a montré une différence du taux de recombinaison entre méioses mâles et femelles, variable le long du chromosome 4, avec une fréquence de recombinaison jusqu'à 6 fois plus important pour la méiose mâle (Drouaud *et al.*, 2007).

La différence entre les taux de recombinaison mâles et femelles (hétérochiasmie) a été largement observée aussi bien chez les animaux que chez les végétaux. Toutefois, les mécanismes évolutifs et de régulation à l'origine de ces différences ne sont pas connus (Drouaud *et al.*, 2007 ; Lenormand et Dutheil, 2005). Les résultats obtenus pour *T. subulata*, *B. oleracea* et *A. thaliana* sont en accord avec les données de Lenormand et Dutheil (2005) qui suggèrent que le taux de recombinaison des méioses femelles est plus faible que celui des mâles pour les espèces autogames, en raison d'une sélection haploïde. Cependant, les résultats obtenus pour la chicorée, le pommier ou *B. nigra* ne peuvent être expliqués au regard des différentes hypothèses qui ont pu être proposées pour expliquer l'hétérochiasmie, que ce soit la sélection neutre, sexuelle ou haploïde, ou encore la pléiotropie (Hedrick, 2007)¹⁸.

Nos résultats ont également montré que les taux de recombinaison sont significativement différents sous le locus *S* (locus *S*-EU03D06), alors que l'intervalle au dessus du locus *S* (TACG-locus *S*) ne montre pas de différences entre les méioses mâles et femelles. Le phénomène d'hétérochiasmie observé ne semble donc pas directement lié au locus *S* ou à un haplotype donné. Cependant, à partir des données dont nous disposons, il n'est pas possible d'exclure qu'il existe un rapport entre ce phénomène et la présence du locus *S* à proximité. Une augmentation de la fréquence de recombinaison, à proximité de régions où celle-ci est supprimée, a été observée dans plusieurs cas, notamment au niveau de la région pseudoautosomale des chromosomes sexuels des mammifères, ou de part et d'autre de la région chromosomique déterminant les spécificités mâles chez la papaye, ou encore du locus de compatibilité sexuelle de *Neurospora tetrasperma* (Menkis *et al.*, 2008 ; Perry et Ashworth, 1999 ; Yu *et al.*, 2009). Cependant, les données de cartographie génétique à elles seules ne permettent pas savoir si les fréquences de recombinaison observées pour les méioses mâles (ou femelles) sont au dessus ou en dessous de la fréquence estimée pour l'ensemble du génome. Ce point de discussion sera repris dans la partie II, en intégrant les informations apportées par la carte physique (*cf.*

18. Définitions d'après Hedrick (2007) :

- Sélection neutre : seule la fréquence moyenne de recombinaison est soumise à sélection.
- Sélection sexuelle : chez les mâles, elle favorise fortement la préservation des combinaisons favorables, entraînant une recombinaison plus faible.
- Sélection haploïde : le sexe étant soumis à la sélection haploïde la plus forte présente la fréquence de recombinaison la plus faible.
- Pléiotropie : la sélection limitant la recombinaison entre les différents chromosomes sexuels entraîne une diminution de la fréquence de recombinaison des autosomes chez les individus hétérogamétiques.

discussion p.173).

Associé à la suppression de la recombinaison dans la région du locus S , on s'attend à trouver un fardeau génétique récessif lié au locus S (Glemin *et al.*, 2001 ; Uyenoyama, 2005), comme cela a été mis en évidence chez *Solanum carolinense* (Stone, 2004) ou *Arabidopsis halleri* (Llaurens *et al.*, 2009).

La construction des cartes a effectivement permis d'observer, dans deux populations, des phénomènes significatifs de distorsion de ségrégation où l'haplotype S_3 était sous représenté. Le fait de disposer de plusieurs croisements dont le génotype des parents est identique dans cette région a permis de conclure, dans le cas de cet haplotype, que ce phénomène de distorsion n'était pas dépendant du génotype dans la région du locus S . Il est donc possible que l'haplotype S_3 soit lié à un locus également présent sur le groupe de liaison 2 mais distant du locus S et soumis à sélection.

Les croisements H et M ayant permis la production d'individus homozygotes S_2S_2 , l'étude de la ségrégation des allèles S pour ces croisements devait potentiellement nous permettre d'observer l'expression d'un fardeau lié à l'haplotype S_2 , mais aucune distorsion de ségrégation n'a été associée au génotype S_2S_2 . La croissance, le développement et la survie des génotypes S_2S_2 n'a pas montré de différences par rapport aux autres génotypes. De plus, pour les croisements réalisés avec ces individus, le nombre de graines obtenues lors des croisements compatibles était similaire à celui obtenu pour des individus hétérozygotes. La première hypothèse pouvant expliquer ce résultat est l'absence de fardeau lié au locus S chez la chicorée. Cependant, l'haplotype S_2 est le plus récessif des 4 haplotypes présents dans nos populations. A l'échelle plus large de la population Koospol dont sont issus les individus K28 et K59, si S_2 est également un des haplotypes les plus récessifs, cet haplotype se retrouve fréquemment à l'état homozygote et le fardeau lié au locus S peut alors être purgé (Llaurens *et al.*, 2009). Enfin, il est également possible qu'un fardeau lié à l'haplotype S_2 existe, mais que celui-ci ne s'exprime pas dans un environnement contrôlé, tel que celui des serres.

Deuxième partie

Carte physique du locus S

Chapitre 6

Vers le clonage positionnel du locus S

La construction d'une carte génétique fine a permis de restreindre la fenêtre autour du locus S à 0,6 cM pour les méioses femelles et 1,24 cM pour les méioses mâles. L'étape suivante du clonage positionnel consiste à mettre en relation la carte génétique avec une carte physique afin d'isoler la région couvrant le locus S . Pour cela, il est nécessaire de disposer de ressources génomiques correspondant à des fragments d'ADN de grande taille afin d'entamer une marche chromosomique vers le locus S . Au début de ce travail, les ressources génomiques disponibles chez la chicorée étaient uniquement constituées de banques d'EST (*cf.* Matériels et méthodes p. 65), la construction de banques BAC a donc été programmée dès le début du projet de clonage positionnel. Une fois cette ressource génomique produite, nous avons pu entreprendre la marche chromosomique vers le locus S .

6.1 Construction de banques BAC de Chicorée

Parmi les caractéristiques communes aux locus S des différentes familles autoincompatibles, une diversité de séquence et de structure a été observée entre haplotypes S , diversité que l'on s'attend donc à observer chez la chicorée. Une fois atteinte la région du locus S , la comparaison de différents haplotypes devrait permettre de restreindre la recherche des déterminants mâles et femelles à la région la plus divergente. Deux banques BAC ont donc été construites à partir de 2 génotypes de *C.intybus* différents au locus

S mais également pour d'autres traits étudiés au laboratoire et liés à la reproduction, comme par exemple la stérilité mâle nucléaire (locus NMS_1). La construction des banques BAC et des macroarrays ainsi que les procédures de criblage sont détaillées dans l'article présenté dans l'annexe C (Gonthier *et al.*, 2010). Seules les principales caractéristiques de ces banques sont donc présentées dans ce chapitre (Tableau 6.1).

La banque CinS2S2 a été construite à partir d'un individu hermaphrodite (NMS_1NMS_1), C15, homozygote au locus S (S_2S_2) ainsi que pour tous les marqueurs compris entre TACG et EU12G02. La banque CinS1S4 a été construite à partir d'un individu mâle-stérile (nms_1nms_1), C30.01, hétérozygote au locus S (S_1S_4) ainsi que pour tous les marqueurs compris entre TACG et EU12G02. Un total de 89 088 clones, contenant des inserts de 90 kb en moyenne, a été obtenu pour la banque CinS2S2. Pour la banque CinS1S4, 81 408 clones contenant des inserts de 120 kb en moyenne ont été obtenus. Ces 2 banques correspondent à une couverture théorique de plus de 12 fois du génome haploïde. Les contaminations en séquences d'ADN mitochondrial ou chloroplastique correspondent aux gammes de pourcentages communément observées lors de la construction de banques BAC (Tableau 6.1). Le criblage des banques avec 2 sondes correspondant à des gènes copies uniques a permis de valider de façon empirique la couverture de 12 X du génome haploïde.

Ces 2 banques représentent donc une ressource génomique de qualité, afin d'étudier le génome de la chicorée et plus particulièrement afin d'identifier les déterminants de l'autoincompatibilité par clonage positionnel.

6.2 Marche chromosomique

6.2.1 Stratégie mise en place afin de réaliser la marche chromosomique

Une approche systématisée a été mise en place pour chacune des sondes utilisées (Figure 6.1). La première étape correspondait à l'hybridation des sondes sur les macroarrays. Les clones hybridés ont été validés par PCR avec les amorces utilisées pour la production des sondes. Les BES (BAC Ends Sequences) de ces clones positifs ont été séquencés puis analysés (*cf.* Matériels et méthodes p. 63) afin d'identifier la présence potentielle de gènes, d'éléments transposables ou de séquences sans homologie. Des amorces ont ensuite été définies sur ces BES, hors des régions contenant des éléments transposables (*cf.* Matériels

TABLEAU 6.1 – Caractéristiques des banques BAC de chicorée d’après Gonthier *et al.* (2010, Annexe C).

Banque	CinS2S2	CinS1S4	Total
Génotype au locus <i>S</i>	S_2S_2	S_1S_4	
Génotype au locus <i>NMS</i> ₁	NMS_1NMS_1	nms_1nms_1	
Nb de clones	89 088	81 408	170 496
Nb de plaques 384	232	212	444
Taille moyenne des inserts (kb)	91,78	118,20	104,40
Clones vides (%)	0,77	3,07	1,87
Clones mtDNA (%)	0,02	0,03	0,03
Clones ptDNA (%)	1,00	2,35	1,65
Probabilité de trouver une séquence	0,9969	0,9987	0,9999
Couverture du génome haploïde	5,74 X	6,50 X	12,26 X
Nombre de clones positifs pour <i>CiSTM</i>	5	7	12
Nombre de clones positifs pour <i>CiAGP</i>	2	11	13

TABLEAU 6.2 – Exemple de construction d’un contig à partir des clones obtenus pour la sonde TTAA. La présence (1) ou l’absence (0) de produit d’amplification des BES (Si39t7 à Si37m13) a permis de positionner les clones les uns par rapport aux autres. Le contig a été construit en considérant que les BES d’un clone devaient impérativement être positionnés aux extrémités de ce clone, une fois le contig assemblé.

Clones	Si39t7	Si38m13	Si36t7	Si35m13	TTAA	Si39m13	Si38t7	Si35t7	Si36m13	Si37m13
Si35	0	0	0	1	1	1	1	1	0	0
Si36	0	0	1	1	1	1	1	1	1	0
Si37	0	0	0	1	1	1	1	1	1	1
Si38	0	1	1	1	1	1	1	0	0	0
Si39	1	1	1	1	1	1	0	0	0	0

et méthodes p. 51) et autour des régions introniques des gènes identifiés, dans la mesure ou cela s’avérait possible. Ces amorces ont d’abord été utilisées afin de créer des contigs à partir des clones positifs pour une sonde donnée. Pour cela, des PCR ont été réalisées sur tous les clones obtenus pour une même sonde avec tous les couples d’amorces définis sur les BES correspondants (*cf.* Matériels et méthodes p. 63). Ces clones devant être au moins partiellement chevauchants, la présence ou l’absence d’amplification a permis de les positionner les uns par rapport aux autres et de construire des contigs (Exemple Tableau 6.2).

Les amorces définies sur des BES ont également été utilisées pour chercher du polymorphisme (*cf.* Matériels et méthodes p. 52), afin de placer les séquences correspondantes

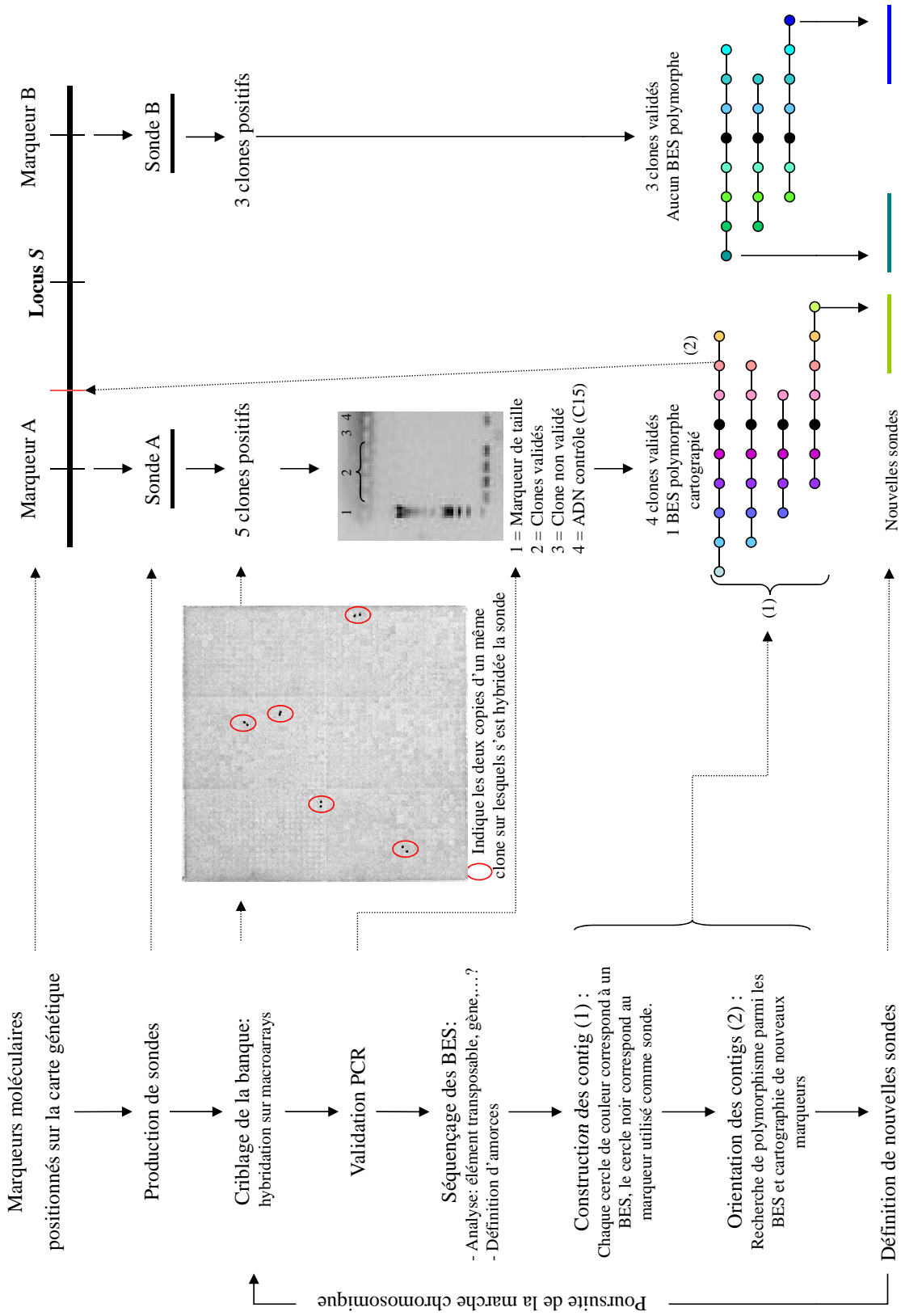


FIGURE 6.1 – Différentes étapes de la marche chromosomique vers le locus *S*.

sur la carte génétique et d'orienter les contigs. Ainsi le BES situé à l'extrémité du contig orientée vers le locus S devait être utilisé pour produire une nouvelle sonde. Dans le cas où le contig ne pouvait pas être orienté, de nouvelles sondes ont été produites en utilisant les BES des 2 extrémités du contig .

6.2.2 Criblage des banques BAC

Dans un premier temps, le criblage des 2 banques BAC en parallèle avait été envisagé, afin d'avancer simultanément le long des chromosomes portant les différents haplotypes S (Tableaux 6.3 et 6.4). Cependant, dès les premières hybridations, nous avons été confronté à des difficultés d'alignement des clones identifiés à partir de la même sonde et correspondants à des haplotypes différents. Par exemple, 5 clones positifs ont été obtenus sur la banque CinS1S4 à partir de la sonde TACG. Trois de ces clones portaient l'allèle a et 2 l'allèle c du marqueur TACG, l'allèle a étant lié à l'haplotype S_1 chez l'individu C30.01 et l'allèle c à l'haplotype S_4 . Des amorces ont été définis sur 9 des 10 BES de ces clones et la présence/absence des produits d'amplification correspondants a été testée pour chacun des clones (Tableau 6.5). La construction d'un contig à partir des clones portant le même allèle TACG n'a pas présenté de difficulté. Par contre, les positions relatives des marqueurs BES les uns par rapport aux autres se sont avérées différentes pour les contigs correspondant à différents haplotypes S . Le même problème s'est présenté avec la sonde EU12G02, lorsque nous avons confronté les résultats obtenus pour les contigs des clones liés à l'haplotype S_1 à ceux de l'haplotype S_2 . Dans un souci d'efficacité, nous avons donc choisi de nous concentrer sur la banque CinS2S2. Celle-ci ayant été construite à partir d'un individu homozygote dans la région du locus S , la construction des contigs a été rendue plus aisée. Les résultats obtenus à partir de la banque CinS1S4 (Tableau 6.4) ne seront donc pas développés par la suite.

Les premières sondes utilisées pour cribler la banque CinS2S2 ont été produites à partir des marqueurs moléculaires suivants : TACG, tgct160, EU08B02, TCAA, TTAA, AACC et EU12G02 (Tableau 6.3, Figure 6.3). La sonde TACG a permis d'identifier 1 seul clone sur lequel le marqueur GGAT a été amplifié, mais pas le marqueur tgct160 qui coségrège également avec TACG. Du fait de sa présence sur ce clone, GGAT n'a pas été utilisé comme sonde car devant être très proche physiquement de TACG. Quatre clones ont été obtenus pour la sonde EU12G02, 5 pour TTAA et 8 pour AACC. Aucun clone positif n'a été obtenu pour le marqueur EU08B02, ni sur la banque CinS2S2, ni sur

Chapitre 6. Vers le clonage positionnel du locus *S*

TABLEAU 6.3 – Résultats des hybridations (Hyb) et des validations PCR (Val) sur la banque CinS2S2. Le nom des clones correspond au numéro de plaque de la banque et à la position du clone dans sur cette plaque. Les noms abrégés (Abr) sont ceux utilisés pour désigner les clones de façon usuelle. Dans le cas des sondes définies à partir des BES (Si37m13 et Si39m13), seuls les nouveaux clones identifiés ont été indiqués. La présence (P) ou l'absence de polymorphisme (np) a été indiquée pour chaque BES. L'identification d'un gène a également été signalée. Lorsqu'un BES polymorphe a pu être placé sur la carte génétique, celui-ci a été surligné en jaune. Le symbole - correspond a une absence de données.

Locus	Sondes	Hyb	Val	Clone	Abr	BES T7	BES M13			
TACG	TACG	2	1	168F16	Si10	np	np			
	Si10m13		0	-	-	-	-			
	Si10t7		1	120A19	Si58	np	np			
tgct160	tgct160	113	40	-	-	-	-			
EU08B02	EU08_294 et EU06B03	0	0	-	-	-	-			
EU12G02	EU12_380 et EU12_160	7	4	68C01	Si20	np	np			
				138M05	Si23	np	np			
				198N16	Si25	np	np			
				216D03	Si26	np	np			
TCAA	TCAA	aspécifique	-	-	-	-	-			
TTAA	TTAA	5	5	38M16	Si35	gène, P	P			
				101N05	Si36	np	gène, np			
				110H01	Si37	np	P			
				187N10	Si38	P	np			
				209G11	Si39	np	P			
				Si37m13	11	11	27M15	Si47	np	P
							32H09	Si48	np	gène, np
							57M13	Si49	np	np
							59P05	Si50	P	np
							70H14	Si51	np	np
							85D17	Si52	np	gène, P
							89H16	Si53	-	P
							171G13	Si54	np	np
							181K09	Si55	P	np
							195H15	Si56	np	P
				197J22	Si57	np	np			
				Si39t7	22	2	29J16	Si40	np	np
144A21	Si42	np	np							
34G12	Si27	np	np							
AACC	AACC	8	8	29O21	Si28	P	np			
				57N02	Si29	P	P			
				75C24	Si30	P	np			
				103E11	Si31	P	np			
				106C17	Si32	np	np			
				128K14	Si33	np	np			
				168B06	Si34	np	np			

6.2. Marche chromosomique

TABLEAU 6.4 – Résultats des hybridations (Hyb) et des validations PCR (Val) sur la banque CinS1S4. L’haplotype (Hapl), auquel est associé le clone, et l’allèle (All) amplifié, correspondant au marqueur utilisé comme sonde, sont indiqués. Le nom des clones correspond au numéro de plaque de la banque et à la position du clone dans cette plaque. Les noms abrégés (Abr) sont ceux utilisés pour désigner les clones de façon usuelle. La présence (P) ou l’absence de polymorphisme (np) a été indiquée pour chaque BES. L’identification d’un gène a également été signalée. Le symbole - correspond à une absence de données.

Locus	Sondes	Hyb	Val	Hapl.	All.	Clone	Abr	BES T7	BES M13
TACG	TACG	5	5	S_1	a	38D10	Si2	np	np
				S_1	a	79M06	Si3	np	np
				S_1	a	116C02	Si7	np	-
				S_4	c	17F01	Si1	np	np
				S_4	c	169C12	Si9	gène, P	P
EU08B02	EU08_294 et EU06B03	4	0	-	-	-	-	-	-
EU12G02	EU12_380 et EU12_160	4	3	S_1	f	87F23	Si15	np	np
				S_1	f	159J09	Si16	np	np
				S_1	f	179C24	Si17	np	np

TABLEAU 6.5 – Comparaison des contigs obtenus pour les clones associés aux allèles a et c du marqueur TACG, pour la banque CinS1S4. L’ordre des marqueurs BES (indiqués sur la première ligne de chaque tableau) les uns par rapport aux autres n’est pas le même pour chaque contig ce qui a rendu la construction d’un contig unique impossible. Quelques marqueurs ont été reliés d’un contig à l’autre pour souligner ces différences.

		Marqueurs BES										
		Clones	Si9t7	Si3m13	Si1m13	Si2m13	TACG	Si7t7	Si9m13	Si3t7	Si1t7	Si2t7
Haplotype S_4	Si1		0	0	1	1	c	1	1	1	1	0
	Si9		1	1	1	1	c	1	1	0	0	0
				⋮								
Haplotype S_7	Si2		0	0	0	1	1	1	a	1	0	0
	Si3		0	1	1	1	1	1	a	1	1	0
	Si7		0	0	0	0	1	1	a	1	1	1

la banque CinS1S4. La sonde correspondant au marqueur tgct160 s’est hybridée à 113 clones dont 40 ont été validés par PCR. Cette sonde a été définie à partir de la séquence du marqueur AFLP et l’amplification de l’ADN génomique à partir d’amorces internes au fragment n’avait pourtant révélé qu’une seule bande sur gel d’agarose. Toutefois, le nombre de clones obtenus même après validation PCR n’était pas cohérent avec la couverture de la banque CinS2S2, cette séquence devait donc correspondre à une séquence répétée dans le génome de la chicorée. La marche chromosomique devant être réalisée à

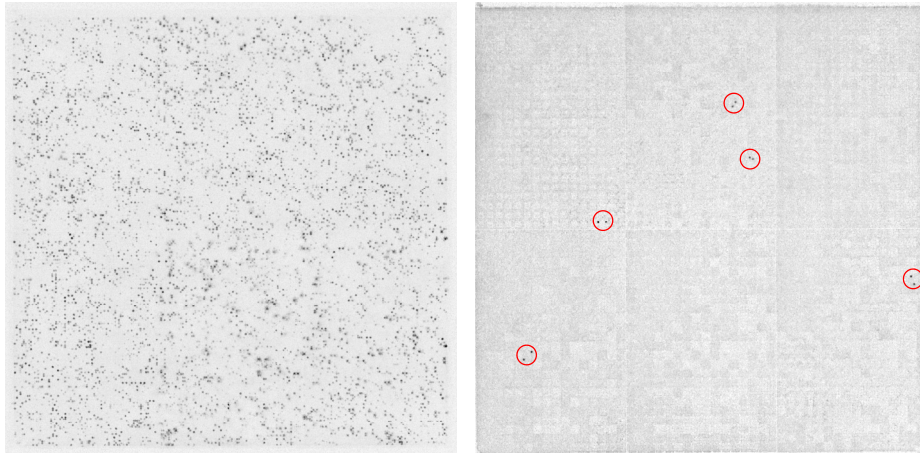


FIGURE 6.2 – Macroarrays. Exemple d’une hybridation aspécifique sur le filtre de gauche (sonde TCAA) et d’une hybridation spécifique sur le filtre de droite (sonde TTAA), avec les 2 copies de chacun des 5 clones hybridés entourées en rouge.

partir de sondes dont la position aux abords du locus S est certaine, ces clones n’ont donc pas été utilisés. L’hybridation de la sonde TCAA s’est révélée aspécifique et donc inexploitable (Figure 6.2). Un profil multibande avait été obtenu à partir de l’amplification PCR sur l’ADN génomique (*cf.* chapitre 3 p.96), nous n’avons donc pas été surpris d’obtenir un grand nombre de clones. Cependant, du fait de sa colocalisation avec le locus S nous avons tout de même tenté l’hybridation.

De plus, 4 sondes ont été définies à partir de BES (Tableau 6.3, Figure 6.3). Deux sondes ont été définies à partir des BES du clone Si10; 1 seul clone (Si58) a été validé pour la sonde Si10t7 et aucun à partir de la sonde Si10m13. Les 2 autres sondes, Si39t7 et Si37m13, ont été définies à partir des BES placés aux extrémités du contig de clones identifiés avec la sonde TTAA. La première a permis d’identifier 2 nouveaux clones (Si40 et Si42) et la seconde 11 nouveaux clones (de Si47 à Si57).

6.2.3 Caractéristiques des BES

Les BES de chacun des clones validés et conservés pour la marche chromosomique ont été séquencés. Sur un total de 32 clones obtenus à partir de la banque CinS2S2, une séquence exploitable a été obtenue pour 63 BES, avec une longueur moyenne de 835 pb. Parmi ces séquences, 9 se sont avérées identiques au BES d’un autre clone appartenant au même contig, indiquant des extrémités de clones communes. Un total de 54 BES

différents a été obtenu, ceux-ci étant composés de 11,7% d'éléments transposables, 1,2% de microsatellites et 2,1% de régions de faible complexité. De plus, 4 gènes ont été identifiés (résultats développés plus en détails dans le chapitre 7). Ces analyses ont permis de définir des amorces hors des régions contenant des éléments transposables ou autour des introns, lorsque cela était possible. Les BES composés uniquement d'éléments transposables ont été automatiquement écartés pour la production de nouvelles sondes. Des amorces ont cependant été définies pour la construction du contig.

Du polymorphisme a été observé entre les produits d'amplification de 16 marqueurs définis sur les BES. Deux de ces marqueurs ont été définis à partir de séquences de gènes et 14 à partir de séquences ne présentant aucune homologie avec des séquences connues (Tableau 6.3). Cependant, les individus recombinants de la carte haute résolution n'ont pu être génotypés sans ambiguïté que pour 2 marqueurs : Si35t7 (2 allèles codominants) et Si30t7 (3 allèles codominants), les autres présentant du polymorphisme de type présence/absence ou n'étant pas exploitables. Le marqueur Si35t7 a colocalisé avec le marqueur TTAA. Le marqueur Si30t7 a été cartographié entre les marqueurs aacc134 et TTAA : 2 événements de recombinaisons ont été observés entre le marqueur Si30t7 et le marqueur aacc134 dans la population M. Le manque d'information apporté par le marqueur aacc134, en raison de son caractère dominant, n'a pas permis de déterminer le génotype de tous les individus de la population M à ce locus (*cf.* chapitre 4 p.114). Pour cette raison, la distance entre Si30t7 et aacc134 n'a pas pu être calculée. Cependant, l'identification de recombinants nous a permis de montrer la progression de la marche chromosomique.

6.2.4 Construction des contigs liés à l'haplotype S_2

Des contigs correspondant aux sondes TACG, TTAA, AACC et EU12G02 ont été construits à partir des amorces définies sur les BES (Tableau 6.2, Figure 6.3). Dans le cas de la sonde TACG, 1 seul clone (Si10) a été identifié et a lui même permis d'en identifier un second (Si58), ces 2 clones formant un contig qui n'a pas pu être orienté par rapport à la carte génétique. Le marqueur Si30t7, positionné entre aacc134 et TTAA a permis d'orienter le contig associé à la sonde AACC. Au contraire, Si35t7, le seul marqueur polymorphe obtenu à partir des BES du contig TTAA, n'a pas permis d'orienter ce contig puisqu'il coségrégait avec TTAA. De même, les BES des clones associés à Si37m13 et Si39t7 n'ont présenté aucun polymorphisme exploitable. Cependant, les clones obtenus

ont permis d'étendre le contig TTAA, avec à l'une des extrémité le BES Si42m13 et à l'autre extrémité le BES Si52t7. Ce dernier ayant été amplifié chez 4 clones du contig AACC, le contig TTAA a pu être relié au contig AACC pour former un supercontig dont l'orientation a été définie (Figure 6.3, *cf.* annexe B.1).

Dans la région placée au-dessus du locus S , 2 clones ont été validés dans la région correspondant au marqueur TACG et aucun à partir du marqueur EU08B02, qui a été cartographié comme le marqueur le plus proche du locus S . La marche chromosomique a donc été temporairement interrompue à partir de ces locus. Dans la région située sous le locus S , les criblages successifs de la banque CinS2S2 ont permis d'orienter le supercontig qui regroupe les clones validés pour les sondes AACC, TTAA, Si37m13 et Si39t7. Le clone Si42 est le clone dont une des extrémités est la plus proche du locus S . La marche chromosomique vers le locus S est donc actuellement poursuivie à partir de ce clone.

Chapitre 7

Caractérisation de régions génomiques au voisinage du locus *S*

La démarche de clonage positionnel permet d'atteindre la région génomique contenant le ou les gènes recherchés, dans notre cas, les déterminants mâle et femelle de l'autoincompatibilité chez la chicorée. Cette démarche est en bonne voie puisque la marche chromosomique est à présent orientée. Cependant, aucune séquence génomique de chicorée n'était disponible jusqu'à ce jour. Par conséquent, aucune information n'existait quant à la nature de son génome. Pour cette raison, il était important de commencer à le caractériser afin d'acquérir l'expérience suffisante pour identifier les déterminants de l'autoincompatibilité, une fois que le locus *S* sera atteint. Les analyses présentées dans ce chapitre avaient pour objectif de déterminer : 1) le degré de complexité des séquences et son impact sur l'assemblage, 2) la relation entre distance physique et distance génétique à proximité du locus *S*, 3) la composition des séquences étudiées et 4) comment utiliser ces différentes informations dans le cadre de la poursuite du clonage positionnel.

7.1 Séquençage des clones BAC

7.1.1 Caractéristiques générales des inserts séquencés

Les clones sélectionnés pour être séquencés ont été choisis en fonction de deux critères : 1) chaque contig construit lors de la marche chromosomique devait être représenté, 2) la plus grande distance physique possible devait être couverte. Six clones issus de la banque CInS2S2 ont donc été sélectionnés de part et d'autre du locus *S*

(Figure 6.3 p.139). La sonde TACG n'ayant permis l'identification que du clone Si10, celui-ci a par conséquent été choisi. Pour les contigs identifiés à partir des sondes Si39t7 et EU12G02, 2 clones (Si42 et Si26) pour lesquels le plus de BES avaient été amplifiés, c'est-à-dire les plus représentatifs de ces contigs, ont été séquencés. Les contigs construits avec les clones obtenus à partir des sondes TTAA, Si37m13 et aacc134 se sont avérés chevauchants. Les clones sélectionnés pour chacun de ces contigs (Si27, Si37 et Si52) devaient alors permettre d'obtenir une séquence allant au minimum du marqueur TTAA au marqueur aacc134 et rendre possible une première estimation de la relation entre la distance génétique et la distance physique. Un 7^e clone, 60A10, également obtenu à partir de la banque CinS2S2, a été séquencé dans le cadre de l'étude de la stérilité mâle nucléaire (*cf.* chapitre 2 p.75). Le locus *NMS*₁ associé à ce trait a été cartographié sur le groupe de liaison 5 de la carte génétique de la chicorée. Ce clone, qui contient le gène potentiellement responsable de la stérilité, a permis d'avoir un point de comparaison avec une autre région du génome qui n'est pas soumise aux mêmes contraintes que celles attendues pour le locus *S*.

En moyenne, chaque clone de la région du locus *S* a été assemblé en 14 contigs (*cf.* Matériels et méthodes p. 63), avec un minimum de 7 contigs pour le clone Si27 et un maximum de 23 contigs pour le clone Si26 (Tableau 7.1). Des résultats similaires ont été obtenus pour le clone 60A10, puisqu'il compte 17 contigs. Par la suite, l'abréviation Cg remplacera le terme de contig, quand celui-ci est associé aux séquences des clones, afin de le distinguer des contigs construits pour la marche chromosomique. Le nombre de Cg obtenus chez la chicorée, à l'issue de l'étape d'assemblage, est supérieur à celui obtenu pour des clones BAC d'autres espèces étudiées au CNRGV (canne à sucre, blé, piment...), séquencés dans les mêmes conditions (Hélène Bergès et Sonia Vautrin, communications personnelles).

Pour le clone Si26, qui présentait le plus grand nombre de Cg (assemblage Newbler), des tests ont été réalisés au laboratoire avec un autre logiciel (CAP3), en faisant varier les différents paramètres d'assemblage. Dans les conditions optimales, le second assemblage s'est avéré similaire au premier, en terme de nombre de Cg et de taille totale de séquence obtenus. Par contre, dans la majorité des cas, la longueur, les points d'initiation et d'interruption de séquence de chaque Cg se sont avérés différents d'une méthode à l'autre (Sylvain Legrand, communication personnelle). Les difficultés d'assemblage rencontrées chez la chicorée ne semblent donc pas dues à un problème méthodologique mais seraient plutôt liées à la nature même des séquences. Les assemblages réalisés par le CNRGV et

ses collaborateurs ont donc été utilisés pour les analyses présentées par la suite.

Des motifs microsatellites ont été observés sur la moitié des extrémités des Cg, ce qui pourrait en partie expliquer pourquoi il a été difficile de réaliser l'assemblage des séquences. Les tailles des inserts obtenues après migration sur gel d'électrophorèse en champ pulsé et après assemblage des séquences se sont révélées du même ordre de grandeur, avec par exemple, respectivement 105 et 98 kb pour le clone Si10 (Tableau 7.1). Cela indique que malgré le nombre important de Cg obtenus, avec une longueur minimale de 500 pb, la quasi totalité de la séquence a été obtenue pour chacun des clones. De plus, la taille moyenne des inserts calculée à partir des 6 clones de la région du locus *S* (95,62 kb) s'est avérée équivalente à celle calculée pour l'ensemble de la banque CinS2S2 (91,78 kb, *cf.* tableau 6.1 p.133). La taille du clone 60A10 s'est quant à elle révélée en dessous de la moyenne de la banque avec une longueur de 55,62 kb, pour une taille estimée sur gel de 55 kb. Pour la région du locus *S*, la taille des séquences obtenues varie de 86,6 kb (Si27) à 105,1 kb (Si52), pour un total de 573,7 kb. Une fois les régions chevauchantes des clones Si37, Si52 et Si27 déduites, la région réellement couverte par nos séquences s'est avérée être de 546,2 kb. La taille du supercontig a été estimée à au moins 360,6 kb, en sommant la taille des séquences des 3 clones chevauchants (273,9 kb) et du clone Si42 (86,6 kb), qui constitue une des extrémités du supercontig (*cf.* figure 6.3 p.139).

TABLEAU 7.1 – Caractéristiques des inserts issus de la banque BAC CinS2S2. Pour chaque clone, le groupe de liaison et la sonde à partir de laquelle il a été identifié sont indiqués (LG_sonde). Les clones Si37, Si52 et Si27 (grisés) sont chevauchants. Les moyennes et totaux ont été calculés pour les 6 clones positionnés sur le LG2 autour du locus *S*. (a) Taille de l'insert estimée sur gel d'électrophorèse et (b) longueur totale après assemblage. (c) Nombre de Cg d'au moins 500 pb. (d) Total ne considérant qu'une seule fois les régions chevauchantes des clones Si37, Si52 et Si27. Le symbole - correspond aux données manquantes.

Région	Clone	kb ^(a)	kb ^(b)	Nb de Cg ^(c)	Extrémités SSR (%)	Cg moyen (kb)	Cg le plus long (kb)
LG2_TACG	Si10	105	98,03	17	67,65	5,77	16,53
LG2_Si39t7	Si42	90	86,66	13	61,54	6,67	25,80
LG2_TTAA	Si37	110	105,11	10	75,00	10,51	57,66
LG2_Si37m13	Si52	110	101,74	16	34,38	6,36	14,39
LG2_AACC	Si27	-	94,54	7	42,86	13,51	36,47
LG2_EU12G02	Si26	100	87,63	23	34,78	3,81	12,07
Moyenne		103	95,62	14,33	50,58	7,77	
Total		-	546,19 ^(d)				
LG5_NMS ₁	60A10	55	55,62	17	38,24	3,27	8,09

7.1.2 Organisation des contigs

7.1.2.1 Validation et assemblage

Avant d'étudier en détail les séquences des clones de la région du locus *S*, nous avons vérifié qu'il s'agissait bien des séquences attendues et essayé de positionner un maximum de Cg les uns par rapport aux autres (Figure 7.1). Pour cela, des Blastn ont été réalisés à partir des séquences des Cg de chaque clone :

- avec la séquence de la sonde à partir de laquelle a été identifié chaque clone,
- avec les BES de tous les clones appartenant au même contig.

Les 6 clones ont été validés et, pour chacun d'entre eux, au moins 2 Cg ont été positionnés et orientés aux extrémités, à partir de l'alignement de la séquence des BES propres à chaque clone. De plus, les séquences des BES des clones qui n'ont pas été séquencés ont été alignés avec les séquences des clones séquencés. Ainsi, tous les Cg contenant au moins un BES ont pu être positionnés les uns par rapport aux autres. Les Cg contenant plusieurs BES ont également pu être orientés. Cependant, il n'a pas été possible de déterminer si ces Cg contenant des BES étaient adjacents ou non, en raison des Cg qui n'ont pas pu être positionnés.

En utilisant cette méthode, pour le clone Si10, seuls les Cg portant les BES du clone ont pu être placés aux extrémités. Les Cg C08 et C11 ont toutefois pu être placés de part et d'autres de C07 grâce aux Blastn et Blastx réalisés par la suite (Figure 7.4). De plus, cette démarche a permis de positionner les marqueurs TACG et GGAT sur un même Cg (C07) à une distance de 5 430 pb l'un de l'autre. Cela a confirmé que ces 2 marqueurs étaient très proche physiquement, comme cela avait été suggéré dans le chapitre 6 (p.135), car aucun évènement de recombinaison n'avait été observé entre eux. Six Cg (C07, C01, C10, C03, C09 et C14), sur les 13 que compte le clone Si42, ont pu être positionnés les uns par rapport aux autres. Quatre Cg (C05, C01, C04, C03), sur les 10 du clone Si37, ont pu être organisés et le marqueur TTAA a été placé sur le Cg C05 qui contient également le BES Si37t7. Pour le clone Si52, 7 Cg (C11, C02, C18, C15, C14, C16 et C05) ont pu être positionnés grâce aux BES. Parmi eux, 2 (C14 et C15) se sont avérés adjacents, une partie du BES Si50m13 s'alignant avec une des extrémité du 1^{er} Cg, l'autre partie du BES s'alignant avec une extrémité du 2nd Cg. De plus, 2 autres Cg (C07 et C03) ont été placés bout à bout grâce aux Blastn et Blastx présentés dans les paragraphes suivants (*cf.* paragraphe 7.3 et figure 7.4). Quasiment tous les Cg du clone Si27 (6 sur 7) ont pu être positionnés les uns par rapport aux autres et le

marqueur AACC a été localisé sur le Cg C01. De plus, une distance d'environ 50 kb a été estimée entre les marqueurs Si30t7 et aacc134, respectivement positionnés sur les Cg 3 et 1. Enfin, 6 Cg (C16, C15, C02, C13, C07 et C12) sur les 23 du clone Si26 ont pu être positionnés les uns par rapport aux autres.

Pour organiser de façon plus précise les séquences de ces clones, il aurait fallu définir des amorces aux extrémités des Cg et tester toutes les combinaisons possibles par PCR. Cependant, étant donné le nombre de Cg obtenus par clone, cette démarche se serait avérée très longue. N'étant pas encore à proximité immédiate du locus *S*, l'obtention de séquences parfaitement organisées n'était pas une priorité. Nous avons donc choisi de travailler à partir de ces séquences partiellement ordonnées. Pour faciliter les analyses qui ont suivies, les Cg ont été concaténés pour ne former qu'une seule séquence par clone. Lorsqu'elles étaient connues, les positions relatives de chaque Cg ont été prises en compte pour la construction de ces séquences (Figure 7.4 p.155 à 157).

7.1.2.2 Relation entre distance physique et distance génétique

Une question importante lorsqu'on entame une démarche de clonage positionnel est de savoir quelle distance physique il va falloir couvrir pour atteindre la région d'intérêt. Le rapport entre distance physique et distance génétique est variable d'une espèce à l'autre et d'une région du génome à l'autre, au sein d'une espèce donnée (*cf.* tableau 5.1 p.121). Puisqu'une réduction de la fréquence de recombinaison est attendue dans la région du locus *S*, on s'attend à ce que, à proximité de ce locus, le rapport Mb/cM soit plus élevé que celui estimé pour l'ensemble du génome, c'est-à-dire 1,5 Mb/cM. L'organisation des séquences des 3 clones chevauchants (Si37, Si52 et Si27) a permis de positionner 2 marqueurs de la carte génétique, TTAA et aacc134 (Figure 7.1). A partir de ces données, une première estimation du rapport distance physique/distance génétique a été établie aux abords du locus *S*. Le marqueur TTAA a été positionné sur le Cg C05 correspondant à une extrémité du clone Si37. Le marqueur aacc134 a été positionné sur le Cg C01 du clone Si27, qui lui n'est pas positionné à une extrémité de séquence. Cependant, le clone Si27 est le clone pour lequel un seul Cg (C05) de 5,9 kb n'a pas pu être positionné. Ces 2 clones sont reliés par le clone Si52 et les régions chevauchantes ont été identifiées sans ambiguïté. La distance physique entre ces 2 marqueurs a donc pu être calculée et les valeurs de 236,5 Kb et 242,4 kb ont été obtenues, en fonction de la position attribuée au Cg C05 du clone Si27. Le rapport Mb/cM a été estimé, à partir des distances génétiques obtenues pour les méioses femelles (0,11 cM) et les méioses mâles

(0,47 cM), entre les marqueurs TTAA et aacc134 (*cf.* chapitre 4, figure 4.7 p.116). Des valeurs de 2,15 et 2,20 Mb/cM ont été obtenues dans le premier cas et de 0,50 et 0,52 Mb/cM dans le second cas.

En considérant que ces rapports étaient constants jusqu'au locus *S*, une première estimation de la distance physique entre TTAA et le locus *S* a été réalisée. Pour les méioses femelles, une distance physique de 0,8 Mb a été estimée, correspondant à une distance génétique de 0,36 cM. Pour les méioses mâles, une distance physique de 0,5 Mb a été obtenue, pour une distance génétique de 1 cM. A partir de ces résultats, et connaissant la taille moyenne des inserts de la banque CinS2S2 (91,78 kb) et la taille du clone Si42 (86,6 kb) qui est positionné à l'extrémité du supercontig TTAA-aacc134, nous avons estimé que l'identification de 4 à 8 clones successifs supplémentaires était nécessaire pour atteindre le locus *S*.

7.2 Composition en éléments répétés : éléments mobiles, microsatellites et autres séquences répétées

7.2.1 Elements répétés identifiés à partir de RepeatMasker et des Blastx

Afin de caractériser la composition des clones séquencés en éléments transposables et motifs répétés, le logiciel RepeatMasker a été utilisé (*cf.* Matériels et méthodes p. 64). La composition des séquences en éléments transposables a été complétée par les Blastx des séquences des clones sur la banque de peptides d'*A.thaliana* (*cf.* paragraphe 7.3). Les positions de ces éléments ont été représentés sur la figure 7.4.

Un total de 5,71% d'éléments mobiles a été obtenu pour la région génomique aux abords du locus *S*, dont 4,87% de rétroéléments et 0,74% d'éléments transposables de type II (Tableau 7.2). Parmi ces éléments, 44 ont été identifiés par RepeatMasker et 5 à partir des Blastx. De plus, 159 régions couvrant 1,74% de ces séquences ont été identifiées comme étant des microsatellites (1 SSR pour 3,6 kb), 37% de ceux-ci étant des polyTA. Cependant, le pourcentage de microsatellites a sans doute été mal estimé, en particulier pour les Cg dont les extrémités se terminent par des SSR (Tableau 7.1). Pour la même raison, le nombre total d'éléments SSR à quant à lui dû être surestimé. Chaque

extrémité de Cg se terminant par un microsatellite a été compté comme un élément mais il est probable qu'une partie de ces extrémités soient en réalité adjacentes. En supposant que toutes ces extrémités SSR soient contiguës 2 à 2, un total de 42 éléments auraient été comptés deux fois. Ce raisonnement a permis d'estimer un nombre minimum de 117 éléments microsatellites contenus dans ces séquences, soit 1 SSR pour 4,9 kb. Les séquences des clones étaient également composées de 2,75% de régions de faible complexité, dont près de 80% se sont avérées être des régions riches en TA. Ces résultats ne semblent pas spécifiques de la région du locus *S* car des valeurs du même ordre de grandeur ont été obtenues pour le clone 60A10 (Tableau 7.2).

Afin de disposer d'éléments auxquels comparer la fréquence des motifs répétés identifiés sur nos séquences, le logiciel ReapeatMasker a été utilisé pour estimer le pourcentage de microsatellites et de régions de faible complexité pour l'ensemble du génome d'*A. thaliana*. Nous avons obtenu 0,40% de SSR et 1,75% de régions de faible complexité. Dans ces conditions, les séquences de chicorée, étudiées dans la région du locus *S* et pour le clone 60A10, se sont révélées au moins 4 fois plus riches en SSR et 1,5 fois plus riches en régions de faible complexité que les séquences d'*A. thaliana*.

7.2.2 Identification d'autres séquences répétées

Pour compléter les informations fournies par ReapeatMasker et les Blastx, les séquences des 6 clones ont été comparées les unes aux autres par Blastn. Ces alignements ont permis d'identifier 150 régions répétées de plus de 100 pb avec une e-value inférieure à 10^{-30} , que ce soit intra- ou inter-clones (Tableau 7.2). Les séquences des 6 clones se sont révélées être composées de plus de 10% de ce type de séquences, longues de 380 pb en moyenne. La plus grande région dupliquée avait une taille de 5 760 pb et présentait plus de 97% d'identité entre les copies identifiées sur les clones Si26 (Cg C02) et Si42 (Cg C07) (Figure 7.4). De la même façon, 13 régions distinctes du clone 60A10 se sont révélées répétées dans au moins un des 6 clones de la région du locus *S*. Parmi ces régions, 2 particulièrement longues (5 053 pb et 1 179pb) étaient partagées avec le clone Si26, avec une identité de plus de 90%. Parmi la totalité de ces séquences, 3 se sont avérées correspondre avec un élément transposable et 5 ont présenté une homologie avec un gène ou une séquence des banques d'EST de *Cichorium sp.* (cf. paragraphe 7.3 et figure 7.4).

7.3. Etude de la fraction potentiellement codante à proximité du locus *S*

TABLEAU 7.2 – Caractérisation des abords du locus *S* en éléments répétés. Les éléments mobiles, microsatellites et régions de faible complexité, ainsi que le contenu en GC ont été déterminés grâce au logiciel RepeatMasker et aux Blastx réalisés sur la banque de peptides d’*A. thaliana*. Les régions dupliquées ont été identifiées par Blastn des séquences des 6 clones sur elles mêmes et entre elles. (a) La couverture correspond au nombre total de pb correspondant aux différents éléments présents dans les 573,7 kb que représentent les 6 clones. (b) Ce pourcentage correspond uniquement aux régions génomiques du clone 60A10 ayant été identifiées sur les 6 clones de la région du locus *S*. (c) Les résultats présentés ont été obtenus à partir de la totalité du génome d’*A. thaliana* (119 Mb, version TAIR10).

		Région du locus <i>S</i>			Clone 60A10	<i>A. thaliana</i> ^(c)	
		Nombre d’éléments	Couverture ^(a) (pb)	%	%	Nombre d’éléments	%
Retroelements	LTR elements :	30	27921	4,87	4,44		
dont :	Ty1/Copia	19	18371				
	Gypsy/DIRS1	9	8470				
	gag protein	2	1080				
Transposons		16	4228	0,74	0		
dont :	hobo-Activator	5	1017				
	En-Spm	2	78				
	MuDR-IS905	4	189				
	Tourist/Harbinger	5	2599				
	Non classés	3	637	0,11	1,38		
Total	Éléments mobiles	49	32786	5,71	5,82		
Microsatellites		159	9979	1,74	2,13	10586	0,4
Région de faible complexité		307	15792	2,75	2,89	41883	1,75
Régions dupliquées		150	57067	10,45	14,53 ^(b)		
Contenu en GC				34,88	35,53		36,04

7.3 Etude de la fraction potentiellement codante à proximité du locus *S*

Le fait de disposer de séquence entières de clones BAC devait notamment nous permettre de mettre en place une stratégie afin d’identifier les régions potentiellement codantes, c’est à dire associées à des EST et/ou des gènes annotés pour d’autres espèces. Ce type de régions potentiellement conservées devraient, à l’avenir, apporter des points d’ancrages permettant de passer d’une banque BAC à l’autre. Cette étude devait également nous informer sur la richesse en gène des séquence aux abord du locus *S* et permettre

de révéler une éventuelle conservation de la synténie entre la chicorée et d'autres espèces dont le génome est séquencé. Les gènes annotés comme des éléments mobiles ont donc été automatiquement écartés des analyses décrites ci-après et ajoutés aux résultats obtenus avec RepeatMasker (*cf.* paragraphe 7.2).

7.3.1 Comparaison des séquences de chicorée et des banques d'EST d'Asteraceae

Dans un premier temps, afin d'identifier des régions potentiellement transcrites au sein de nos séquences, des Blastn ont été réalisés sur les banques d'EST de chicorée, de *C. endivia*, de laitue (*Lactuca sativa*) et de pissenlits (*Taraxacum sp.*), qui appartiennent à la tribu des Lactuceae (sous-famille des *Cichorioideae*), ainsi que du tournesol (*Helianthus annuus*), qui appartient à une sous-famille différente (*Asteroideae*, *cf.* Matériels et méthodes p. 64). De part la proximité des séquences et du nombre important d'EST qu'elle contient, la banque de chicorée a été la plus informative et 112 séquences uniques (SU¹⁹) ont été identifiées. De plus, 6 SU ont été obtenues à partir de la banque d'EST de *C. endivia* (Figure 7.2). Les SU issues des autres banques se sont avérées redondantes avec les 118 SU du genre *Cichorium*, tout en ayant un pourcentage d'identité plus faible. Elles n'ont donc pas été prises en compte par la suite.

Les 118 SU ont été obtenues à partir de la même espèce (*C. intybus*) ou d'une espèce très proche (*C. endivia*). Nous nous attendions donc à ce qu'elles s'alignent avec nos séquences génomiques sur toute leur longueur, en dehors des régions introniques. Pourtant, dans près de 50% des cas, l'alignement observé représentait moins de 85 % de la SU (Figure 7.3), cela malgré un pourcentage d'identité supérieur à 95% et en ne considérant pas les SU potentiellement tronquées car localisées à l'extrémité des clones. Il est donc probable qu'une partie des SU identifiées par Blastn se soit alignée sur nos séquences en raison de la présence de motifs conservés. Nous avons donc choisi de ne considérer comme séquence homologue que les SU dont plus de 85% de la séquence étaient alignés avec celle de nos clones BAC. En ajoutant ce critère de sélection, nous avons réduit par 2 le nombre de SU différentes. Au final, 59 SU ont été conservées, 55 (36 contigs et 19 singletons) étant issues de la banque d'EST de la chicorée et 4 provenant de celle de *C. endivia*.

19. SU : cette abréviation est utilisée pour désigner les séquences uniques au sein des banques d'EST, c'est-à-dire, les contigs et singletons obtenus après assemblage des EST.

7.3. Etude de la fraction potentiellement codante à proximité du locus *S*

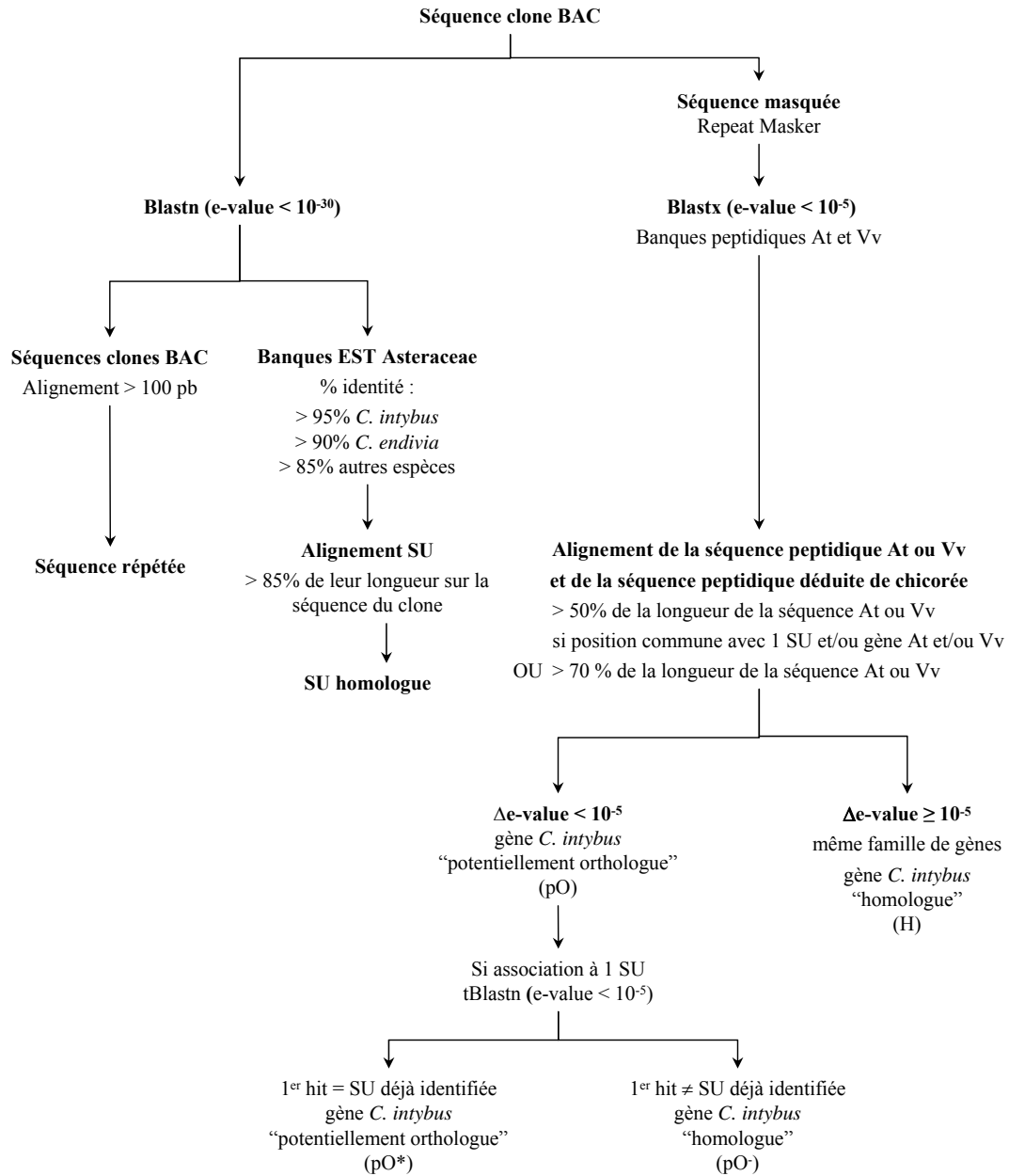


FIGURE 7.2 – Critères de sélection utilisés pour les différents types de Blast : Blastn, Blastx et tBlastn. At : *A. thaliana*, Vv : *V. vinifera*, Δe -value : rapport des e-values des 2 meilleurs résultats de Blast.

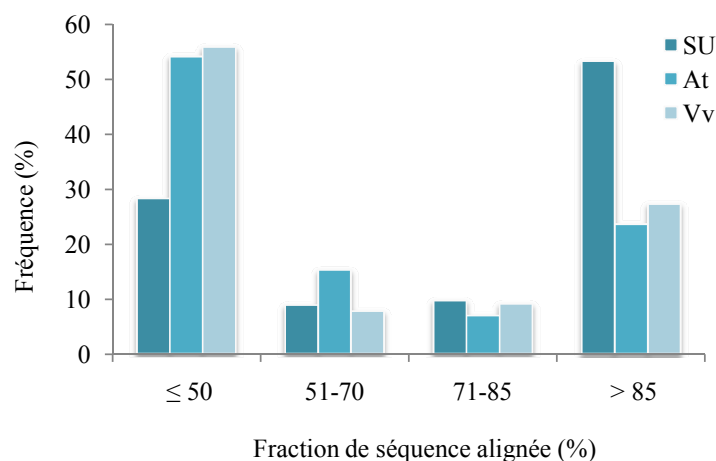


FIGURE 7.3 – Fréquences des portions de séquences des SU, des séquences peptidiques déduites des gènes d’*A. thaliana* et de vigne alignées sur les séquences des clones BAC. SU : alignement de séquences d’ADN, At et Vv : alignement des peptides d’*A. thaliana* et de *V. vinifera* sur les peptides déduits à partir des séquences des clones BAC.

7.3.2 Composition en gènes des séquences de chicorée dans la région du locus *S*

En plus de l’étude des banques d’EST, nous avons complété notre jeu de données en associant à nos séquences des gènes annotés dans d’autres espèces de Dicotylédones vraies. Dans le clade des Astéridées, le séquençage du génome d’une espèce appartenant à la famille des Asteraceae, le tournesol, (*Helianthus annuus*), vient de débuter²⁰. Les génomes de 2 espèces appartenant à la famille des Solanaceae, la pomme de terre²¹ (*Solanum tuberosum*), et la tomate²² (*Solanum lycopersicum*), sont en cours de séquençage. Bien que l’assemblage du génome de la tomate ne soit pas encore achevé, un nombre important de séquences annotées et couvrant les 12 chromosomes de ce génome sont accessibles depuis le début de l’année 2011. Le génome d’une troisième espèce (*Mimulus guttatus*), appartenant à la famille des Phrymaceae (anciennement Scrophulariaceae, Beardsley et Olmstead, 2002), a été séquencé mais ce génome est actuellement assemblé sous la forme de plus de 2000 scaffolds. Par conséquent, nous avons choisi dans un premier temps 2 espèces appartenant au clade des Rosidées dont le génome a été séquencé et assemblé en chromosomes. La première, *A. thaliana*, a été choisie parce qu’il s’agissait

20. Plant & Animal Genomes XVIII Conference, http://www.intl-pag.org/18/abstracts/W22_PAGXVIII_170.html

21. http://www.potatogenome.net/index.php/Main_Page

22. http://solgenomics.net/genomes/Solanum_lycopersicum/index.pl

7.3. Etude de la fraction potentiellement codante à proximité du locus *S*

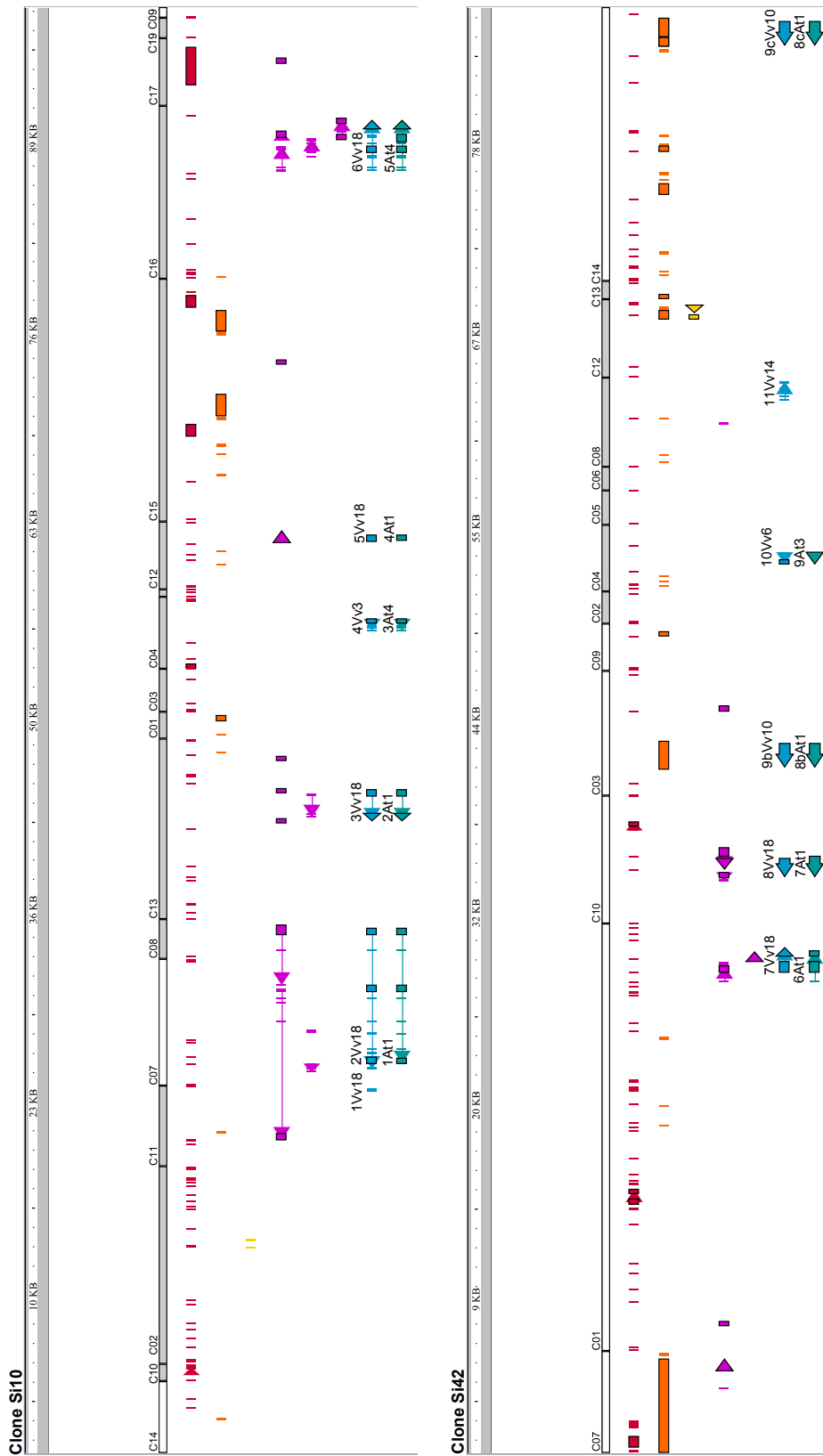


FIGURE 7.4 – Distribution des différents types d'éléments composant les séquences génomiques des 6 clones séquencés. Une échelle de taille en kb est associée à chaque clone. Dans l'ordre suivant, sont représentés les positions : des Cg ordonnés en fonction des BES (blanc) et des Cg placés de façon aléatoire (gris), des régions masquées par RepeatMasker (rouge), des régions répétées entre les 6 clones (orange) et des régions communes avec le clone 60A10 (jaune), des SU (violet), des gènes homologues d'*A.thaliana* (vert) et de *V. vinifera* (bleu). Les notations utilisées pour les gènes correspondant aux abréviations présentées dans le tableau 7.3. Suite de la figure p.156 et 157.

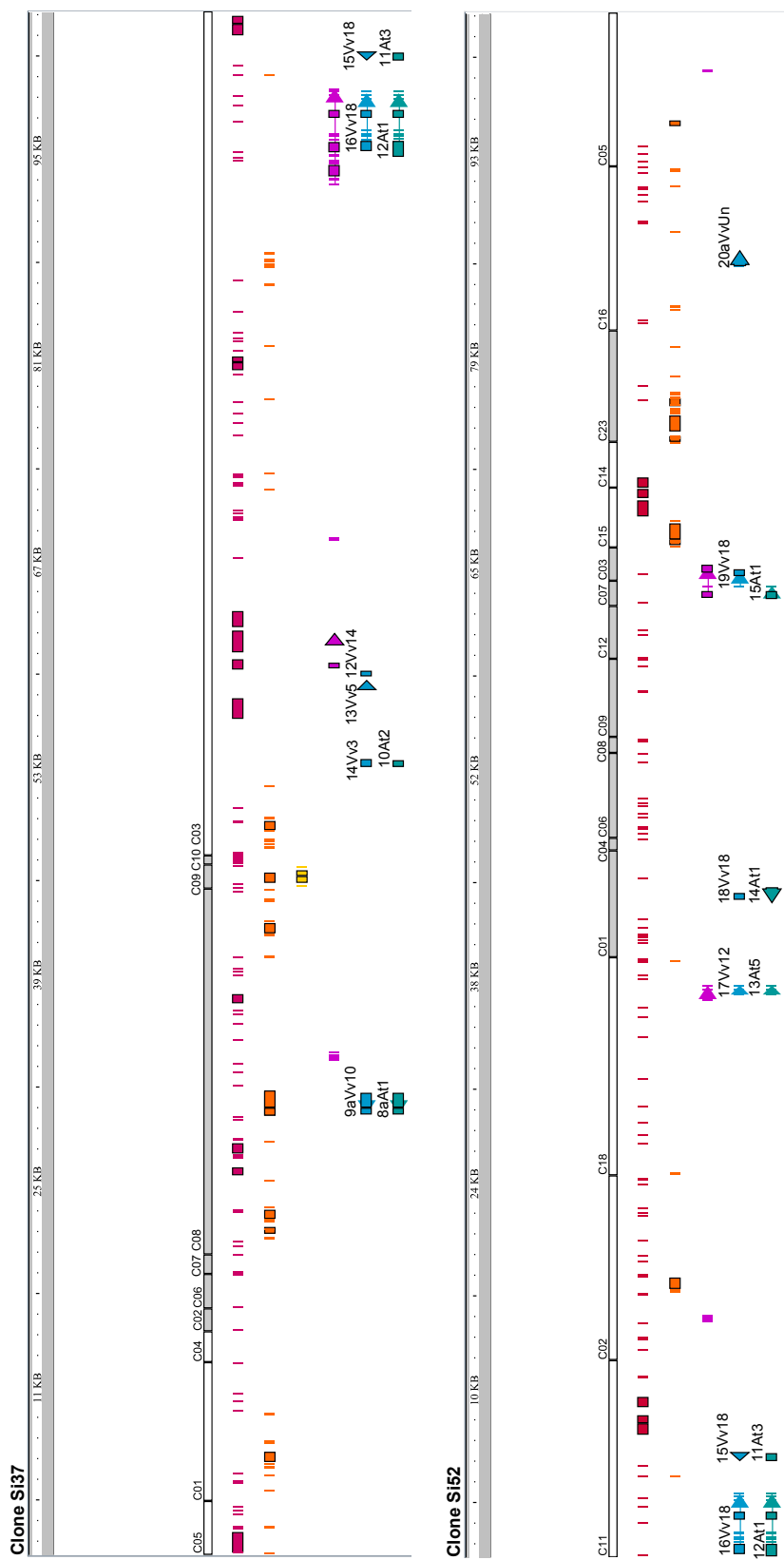


FIGURE 7.4 – Suite – Distribution des différents types d'éléments composant les séquences génomiques des 6 clones séquencés.

7.3. Etude de la fraction potentiellement codante à proximité du locus *S*

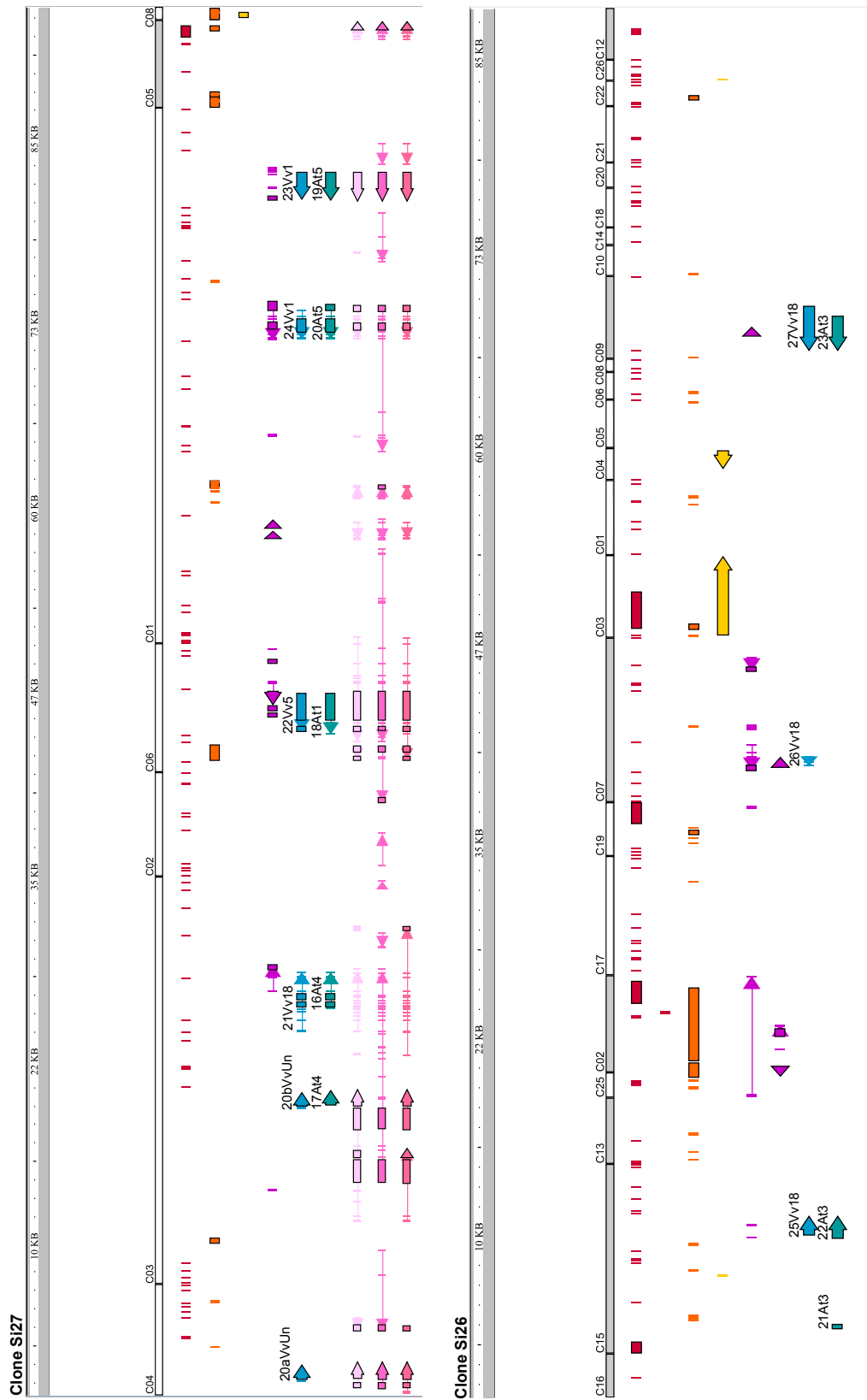


FIGURE 7.4 – Suite - Distribution des différents types d'éléments composant les séquences génomiques des 6 clones séquencés. Pour le clone Si27, exemples de prédictions de gène (en rose, avec du plus clair au plus foncé : prédictions Fgenesh puis GenScan avec *A. thaliana* comme organisme de référence, puis prédictions Fgenesh avec *V. vinifera* comme organisme de référence).

de l'espèce modèle chez les Dicotylédones et que son génome a largement été annoté. La seconde, la vigne (*V. vinifera*), a été sélectionnée car elle semblait posséder le génome le plus proche du génome ancestral des Dicotylédones et le moins dupliqué, en comparaison des génomes séquencés de différentes Rosidées (Jaillon *et al.*, 2007 ; Van de Peer *et al.*, 2009). De plus, de même que pour *A. thaliana*, le génome de la vigne est presque totalement assemblé en chromosomes et annoté.

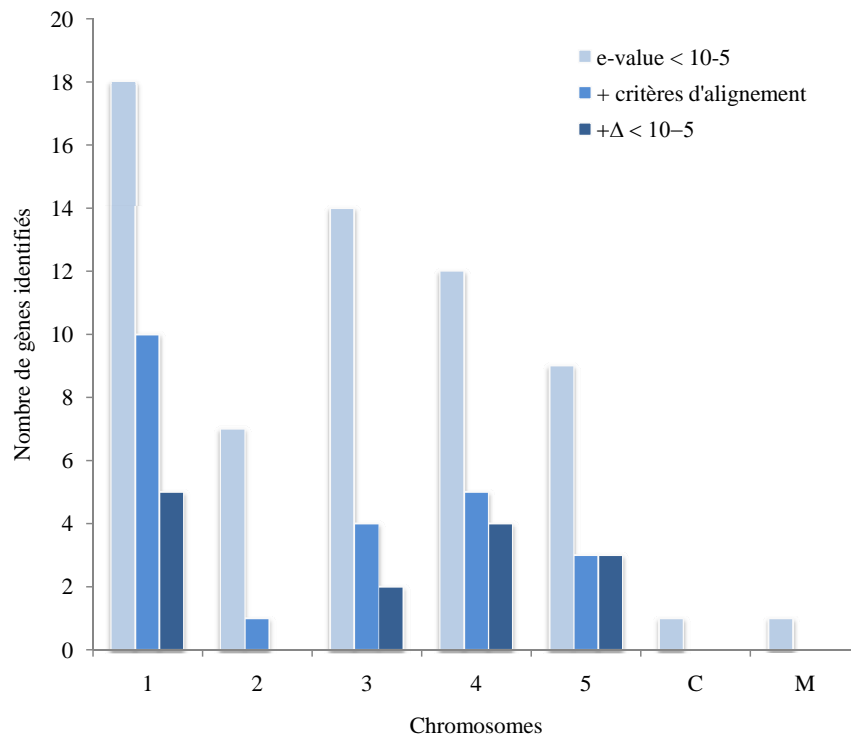
7.3.2.1 Identification de gènes homologues

Etude des banques de peptides d'*A. thaliana* et de *V. vinifera* Après avoir masqué les éléments mobiles et répétés avec le logiciel RepeatMasker, des Blastx des séquences des clones BAC ont été réalisés sur les banques de peptides d'*A. thaliana* et de la vigne. L'objectif était d'identifier des gènes homologues et parmi eux de potentiels orthologues (Tableau 7.2). En utilisant comme critère de sélection, une e-value maximum de 10^{-5} , 62 gènes ont été identifiés chez *A. thaliana* et 65 chez la vigne (Figure 7.5). Cependant, comme pour les EST, l'alignement des séquences peptidiques déduites des séquences génomiques avec celles des bases de données s'est avéré très partiel, avec environ 55% des ces séquences alignés sur moins de 50% de leur longueur (Figure 7.3). Nous avons donc défini des critères supplémentaires de sélection, afin d'éliminer les résultats dus uniquement à des motifs conservés. Dans les régions où des SU avaient été préalablement identifiées, les gènes ont été retenus lorsque leur séquence peptidique s'alignait sur plus de 50% de leur longueur avec la séquence traduite du clone. En l'absence de SU, seules les séquences peptidiques s'alignant sur plus de 70% de leur longueur ont été retenues.

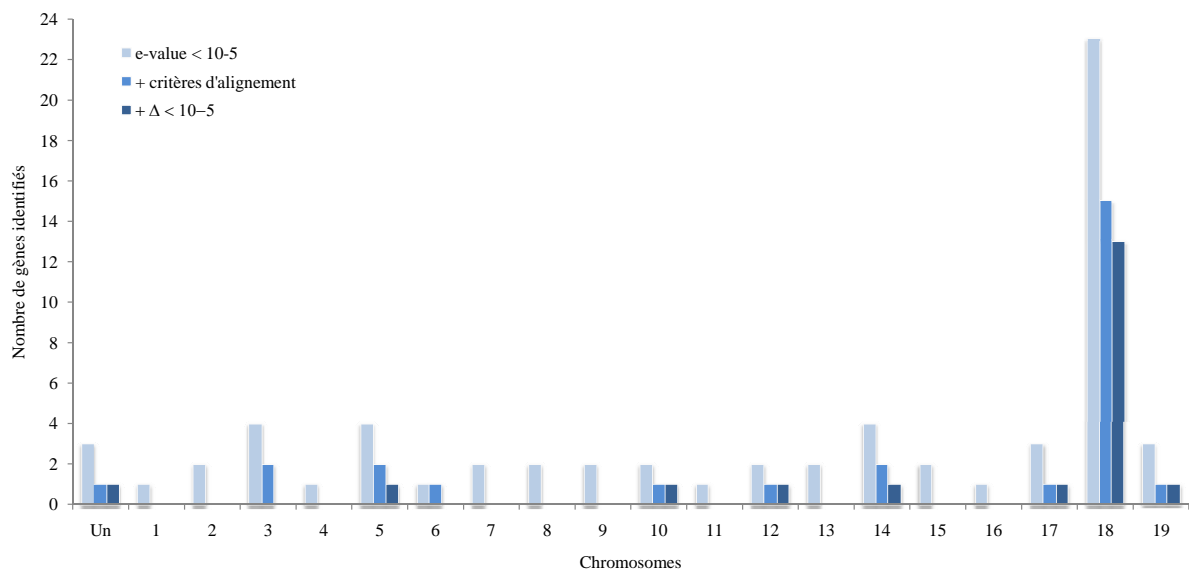
Après avoir filtré les résultats de Blastx avec ces critères supplémentaires, 23 gènes différents ont été conservés pour *A. thaliana* et 27 pour la vigne (Tableau 7.3, Figures 7.4 et 7.5). Ainsi, 15 gènes d'*A. thaliana* (60,8% des gènes identifiés) et 30 SU ont été positionnés dans des régions communes sur nos séquences. De même, 16 gènes de vigne (55,5% des gènes identifiés) et 32 SU ont été positionnées dans des régions communes sur nos séquences. Au total, près de 50% des SU ont été identifiées aux mêmes positions que des gènes d'*A. thaliana* ou de vigne.

Un Blastx a ensuite été réalisé sur les bases de données de peptides d'*A. thaliana* et de *V. vinifera* à partir des séquences des 59 SU préalablement identifiées par Blastn. Trente-trois d'entre elles (56%) ont présentées une homologie pour un gène, ce qui est un pourcentage similaire à celui obtenu pour la banque totale d'EST de *C. intybus* (*cf.*

7.3. Etude de la fraction potentiellement codante à proximité du locus *S*



(a) *A. thaliana*



(b) *V. vinifera*

FIGURE 7.5 – Distribution des gènes identifiés par Blastx : (a) sur les chromosomes 1 à 5 d'*A. thaliana* ou les génomes chloroplastique (C) et mitochondrial (M), et (b) sur les chromosomes 1 à 19 de la vigne, selon les différents critères de sélection appliqués (*cf.* figure 7.2).

Matériels et méthodes p. 65). Une partie des SU a présenté la plus forte homologie avec les gènes d'*A. thaliana* ou de vigne associés aux mêmes régions génomiques. Ainsi, 24 d'entre elles correspondaient au gène d'*A. thaliana* et 25 au gène de vigne, identifiés à position identique. Deux des 30 SU associées à des gènes d'*A. thaliana* et 5 des 32 associées à des gènes de *V. vinifera* ont présenté la meilleure homologie avec un gène différent de celui identifié à partir de la séquence génomique. De plus, 3 et 4 SU, peu ou pas chevauchantes avec des régions génomiques respectivement associées à un gène de vigne ou d'*A. thaliana*, n'ont présenté d'homologie avec aucun gène. Parmi les SU identifiées dans des régions n'ayant été associées à aucun gène d'*A. thaliana* ou de vigne, 3 ont présenté une homologie avec un gène d'*A. thaliana* et 2 avec un gène de vigne. Ainsi, parmi les SU montrant une homologie avec une séquence d'*A. thaliana* et/ou de vigne, 76% ont présenté la même homologie que la séquence génomique de chicorée avec laquelle ces SU étaient alignées.

Etude de banques de peptides supplémentaires La réalisation de Blastx à partir des séquences complètes de 6 clones BAC nous a permis d'identifier 23 gènes chez *A. thaliana* et 27 chez la vigne sur un total de 546,2 kb. Nous avons également pu constater que 56% des SU présentaient une homologie avec un gène d'au moins une de ces 2 espèces. Afin de déterminer si d'autres espèces pouvaient nous apporter un supplément d'information par rapport à *A. thaliana* et la vigne, des Blastx des 59 SU positionnées sur nos séquences ont été réalisés sur les banques de peptide de 2 espèces d'Astériidées (*M. guttatus* et *S. lycopersicum*) et de 2 Rosidées (*Medicago truncatula* et *Populus trichocarpa*). Le fait d'avoir augmenté le nombre d'espèces n'a pas augmenté le nombre de gènes identifiés. Nous avons donc considéré que la recherche d'informations, quant à la richesse en gènes des séquences de chicorée, pouvait être basée uniquement sur les banques de peptides des espèces *A. thaliana* et *V. vinifera*, qui sont probablement les plus complètes en terme d'annotation.

7.3.2.2 Identification de gènes potentiellement orthologues

Pour identifier les 23 gènes d'*A. thaliana* et les 27 gènes de vigne, seul le gène ayant la e-value la plus faible avait été sélectionné pour une position donnée. Dans le but d'identifier des gènes potentiellement orthologues, nous avons introduit un paramètre supplémentaire correspondant au rapport des e-values des 2 meilleurs résultats pour une position donnée, nommé Δ e-value (Figure 7.2). Ce paramètre nous a permis de distinguer

7.3. Etude de la fraction potentiellement codante à proximité du locus *S*

TABLEAU 7.3 – Gènes identifiés sur les séquences génomiques positionnées à proximité du locus *S* (Blastx réalisés sur les banques de peptides de *A. thaliana* et *V. vinifera*). La position relative des gènes sur les différents clones (Pos), le nombre de SU identifiés pour cette même position (SUtot) et le nombre de SU présentant une homologie avec le même gène que la séquence génomique (SU) sont indiqués pour chaque gène. Chr : chromosome sur lequel chaque gène est présent. Hits : nombre de résultats obtenu par position. Δ : Δ e-value, rapport des e-values du 1^{er} et 2^e résultat, sauf (a) : e-value du 2^e résultat. Alias : abréviations utilisées sur la figure 7.4. Cat : catégorie dans laquelle a été classé chaque gène par rapport à la séquence de chicorée : gène homologue (H) ou potentiellement orthologue (pO). Un tBlastn a été réalisé pour certains gènes pO : la meilleure homologie a été obtenue soit pour la SU identifiée à la même position (pO*), soit une SU différente (pO⁻). La fonction des gènes d'*A. thaliana* est précisée en annexe (p.223). (b) : différentes positions correspondant à un même gène chez *A. thaliana* ou *V. vinifera*.

Pos	Clone (Cg)	SUtot	<i>Arabidopsis thaliana</i>				<i>Vitis vinifera</i>											
			Gène	Chr	e-value	Hits	Δ	SU	Alias	Cat	Gène	Chr	e-value	Hits	Δ	SU	Alias	Cat
1	Si10(C11)																	
2	Si10(C11-C07-C08)	3	AT1G22870.1	1	5E-65	2	5E-01	3	1At1	pO*	GSVIVT01009308001	18	2E-23	1	1Vv18	pO		
3	Si10(C13)	3	AT1G10720.1	1	9E-25	3	2E-05	1	2At1	H	GSVIVT01037227001	18	2E-69	1	2Vv18	pO*		
4	Si10(C04)		AT4G34660.1	4	7E-23	1			3At4	pO	GSVIVT01009312001	18	6E-30	3	3Vv18	pO ⁻		
5	Si10(C12)	1	AT1G10140.1	1	4E-18	2	1E-02	1	4At1	pO*	GSVIVT01024196001	3	2E-23	2	4Vv3	H		
6	Si10(C16)	4	AT4G07410.1	4	1E-176	2	1E-05	4	5At4	H	GSVIVT01009317001	18	8E-20	1	5Vv18	pO*		
7	Si42(C01)	3	AT1G76880.1	1	7E-39	6	7E-04	3	6At1	H	GSVIVT01009321001	18	1E-107	1	6Vv18	pO*		
8	Si42(C10)	2	AT1G21200.1	1	1E-112	3	3E-26	2	7At1	pO*	GSVIVT01034523001	18	4E-42	5	7Vv18	pO ⁻		
9	Si42(C03)		AT1G30760.1(b)	1	1E-112	28	1E-01		8bAt1	H	GSVIVT01034521001	18	1E-106	4	8Vv18	pO*		
10	Si42(C04)		AT3G43700.1	3	1E-35	9	3E-06		9At3	pO	GSVIVT01021726001(b)	10	1E-111	29	9bVv10	H		
11	Si42(C08)										GSVIVT01037518001	6	1E-30	5	10Vv6	H		
12	Si42(C14)		AT1G30760.1(b)	1	1E-119	28	1E-02		8cAt1	H	GSVIVT01031016001	14	2E-17	1	11Vv14	pO		
13	Si37(C08)		AT1G30760.1(b)	1	1E-111	28	1E-02		8aAt1	H	GSVIVT01021726001(b)	10	1E-117	30	9cVv10	pO		
14	Si37(C03)										GSVIVT01021726001(b)	10	1E-115	21	9aVv10	pO		
15	Si37(C03)										GSVIVT01021964001	14	7E-13	4	12Vv14	H		
16	Si37(C03)		AT2G34440.1	2	2E-21	>50	1E-01		10At2	H	GSVIVT01038208001	5	4E-14	4	13Vv5	H		
17	Si37(C03)&Si52(C11)		AT3G43630.1	3	2E-59	5	7E-01		11At3	H	GSVIVT01037034001	3	2E-24	>50	14Vv3	H		
18	Si37(C03)&Si52(C11)	1	AT1G76810.1	1	1E-138	6	1E-01	1	12At1	H	GSVIVT01034896001	18	9E-62	4	15Vv18	H		
19	Si52(C18)	1	AT5G58740.1	5	3E-29	1		1	13At5	pO*	GSVIVT01034895001	18	2E-96	3	16Vv18	pO*		
20	Si52(C01)		AT1G76920.1	1	1E-114	2	5E-106		14At1	pO	GSVIVT01038735001	12	3E-30	1	17Vv12	pO*		
21	Si52(C07-C03)	1	AT1G43700.1	1	2E-18	2	1E-11	1	15At1	pO ⁻	GSVIVT01034899001	18	1E-64	2	18Vv18	pO		
22	Si27(C04)&Si52(C05)										GSVIVT01034897001	18	3E-16	6	19Vv18	H		
23	Si27(C03)	1	AT4G38470.1	4	5E-54	>100	8E-07		16At4	pO ⁻	GSVIVT01004252001(b)	Un	1E-52	6	20aVvUn	pO		
24	Si27(C03)		AT4G29090.1	4	1E-19	3	2E-08		17At4	pO	GSVIVT01001690001	18	3E-90	>100	21Vv18	pO ⁻		
25	Si27(C06)	3	AT1G76950.1	1	0	>50	0(a)	3	18At1	H	GSVIVT01004252001(b)	Un	2E-41	7	20bVvUn	pO		
26	Si27(C01)	4	AT5G42120.1	5	1E-154	>500	1E-47	1	19At5	pO ⁻	GSVIVT01027823001	5	0	7	22Vv5	pO ⁻		
27	Si27(C01)	1	AT5G50320.1	5	7E-69	1		1	20At5	pO*	GSVIVT01014693001	19	1E-107	>500	23Vv19	pO		
28	Si26(C15)		AT3G29785.1	3	1E-17	1			21At3	pO	GSVIVT01007544001	17	0	1	24Vv17	pO*		
29	Si26(C15)	1	AT3G21470.1	3	1E-134	>500	5E-41	1	22At3	pO ⁻	GSVIVT01010017001	18	1E-162	>500	25Vv18	pO ⁻		
30	Si26(C07)	2									GSVIVT01010015001	18	7E-53	13	26Vv18	pO*		
31	Si26(C09)	1	AT3G09040.1	4	0	>500	1E-124(a)	1	23At3	H	GSVIVT01010031001	18	0	>500	27Vv18	pO ⁻		

2 catégories de gènes :

- des gènes potentiellement orthologues (pO), c'est à dire des gènes d'*A. thaliana* ou *V. vinifera* présentant la meilleure homologie avec les gènes identifiés sur nos séquences génomiques. Il s'agit des gènes ayant un Δe -value $< 10^{-5}$.
- des gènes homologues (H) associés à un Δe -value $\geq 10^{-5}$, c'est à dire des gènes appartenant à la même famille de gènes, mais dont nous n'avons pas pu déterminer s'il s'agissait de la copie la plus proche de la séquence de chicorée chez la vigne et/ou *A. thaliana*.

Quatre gènes d'*A. thaliana* et 7 de vigne se sont avérés être les seuls gènes présentant une homologie avec la région considérée et, dans un premier temps, ont été considérés comme des gènes pO (Tableau 7.3). Pour les autres positions, de 2 à plus de 500 gènes présentant une forte homologie avec la séquence de chicorée ont été obtenus. En tenant compte de la valeur de Δe -value, 13 gènes supplémentaires ont été considérés comme pO chez la vigne et 8 chez *A. thaliana*. Deux gènes supplémentaires, appartenant à des régions dupliquées d'*A. thaliana*, ont été rajoutés malgré un Δe -value supérieur à 10^{-5} (cf. positions 2 et 5 du tableau 7.3). En effet, le second résultat qui leur était associé correspondait à l'autre copie de ce gène dupliqué²³. Nous avons donc obtenu 20 gènes classés comme potentiellement orthologues pour la vigne et 14 pour *A. thaliana* (Figure 7.5).

N'ayant pas accès à la totalité du génome de la chicorée, il n'était pas possible de déterminer si les gènes identifiés sur les séquences de chicorée étaient les orthologues de ceux d'*A. thaliana* et de vigne. En effet, en l'absence de données sur la totalité du génome de la chicorée, il n'était pas possible d'identifier le RBH (Reciprocal Best Hits) de ces gènes, qui est un des critères fréquemment utilisé pour procéder à l'identification d'orthologues (Darracq *et al.*, 2010 ; Koonin, 2005). Toutefois, un tblastn (e-value $< 10^{-5}$) des gènes pO associés à des SU a été réalisé sur les banques d'EST de *Cichorium*. Cette démarche a été possible pour 9 gènes d'*A. thaliana* sur 14 gènes pO et pour 14 gènes de vigne sur 20 pO, les autres gènes pO n'étant pas associés à des SU. Pour l'ensemble de ces gènes, au minimum 2 et jusqu'à 850 SU ont été obtenues et toutes les SU déjà associées ont été retrouvées. Les SU associées à 5 gènes sur 9 pour *A. thaliana* et à 8 gènes sur 14 pour la vigne ont quant à elles correspondu au meilleur résultat. Ces gènes ont donc été considérés comme les orthologues les plus probables de nos séquences

23. Informations obtenues sur le site Plant Genome Duplication Database, <http://chibba.agtec.uga.edu/duplication/>

(pO*, tableau 7.3). Pour 4 gènes d'*A. thaliana* et 6 de *V. vinifera*, les SU associées n'ont pas correspondu au meilleur résultat (pO⁻). Cela n'a pas invalidé le fait que ces gènes étaient les homologues les plus proches de nos séquences. Toutefois, l'existence d'autres copies de ces gènes chez la chicorée a été mise en évidence, ces copies présentant une plus forte homologie avec les gènes d'*A. thaliana* ou de la vigne.

7.3.3 Les logiciels de prédiction de gènes peuvent-ils être utilisés dans le cas de la chicorée ?

Les informations obtenues à partir des bases de données d'EST et de peptides d'*A. thaliana* et *V. vinifera* ont été confrontées aux résultats obtenus en utilisant 2 logiciels de prédiction de gènes (Genscan et Fgenesh, cf. Matériels et méthodes p. 64). Cette démarche avait pour double objectif d'éventuellement compléter les résultats obtenus précédemment et d'évaluer la précision de ces logiciels sur des séquences de chicorée. En moyenne, quand *A. thaliana* était utilisée comme organisme de référence, 19 gènes ont été prédits par clone (Fgenesh et Genscan). Dans le cas de la vigne comme organisme référent, 12 gènes par clone ont été obtenus en moyenne avec Fgenesh. Pour chacun des clones, les gènes identifiés à partir des prédictions se sont avérés positionnés les uns après les autres, couvrant quasiment toute la longueur des séquences de chicorée. L'utilisation de 2 espèces de références avec Fgenesh a permis d'obtenir un nombre de gènes prédits différents. Cependant, dans la plupart des cas, les régions exoniques prédites étaient identiques et seul le nombre d'exons par gène variait d'une espèce modèle à l'autre. Les gènes prédits en fonction d'*A. thaliana* contenaient moins d'exons que ceux prédits par rapport à la vigne. En moyenne, par prédiction et par clone, 3 gènes dont les positions exons/introns montraient une correspondance avec des SU, au moins pour partie, ont été obtenus. Il s'agit généralement de SU alignées avec la séquence génomique en plusieurs fragments (probablement des exons) et pour lesquelles une correspondance a été trouvée avec un gène d'*A. thaliana* et/ou de vigne.

Le fait d'utiliser des séquences correspondant aux Cg concaténés pour chacun des clones a pu fausser en partie les prédictions de ces logiciels. Ainsi, de nombreuses prédictions ont été faites à cheval sur des Cg dont la continuité ou l'orientation n'ont pas pu être déterminées. Elles ne peuvent donc être prises en considération. Il n'est pas surprenant que le meilleur résultat ait été obtenu avec le clone le mieux assemblé (Si27, cf. paragraphe 7.1.2.1). Parmi les gènes prédits par l'un ou l'autre des logiciels, 6 d'entre eux

en moyenne correspondaient totalement ou en partie avec la position de SU et/ou de gènes identifiés chez *A. thaliana* et la vigne (Figure 7.4). Par conséquent, dans la mesure où les banques d'EST et de peptides ont apporté de nombreuses informations, ces résultats n'ont pas été exploités plus en détails pour l'analyse des séquences génomiques de chicorée. Pour une optimisation de l'utilisation de ces logiciels, leur adaptation aux spécificités de la chicorée (au fur et à mesure des acquisitions de données de séquences génomiques de chicorée et/ou d'Asteraceae) devra être envisagée, ainsi que l'utilisation de logiciels prenant en compte les informations relatives aux EST.

7.3.4 Etude préliminaire de la conservation de la synténie

Lors de l'étude de la fraction codante des séquences de chicorée, des gènes d'*A. thaliana* et de vigne potentiellement orthologues (pO) et homologues (H) ont été identifiés (Tableau 7.3), en fonction des critères mis en place (Figure 7.2). Cette étude avait pour but d'estimer la richesse en gènes de nos séquences. Au delà de cet aspect, l'objectif était également de pouvoir déterminer quels gènes pouvaient être utilisés pour étudier la conservation de la synténie. Nous avons donc cherché à savoir s'il existait de la colinéarité entre la position des gènes sur la carte physique de la chicorée et leurs homologues chez 6 espèces dont le génome est séquencé (Figure 7.6).

7.3.4.1 Conservation de la synténie avec *A. thaliana* et *V. vinifera*

Chez *A. thaliana*, lors de l'analyse préliminaire des résultats de Blastx, une prédominance de gènes du chromosome 1 (30% des gènes identifiés) a été observée (Figure 7.5). Après avoir appliqué les différents critères de sélection donnés précédemment, parmi les 14 gènes d'*A. thaliana* considérés comme pO, 5 étaient positionnés sur le chromosome 1, 4 sur le chromosome 3, 2 sur le chromosome 4 et 3 sur le chromosome 5. La prédominance du chromosome 1 semblait avoir été conservée, mais avec une différence moins marquée par rapport aux chromosomes 3, 4 et 5.

Chez la vigne, les résultats préliminaires ont montré une proportion importante de gènes positionnés sur le chromosome 18 (23 sur 65 gènes identifiés). Après analyse des résultats de Blastx, 13 des 20 gènes considérés comme pO (65%), dont 11 pO*, se sont avérés positionnés sur le chromosome 18, les autres gènes pO étant chacun sur un chromosome différent. La distribution de ces 13 gènes sur ce chromosome long de 30 Mb a donc été étudiée (Figure 7.7).

7.3. Etude de la fraction potentiellement codante à proximité du locus *S*

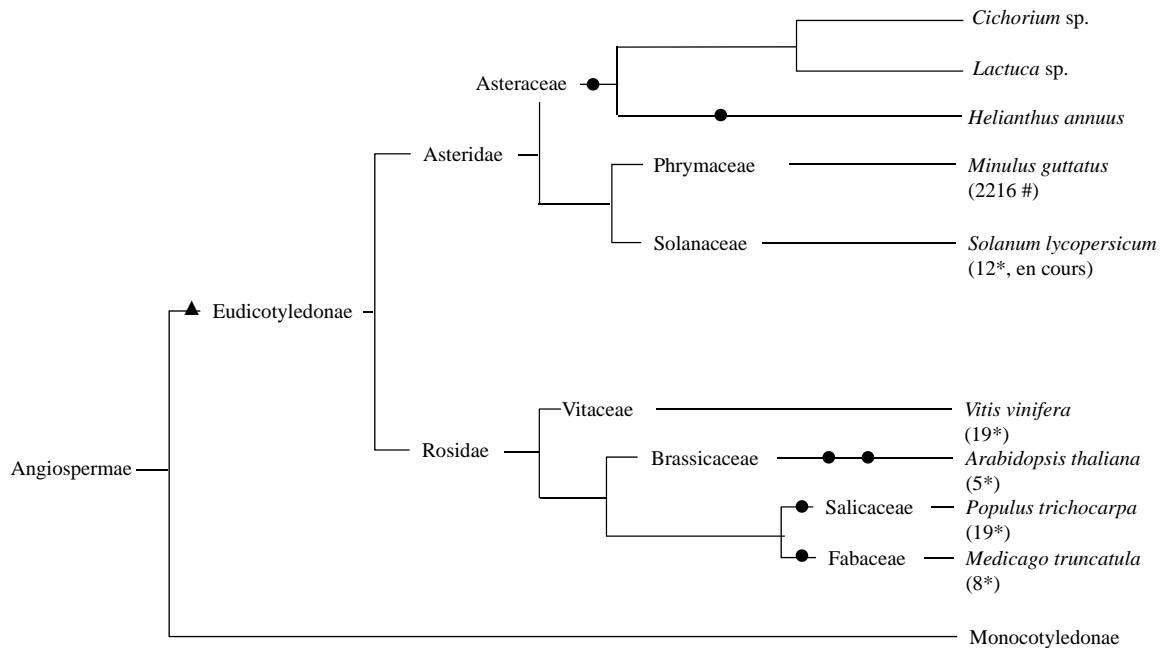


FIGURE 7.6 – Arbre phylogénétique des Dicotylédones dont les données de séquence ont été utilisées et positionnement de la chicorée, de la laitue et du tournesol. Lorsque le séquençage est achevé, le type d’assemblage est précisé : * indique le nombre de chromosomes ou groupes de liaison obtenu, # le nombre de scaffolds. Les événements de polypléidisation inférés récemment (●) et d’hexaploïdisation de l’ancêtre commun le plus récent (▲) ont également indiqués d’après Van de Peer *et al.* (2009) et Barker *et al.* (2008).

Ces gènes se sont avérés associés à 5 régions du chromosome 18 et répartis en groupes de 2 à 4 gènes, plus un gène isolé à une extrémité du chromosome. Cette représentation a permis de constater que la totalité des gènes pO identifiés, pour les clones Si26, Si42 et les 3 clones chevauchants, étaient regroupés sur des intervalles allant de 22 kb (clone Si42) à 1 900 kb (clones Si37-Si52-Si27) sur ce chromosome. De plus, 4 des 5 gènes identifiés sur le clone Si10 se sont révélés positionnés dans un intervalle de 180 kb, le 5^e gène correspondant au gène isolé à l’extrémité du chromosome 18 de la vigne.

En s’intéressant à l’organisation des différents blocs de gènes, nous avons pu constater que l’ordre des blocs correspondant aux clones Si42 à Si26 était conservé. Les bloc de gènes correspondant au clone Si10 et au supercontig se sont avérés positionné de part et d’autre du bloc correspondant au clone Si26. Les groupes de gènes correspondant aux clones Si10 et Si42, qui sont physiquement les plus proches sur la carte physique de la chicorée, se sont donc révélés les plus distants sur le chromosome 18 de la vigne. Les distances physiques entre les clones Si42, Si37-Si52-Si27 et Si26 ne sont pas connues.

Toutefois, la distance séparant le clones Si42 (86,6 kb) et le groupe de clones Si37-Si52-Si27 (273,9 kb) a pu être estimée à la taille d'un clone BAC au maximum, c'est-à-dire environ 100 kb (*cf.* figure 6.3 p.139). La région couverte par ces clones chez la chicorée a donc été estimée à moins de 500 kb. Sur le chromosome 18 de la vigne, les groupes de gènes correspondant à ces clones sont répartis sur un intervalle de 5,59 Mb.

Si les gènes identifiés sur les 6 clones séquencés se sont révélés représentatifs du chromosome 18, les échelles de distances entre les différents groupes de gènes se sont donc avérées très différentes.

7.3.4.2 Conservation de la synténie avec *M. guttatus*, *S. lycopersicum*, *P. trichocarpa* et *M. truncatula*

Afin de déterminer si d'autres espèces pouvaient apporter des informations complémentaires de celles obtenues à partir du génome de la vigne, nous nous sommes également intéressés aux résultats générés par les Blastx des SU sur les banques de peptides de *M. guttatus*, *S. lycopersicum*, *P. trichocarpa* et *M. truncatula*. Pour chaque position de nos séquences génomiques, correspondant à l'alignement d'un gène d'*A. thaliana* et/ou de vigne, nous avons sélectionné la SU qui présentait la plus forte homologie parmi toutes les SU alignées à cette position. Nous avons ensuite étudié la distribution des gènes obtenus sur les groupes de liaison et scaffolds de ces 4 espèces.

Chez *M. guttatus*, 16 gènes ont été identifiés, 11 pouvant être considérés comme des gènes pO ($\Delta e\text{-value} < 10^{-5}$) et 5 comme des gènes H ($\Delta e\text{-value} > 10^{-5}$). Les gènes identifiés chez *M. guttatus* se sont avérés localisés sur 8 scaffolds différents. Sur 3 de ces 8 scaffolds, plusieurs gènes ont été positionnés :

- 4 gènes pO et 2 gènes H sur le scaffold 2 (4,6 Mb),
- 1 gène pO et 2 gènes H sur le scaffold 98 (1,2 Mb),
- 1 gène pO et 1 gène H sur le scaffold 58 (1,3 Mb).

De plus, en considérant qu'il n'était pas possible de trancher entre 1^{er} et 2nd résultat, lorsque la valeur de $\Delta e\text{-value}$ était inférieure à 10^{-5} , 1 gène supplémentaire positionné sur le scaffold 98 et 2 sur le scaffold 2 ont été identifiés. Les gènes du scaffold 2 se sont révélés représentatifs de ceux présents sur le clone Si10 et le supercontig et ont été positionnés dans un intervalle total de 2,13 Mb. Les gènes du scaffold 98 ont été associés uniquement aux gènes présents sur le supercontig dans un intervalle de 190 kb.

La distribution des gènes sur les scaffolds 2 et 98 a montré que ceux-ci étaient regroupés en blocs de 2 à 4 gènes (Figure 7.8), sur des intervalles allant de 30 à 200 kb. Deux

7.3. Etude de la fraction potentiellement codante à proximité du locus *S*

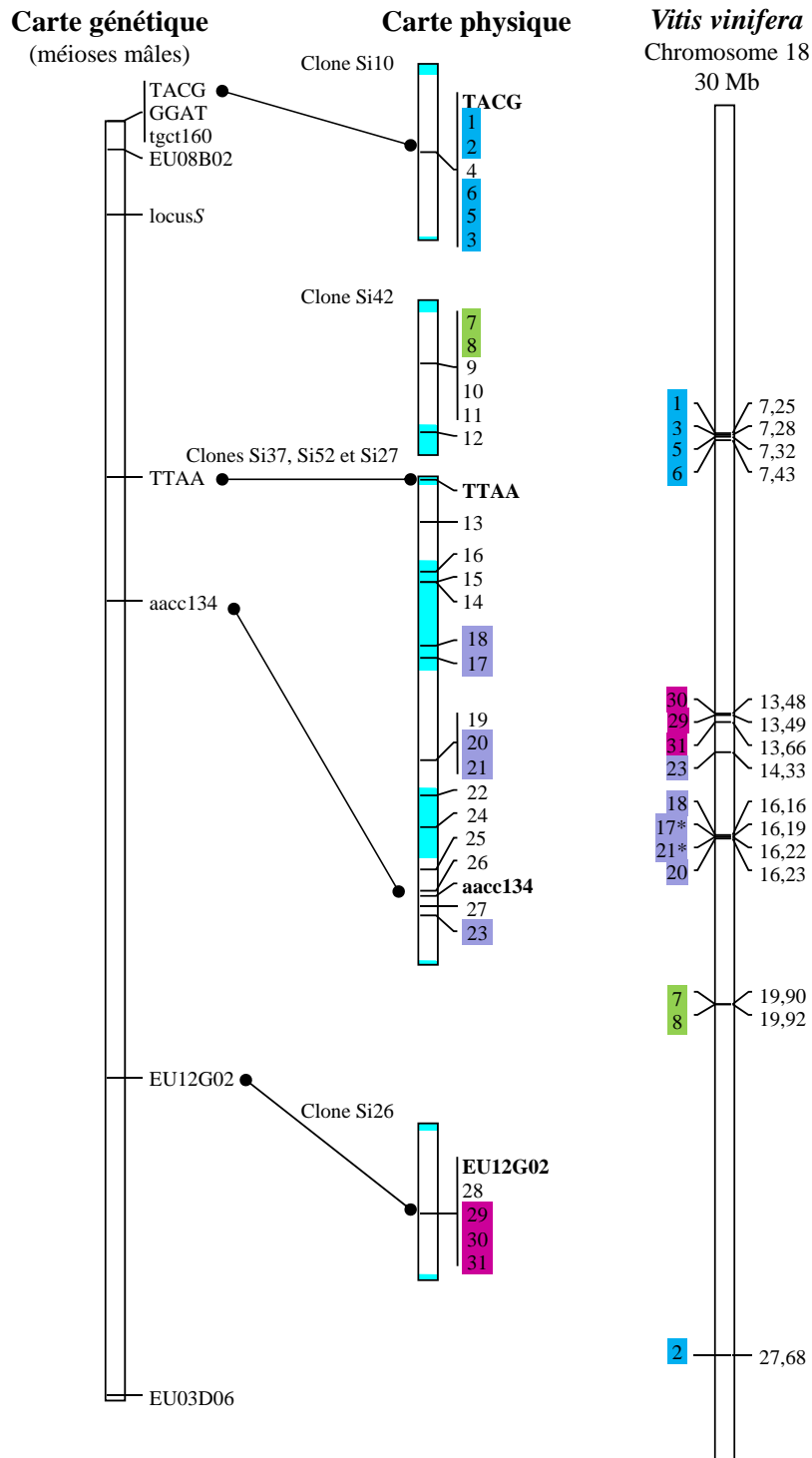


FIGURE 7.7 – Distribution des gènes pO et H(*) du chromosome 18 de la vigne et comparaison avec leur distribution sur les clones de la région du locus *S*. Chaque numéro correspond à 1 gène (cf. Position dans le tableau 7.3). Un code couleur a été défini pour chaque clone ou groupe de clones chevauchants, seuls les gènes présents sur le chromosome 18 étant surlignés sur la carte physique de la chicorée. Les clones Si27, Si37 et Si52 ont été réunis pour ne former qu'une seule séquence. Les régions des clones surlignées en bleu correspondent aux Cg ayant été positionnés et orientés.

gènes associés au clone Si42 ont été identifiés, chacun sur un scaffold, à l'extrémité d'un bloc correspondant aux gènes associés aux clones Si37-Si52-Si27. Des distances de 50 kb (scaffold 2) et 60 kb (scaffold 98) ont été calculées entre les gènes homologues du clone Si42 et des clones Si37-Si52-Si27. De plus, alors que sur le scaffold 98, les gènes présents sur les 3 clones chevauchants ont été observés dans un intervalle de 130 kb, sur le scaffold 2, ils se sont avérés répartis en 2 blocs distants de 1, 24 Mb. Enfin, comme chez la vigne, les gènes correspondant au clone Si10 ont été positionnés à l'opposé du gène homologue sur le clone Si42.

Chez *S. lycopersicum*, 14 gènes différents ont été identifiés (8 pO et 6 H), localisés sur 6 chromosomes différents. Sur 2 d'entre eux, plusieurs gènes ont été identifiés :

- 5 gènes pO et 1 gènes H sur le chromosome 4 (64,06 Mb),
- 2 gènes pO et 1 gène H sur le chromosome 12 (65,49 Mb).

De plus, en considérant les 2^e résultats de blast des gènes H, 3 gènes supplémentaires positionnés sur le chromosome 4 ont été identifiés. Les gènes représentés sur le chromosome 12 correspondent à des gènes sur les clones Si10 et Si42, chaque gène étant distant de plusieurs Mb. Les homologues des gènes positionnés sur le chromosome 4 se sont avérés représentatifs des gènes présents sur les clones séquencés, excepté pour le clone Si42. Toutefois, un gène correspondant à ce clone a été identifié en 2^e résultat d'un gène pO positionné sur le chromosome 12.

La disposition des gènes du chromosome 4 a révélé 3 blocs de 2 à 5 gènes (Figure 7.8). La taille des blocs de gènes variait entre 70 kb (gènes homologues à ceux des clones Si10 et Si26) et 210 kb (gènes homologues à ceux des 3 clones chevauchants). Les blocs correspondant aux gènes des clones Si26 et Si37-Si52-Si27 se sont révélés adjacents, distants de 450 kb et séparés par 4,46 Mb du bloc de gènes présents sur le clone Si10 . Le gène représentant le clone Si42, à quant à lui, été positionné à 260 kb du bloc de gènes homologues sur les clones Si37-Si52-Si27. Sur le chromosome 4, la disposition des blocs de gènes s'est révélée analogue à la disposition des gènes sur la carte physique de la chicorée.

Chez *P. trichocarpa*, 16 gènes ont également été identifiés (8 pO et 8 H) sur 7 groupes de liaison différents. Douze des 16 gènes ont été positionnés sur 3 groupes de liaisons :

- 7 gènes (2 pO et 5 H) sur le groupe de liaison 2 (23,3 Mb),
- 3 gènes (2 pO et 1 H) sur le groupe de liaison 5 (25,5 Mb),
- 2 gènes (1 pO et 1 H) sur le groupe de liaison 14 (17,5 Mb).

7.3. Etude de la fraction potentiellement codante à proximité du locus *S*

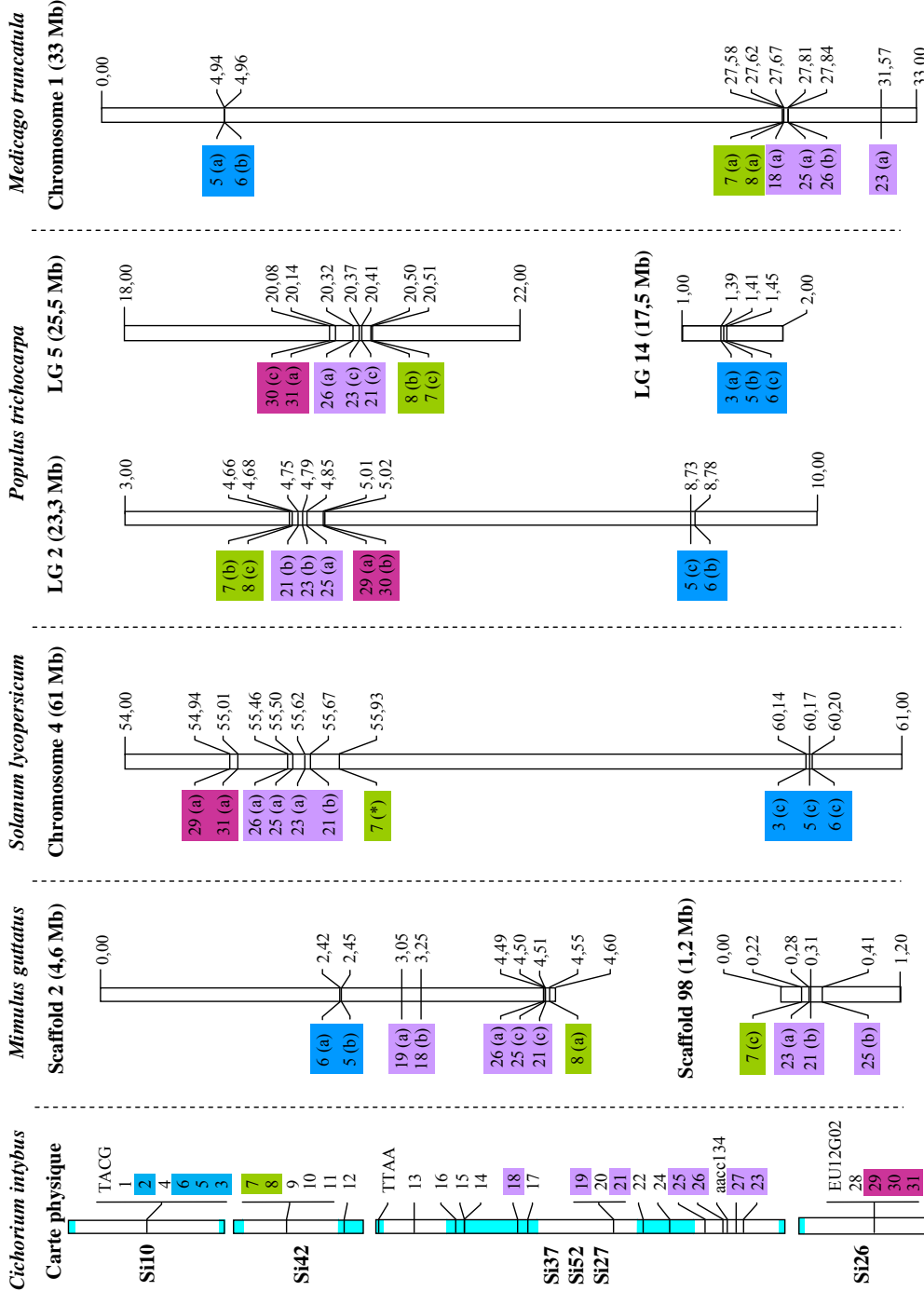


FIGURE 7.8 – Distribution des gènes pO et H identifiés à partir de SU de chicorée sur les chromosomes, groupes de liaison et scaffolds de *M. guttatus*, *S. lycopersicum*, *P. trichocarpa* et *M. truncatula*. Les clones de chicorée sont alignés les uns au dessus des autres (carte physique à gauche). Les clones Si27, Si37 et Si52 ont été réunis pour ne former qu'une seule séquence. Sur la carte physique, chaque numéro correspond à 1 gène (cf. position dans le tableau 7.3) et seuls les gènes associés à des SU ont été surlignés, un code couleur étant défini pour chaque clone ou groupe de clones chevauchants. (a) Gène pO, (b) gène H, (c) gène correspondant au 2^e résultat de Blast. Les régions des clones surlignées en bleu correspondent au Cg ayant été positionnés et orientés.

De plus, 1 gène positionné sur le groupe de liaison 14, 2 sur le groupe de liaison 2 et 4 sur le groupe de liaison 5 ont également été identifiés, en considérant les seconds résultats de Blast pour des valeurs de Δe -value $< 10^{-5}$. Les gènes positionnés sur le groupe de liaison 2 se sont avérés représentatifs de leurs homologues identifiés sur les 6 clones de chicorée. Les gènes positionnés sur les groupes de liaison 5 et 14 correspondaient à des gènes présents sur le supercontig et le clone Si26 (LG5), et sur le clone Si10 (LG14). Sur ces 3 groupes de liaison, la distribution des gènes a mis en évidence des blocs de 2 à 3 gènes couvrant des intervalles compris entre 10 et 100 kb (Figure 7.8). Sur le groupe de liaison 14, 3 gènes (clone Si10) ont été localisés dans un intervalle de 60 kb. Les 3 blocs de gènes correspondant aux clones du supercontig et au clone Si26, sous le locus *S*, ont présentés un ordre similaire à celui des clones sur la carte physique de la chicorée. Ces 3 blocs sont répartis sur 360 kb sur le groupe de liaison 2 et 430 kb sur le groupe de liaison 5, les intervalles entre les blocs étant compris entre 70 et 180 kb. Sur le groupe de liaison 2, le bloc de 2 gènes homologues des gènes du clone Si10 a été positionné dans un intervalle de 40 kb, distant de 3,71 Mb du bloc correspondant aux gènes du clone Si26. Là encore, le bloc de gène correspondant au clone Si10 a été positionné à l'opposé de celui correspondant aux gènes du clone Si42.

Chez *M. truncatula*, 14 gènes ont été identifiés dont la moitié (5 pO et 2 H) a été positionnée sur le chromosome 1. Ces 7 gènes se sont avérés homologues des gènes identifiés sur le clone Si10 et le supercontig. Un bloc de gènes de 20 kb (Si10), ainsi qu'un bloc de 260 kb, correspondant à des gènes présents sur le supercontig, ont été observés (Figure 7.8). Une distance de 22,62 Mb sépare les 2 blocs de gènes.

L'observation de la distribution des gènes sous forme de blocs compris dans des intervalles de 10 à plusieurs centaines de kb a donc mis en évidence l'existence d'une colinéarité entre les séquences de chicorée et des régions chromosomiques de *V. vinifera*, *M. guttatus*, *S. lycopersicum*, *P. trichocarpa* et *M. truncatula*. Pour chaque espèce, sur un même chromosome, groupe de liaison ou scaffold, au moins 3 blocs correspondant à des gènes identifiés sur 3 clones différents ont été trouvés. Ces observations suggèrent qu'il existe une relative conservation de la synténie entre la région du locus *S* de la chicorée et des régions chromosomiques de ces 5 espèces.

Chez, *M. guttatus*, *S. lycopersicum* et *P. trichocarpa*, les blocs de gènes associés aux clones qui encadrent le locus *S* ont de plus été identifiés dans des intervalles allant de 2,22 Mb à 5,26 Mb, ce qui correspond à des distances physiques vraisemblablement

7.3. Etude de la fraction potentiellement codante à proximité du locus *S*

comparables avec la région étudiée chez la chicorée. En effet, la région du locus *S* (3,16 cM en moyenne) correspond à environ 5% de la longueur totale du groupe de liaison 2. Si on tient pour constant le rapport Mb/cM, la taille physique de cette région est estimée à 4,7 Mb. De plus, parmi ces 3 espèces, seul le chromosome 4 de la tomate présente une organisation des blocs de gènes similaire à celle observée sur la carte physique de la chicorée.

Chapitre 8

Discussion

Deux étapes principales sont nécessaires au clonage positionnel d'un locus : une carte génétique fine et une carte physique de la région d'intérêt. La première carte est nécessaire à l'initiation de la seconde, qui permet d'enrichir la première. De nombreux allers-retours entre ces cartes sont réalisés avant d'atteindre le locus recherché. A partir de là seulement, chacune de ces cartes peut être considérée comme réellement achevée.

Dans le cadre du clonage positionnel du locus *S* chez la chicorée, la carte génétique fine présentée dans la partie I a constitué le point d'ancrage permettant d'initier la marche chromosomique. Deux contigs et un supercontig ont été construits (*cf.* figure 6.3 p.139). En se référant à la carte génétique obtenue à partir des méioses mâles, les contigs correspondants aux marqueurs TACG et EU12G02 ont été positionnés de part et d'autre du supercontig.

Du fait de l'importance des distances génétiques séparant les locus TACG et TTAA (1,34 cM) d'une part, et aacc134 et EU12G02 (3,43 cM) d'autre part, il n'est pas surprenant de n'avoir pas pu relier physiquement les 3 contigs de clones. Le supercontig a permis de couvrir la distance génétique séparant les marqueurs TTAA et aacc134, soit un intervalle de 0,47 cM. De plus, il s'étend au delà de chacun de ces marqueurs. Ce supercontig correspond à une distance physique d'au moins 360 kb (somme des tailles des séquences des 3 clones chevauchants, Si37, Si52 et S27, et du clone Si42). Le marqueur Si30t7, défini à partir d'un BES et cartographié sur la carte haute résolution, a permis d'orienter le supercontig par rapport au locus *S*. Les marqueurs Si30t7 et aacc134 sont présents sur un même clone (Si27) et sont distants de moins de 50 kb. Au moins 2 évènements de recombinaison ont été observés dans cet intervalle.

Les séquences obtenues à partir des clones correspondant à des régions génomiques proches du locus *S* ont permis de déterminer les premiers rapports distance physique/distance génétique (Mb/cM) dans cette région.

Entre les marqueurs TTAA et Si35t7, présents sur le clone Si37 et séparés d'au moins 77 kb, aucun évènement de recombinaison n'a été observé. Comme le marqueur Si35t7 est peu informatif (2 allèles codominants) et qu'il est monomorphe dans les populations E, F et G ($n = 787$), le génotype des individus recombinants de ces populations n'a pu être déterminé. Ces résultats n'ont donc pas permis de conclure quant à la nature du rapport Mb/cM dans cette région. On peut toutefois supposer que, puisqu'aucun évènement de recombinaison n'a été observé dans les populations B, H, M (pour les méioses mâles et femelles, $n = 1720$) et D (méioses femelles, $n = 317$), la fréquence de recombinaison entre TTAA et Si35t7 est probablement très faible.

Les marqueurs Si30t7 et aacc134, présents sur le même clone (Si27), sont distants de moins de 50 kb. Cependant, du fait du caractère dominant de aacc134, le génotype complet de certains individus recombinants n'a pu être déterminé pour ce marqueur. Néanmoins, un minimum de 2 évènements de recombinaison a été observé. Un rapport distance génétique/distance physique maximal a donc été calculé pour les méioses mâles (0,7 Mb/cM), aucun recombinant n'ayant été observé à partir des méioses femelles.

Enfin, l'estimation la plus fiable du rapport Mb/cM a été obtenue à partir de l'intervalle compris entre les marqueurs TTAA et aacc134 : 0,5 Mb/cM pour les méioses mâles et 2,2 Mb/cM pour les méioses femelles. La moyenne de ces 2 rapports est de 1,35 Mb/cM, ce qui est similaire à l'estimation de 1,5 Mb/cM faite sur l'ensemble du génome, à partir de la carte génétique consensus, qui a été construite sans distinction des méioses mâles et femelles (Cadalen *et al.*, 2010). La différence observée entre les cartes génétiques mâles et femelles (*cf.* chapitre 4) se répercute ici sur le rapport distance physique/distance génétique.

A partir de ces résultats, en considérant que le rapport Mb/cM reste constant dans cette région chromosomique, la distance entre TTAA et le locus *S* a été estimée à plusieurs centaines de kb (entre 0,5 et 0,8 Mb). En prenant en compte le clone Si42 (86,6 kb), identifié au delà du marqueur TTAA, la distance restant à parcourir correspondrait donc à 4 à 8 clones BAC supplémentaires.

Les résultats du chapitre 4 ont montrés que la différence des fréquences de recombinaison entre méioses mâles et femelles n'était pas causée par le locus *S*. La question qui

se pose maintenant est de savoir si le rapport Mb/cM restera constant jusqu'au locus *S*, indépendamment de la différence de fréquence de recombinaison observée entre les méioses mâles et les méioses femelles. L'évolution du rapport distance physique/distance génétique peut dépendre de la position du locus *S* sur le chromosome et de la taille de la région hypervariable encadrant le locus *S*. Selon les espèces considérées, dans la région du locus *S*, 1 cM peut correspondre quelques centaines de kb chez *Ipomoea trifida* (313 kb pour 0,57 cM) ou *Brassica oleracea* (118 kb pour 0,3 cM), ou bien à plusieurs Mb chez *Petunia inflata* (4,4 Mb pour 0,25 cM) (Casselman *et al.*, 2000 ; Suzuki *et al.*, 2004 ; Tomita *et al.*, 2004b ; Wang *et al.*, 2003). Pour les 2 premières espèces citées, le locus *S* a été positionné à l'extrémité d'un chromosome, alors qu'il est localisé à proximité du centromère pour *P. inflata* (*cf.* tableau 5.2 p.122). Cette position sub-centromérique explique ce rapport extrêmement élevé.

Le marqueur le plus distal du groupe de liaison 2 sur la carte consensus (Cadalen *et al.*, 2010), EU08B02, a été cartographié à 0,24 cM du locus *S*. Plusieurs observations cytologiques des chromosomes de chicorée ont permis de constater qu'aucun des chromosomes n'était télocentrique (Doré *et al.*, 1996 ; Ge *et al.*, 2002 ; Haque et Godward, 1984 ; Ma *et al.*, 1984 ; Rambaud *et al.*, 1992). D'autre part, le rapport Mb/cM calculé dans la région située à 1 cM sous le locus *S*, est similaire au rapport estimé sur l'ensemble du génome. Ces éléments suggèrent que le locus *S*, positionné à l'extrémité du groupe de liaison 2 chez la chicorée, ne se trouve pas au niveau d'un centromère.

Dans ce contexte, on s'attend cependant à une augmentation du rapport Mb/cM à mesure que la marche chromosomique avancera vers le locus *S*, ainsi qu'à une suppression de la recombinaison au niveau du locus *S*. Néanmoins, la distance physique délimitant le locus *S* devrait être du même ordre que chez *Ipomoea trifida* ou *Brassica oleracea*, soit quelques centaines de kb.

L'accès à des séquences génomiques nous a permis de réaliser la première étude génomique de la chicorée sur 546 kb, soit 0,04% de son génome. Une des hypothèses formulées pour expliquer l'absence de recombinaison au niveau du locus *S* est l'hétérogénéité de structure des séquences des différents haplotypes. Cette hétérogénéité pourrait notamment être causée par la présence d'éléments répétés, comme cela a été observé chez *P. inflata*, *Papaver rhoeas*, *I. trifida* ou les *Brassicaceae*, que ce soit des éléments transposables ou des microsatellites (McCubbin *et al.*, 2000 ; Sherman-Broyles et Nasrallah, 2008 ; Tomita *et al.*, 2004b ; Wheeler *et al.*, 2003). Pour cette raison, nous nous sommes

intéressés à la composition des séquences en éléments répétés.

La composition en éléments transposables à proximité du locus *S* a été estimée à 5,71% pour les séquences des inserts, et 11,7% pour les BES. Les résultats obtenus pour d'autres espèces (Tableau 8.1) sont compris entre 8,95% pour les BES de clémentines (Terol *et al.*, 2008) et 41,4% sur l'ensemble du génome de la vigne (Jaillon *et al.*, 2007). La composition de nos séquences en éléments transposables ne semble donc pas particulièrement élevée, dans la mesure où l'ensemble des éléments présents a pu être identifié. Il est cependant difficile d'extrapoler à l'ensemble du génome, étant donné la faible proportion de séquences analysée. En effet, le pourcentage en éléments transposables varie du simple au double lorsqu'on étudie soit les séquences des inserts, soit les BES. De plus, la répartition de ces éléments n'est probablement pas homogène sur le génome, comme cela a été montré chez *A. thaliana*. Chez cette espèce, les séquences situées au niveau du centromère sont composées de 27% d'éléments transposables alors que par ailleurs, ils représentent seulement 3% des séquences (Wright *et al.*, 2003). Enfin, les éléments transposables évoluent très rapidement. Ainsi, au sein d'une espèce donnée, des éléments qui lui sont spécifiques peuvent être identifiés (Cavagnaro *et al.*, 2009). L'utilisation d'*A. thaliana* comme espèce de référence lors de la recherche de ces éléments a donc permis d'identifier la fraction d'éléments mobiles homologues entre *A. thaliana* et la chicorée, mais probablement pas tous les éléments transposables.

Cette étude a également révélé plus de 10% de séquences répétées sur nos 6 clones. Il en est de même pour 14,5 % de la séquence du clone 60A10 (sur le groupe de liaison 5) qui contient 2 régions de plus de 1 000 pb retrouvées sur les clones séquencés au voisinage du locus *S*. Parmi ces régions répétées, seulement 3 présentent une homologie avec un élément transposable identifié à partir des banques de données d'*A. thaliana*. Il est donc possible que parmi ces séquences, une partie corresponde à des éléments mobiles spécifiques de *C. intybus*.

Le nombre de Cg obtenu pour les clones de chicorée (14 Cg en moyenne par clone) s'est toujours révélé supérieur à celui obtenu pour des espèces dont le séquençage de clones BAC a été réalisé simultanément par le CNRGV. Nous avons également constaté que la moitié des extrémités des Cg s'achevait par un motif microsatellite. La fréquence d'éléments microsatellites dans la région étudiée (1 SSR pour 4,9 kb minimum) semble du même ordre que celle observée pour d'autres espèces (de 1 SSR pour 1,2 kb chez le riz à 1 SSR pour 6,3 kb chez *A. thaliana*, cf. tableau 8.1). Cependant, le pourcentage de ces éléments (1,75%) est de 2 à plus de 4 fois plus élevé chez la chicorée par rapport à

TABLEAU 8.1 – Composition génomique en éléments répétés pour différentes espèces. * Estimation de la distance moyenne entre SSR (kb). GSS : Genome Survey Sequences, g : génome, chr : chromosome.

Espèce	Microsatellites		Éléments répétés(%)	Source des séquences	Référence
	%	Fréquence*			
<i>C. intybus</i>	1,75	3,6 à 4,9	5,71	BAC (~ 0,4% g)	Cette étude
<i>C. intybus</i>	1,2	-	11,7	BES (~ 0,004% g)	Cette étude
<i>A. thaliana</i>	0,4	11,2	-	Génome	Cette étude
<i>A. thaliana</i>	-	-	10,00	Génome	Initiative (2000)
<i>A. thaliana</i>	-	3,0	10	Génome	Terol <i>et al.</i> (2008)
<i>A. thaliana</i>	-	1,1	-	Génome	Lawson et Zhang (2006)
<i>A. thaliana</i>	0,75	2,4	-	Génome	Grover et Sharma (2007)
<i>A. thaliana</i>	-	3,3	-	GSS	Cavagnaro <i>et al.</i> (2009)
<i>A. thaliana</i>	-	6,3	-	GSS	Hsu <i>et al.</i> (2011)
<i>Solanum lycopersicum</i>	-	5,3	-	GSS	Cavagnaro <i>et al.</i> (2009)
<i>Solanum tuberosum</i>	-	3,7	-	GSS	Cavagnaro <i>et al.</i> (2009)
Riz	-	1,2	-	Génome	Lawson et Zhang (2006)
<i>Citrus clementina</i>	-	5,0	8,95	BES (~ 8% g)	Terol <i>et al.</i> (2008)
<i>Carica apaya</i>	-	2,3	16,2	BES (~4,7% g)	Lai <i>et al.</i> (2006)
<i>Brassica rapa</i>	-	5,6	14	BES (~ 1,5% g)	Hong <i>et al.</i> (2006)
<i>V. vinifera</i>	0,86	2,1	41,4	Génome	Jaillon <i>et al.</i> (2007)
<i>V. vinifera</i>	-	2,9	-	GSS	Cavagnaro <i>et al.</i> (2009)
<i>V. vinifera</i>	-	5,9	-	GSS	Hsu <i>et al.</i> (2011)
<i>Daucus carota</i>	0,18	7,6	11,87	BES (~ 0,3% g)	Cavagnaro <i>et al.</i> (2009)
<i>Phalaenopsis equestris</i>	-	4,8	-	BES	Hsu <i>et al.</i> (2011)
<i>Oryza sativa</i>	-	9,1	-	GSS	Hsu <i>et al.</i> (2011)
<i>Triticum aestivum</i>	0,2	-	76,3	BES (~ 1,1% chr 3B)	Hsu <i>et al.</i> (2011)

d'autres espèces. Ces résultats suggèrent que, pour les séquences étudiées, les microsatellites seraient plus longs que pour les autres espèces citées dans le tableau 8.1. Toutefois, il existe une hétérogénéité importante des résultats obtenus pour la même espèce, selon la méthode de détection ou l'origine des séquences utilisées (par exemple 1 SSR pour 1,1 à 6,3 kb pour *A. thaliana*, cf. tableau 8.1). Le génome d'*A. thaliana* a donc été analysé dans les mêmes conditions que les séquences de chicorée, afin de comparer les pourcentages obtenus pour ces 2 espèces. Les résultats obtenus pour *A. thaliana* ont révélé un pourcentage de SSR plus faible que ceux présentés dans le tableau 8.1 (0,4% avec 1 SSR pour 11,2 kb). Cela indique donc que, même si le logiciel RepeatMasker ne détecte pas la totalité des régions contenant des microsatellites, les séquences de chicorées à proximité du locus *S*, comme au voisinage du locus de la stérilité mâle (clone 60A10), semblent particulièrement riches en éléments de ce type.

Le pourcentage, apparemment élevé de ces éléments dans nos séquences, peut sans

doute expliquer en partie les difficultés rencontrées lors de l'assemblage des séquences. Cependant, la technique de séquençage et le processus d'assemblage des séquences peuvent également avoir introduit un biais dans l'estimation du pourcentage de SSR observé dans les séquences de chicorée. Afin de pouvoir conclure sur la composition exacte de nos séquences en éléments SSR, il faudrait, pour une même séquence, pouvoir disposer de résultats obtenus à partir de différentes techniques de séquençage.

L'étude des éléments répétés présents dans les séquences génomiques de chicorée nous a également permis de reconsidérer les résultats obtenus lors de la marche chromosomique. Dans un premier temps, nous avons pour objectif d'avancer conjointement sur les 2 banques BAC, l'une construite à partir d'un génotype homozygote au locus *S*, l'autre provenant d'un génotype hétérozygote. Lorsqu'un seul haplotype était considéré, la construction de contigs n'a posé aucun problème. Nous avons constaté que, pour des clones obtenus à partir d'haplotypes différents, cette construction était complexe et que les résultats obtenus pouvaient être incohérents. Ces problèmes ont confirmés l'importance de disposer d'un génotype homozygote pour au moins une des 2 banques BAC. Pour construire les contigs, nous ne disposions que de l'information limitée contenue dans les BES. Nous avons tout d'abord supposé que cette complexité d'assemblage pouvait être liée à l'une des caractéristiques associées au locus *S* en général, c'est à dire la localisation dans une région fortement divergente entre haplotypes. Il est certain que le polymorphisme entre haplotype est pour partie responsable de ces difficultés. En effet, l'amplification de certains BES s'est avérée spécifique d'un haplotype donné, ce qui peut s'expliquer par la présence de SNP, ou l'absence des séquences cibles au niveau du site d'hybridation des amorces. Dans certains cas, cependant, des BES, présents uniquement sur les clones d'un haplotype donné, ont été amplifiés sur l'ADN génomique d'individus de différents haplotypes et se sont révélés monomorphes. La composition des BES en éléments transposables, près de 12%, soit le double de l'estimation faite sur les séquences des inserts totaux, pourrait expliquer ces résultats. La présence de séquences répétées, qui représentent 10,45% des 546 kb séquencés, pourraient également être associées à ces résultats. Ainsi, les BES pour lesquels aucune homologie avec un élément transposable n'a été trouvée peuvent tout de même correspondre à des régions répétées. Avant même de disposer d'informations à propos des séquences génomiques, des critères stricts de sélection des clones ont été appliqués :

- étude de la présence/absence des BES de la totalité des clones identifiés avec une

- même sonde afin de construire un contig,
- dans un contig, positionnement obligatoire des BES d'un clone aux extrémités de celui-ci,
- utilisation des BES pour lesquels un seul produit d'amplification est obtenu à partir de l'ADN génomique.

Cette stratégie de construction des contigs s'est révélée adaptée à la complexité de la région génomique étudiée. Les clones obtenus à partir de la banque CinS1S4 et provenant de 2 haplotypes ne permettant pas de respecter les critères de sélection définis, nous nous sommes concentrés sur la banque CinS2S2. Un facteur supplémentaire dans la validité des contigs est le nombre de clones à partir desquels ils ont été construits. Plus le nombre de clones obtenus est important, plus il est facile d'identifier les incohérences lors de la construction des contigs. Si le nombre de clones identifiés à partir d'une sonde dépasse la couverture de la banque, comme cela a été le cas pour la sonde tgct160, alors ces clones doivent être utilisés avec prudence, voir exclus de la marche chromosomique, comme nous avons choisi de le faire.

La prise en compte de tous ces paramètres, et l'utilisation d'une banque construite à partir d'un génotype homozygote dans la région du locus *S*, ont permis d'obtenir des résultats fiables, comprenant la construction d'un supercontig et l'orientation de la marche chromosomique vers le locus *S*.

Les déterminants potentiels de l'AI identifiés pour *I. trifida* et le déterminant mâle des Papaveraceae ne présentent aucune homologie avec des gènes identifiés chez d'autres espèces (Rahman *et al.*, 2007b ; Wheeler *et al.*, 2010). Le déterminant femelle des Papaveraceae, quant à lui, appartient à une famille de gènes référencée, mais dont la fonction est inconnue (Wheeler *et al.*, 2010). Toutefois, des gènes appartenant à une famille connue ont également été recrutés pour remplir la fonction de déterminant mâle ou femelle de l'AI, comme par exemple chez les *Brassicaceae*, où le gène *SRK* appartient à la famille des récepteurs kinases (Stein *et al.*, 1991). Les déterminants de l'AI chez la chicorée peuvent donc ne présenter aucune homologie avec les gènes présents dans les bases de données, ou appartenir à une famille de gènes connue.

Jusqu'à ce jour, il n'existait pas d'informations sur la composition des séquences génomiques de la chicorée et aucun génome d'Asteraceae n'est encore séquencé. Par conséquent, il était nécessaire de déterminer quels outils pouvaient être utilisés afin d'identifier des gènes à partir de séquences génomiques de chicorée.

A partir des Blastn et Blastx, réalisés à partir des séquences génomiques, 31 régions

distinctes présentant une forte homologie avec un gène d'*A. thaliana* et/ou de *V. vinifera* ont été identifiées, ce qui correspond à 23 et 27 gènes différents respectivement pour chacune de ces espèces (*cf.* tableau 7.3 p.161). Un plus grand nombre de gènes homologues et potentiellement orthologues ont été trouvés chez la vigne, ce qui en fait une espèce au moins aussi propice qu'*A. thaliana* pour identifier des séquences potentiellement codantes chez la chicorée. Un nombre important d'informations a donc été apporté, malgré la distance phylogénétique importante entre la chicorée et les 2 espèces choisies comme références.

Lors de l'utilisation de logiciels de prédiction de gènes, peu des gènes prédits se sont trouvés associés à des SU, si celles-ci ne présentaient pas par ailleurs d'homologie avec des gènes des espèces de référence. De plus, un grand nombre de prédictions a été obtenu dans des régions génomiques où aucun gène et SU n'ont été identifiés. Cela a mis en évidence les limites associées à cette approche, lorsqu'elle était utilisée seule. De plus, selon le logiciel et l'espèce de référence choisie, la structure des gènes prédits s'est avérée différente. Enfin, le fait de ne disposer que d'assemblages partiels des clones BAC a également généré un biais dans la prédiction des gènes. Ainsi, l'optimisation du paramétrage de ce type de logiciels pour prédire des gènes à partir du génome de la chicorée sera donc nécessaire. En particulier, il sera important d'intégrer les données d'alignement des SU sur les séquences génomiques afin de mieux définir, par exemple, les sites de jonction exon/exon.

Par ailleurs, nous avons recherché des gènes potentiellement orthologues, c'est-à-dire des gènes ayant divergé en raison d'un évènement de spéciation, regroupés sur un même chromosome ou sur une portion de chromosome. L'objectif était d'identifier une potentielle conservation de l'ordre de ces gènes (synténie) entre la chicorée et 6 espèces de Dicotylédones. La conservation de la synténie a été observée à de nombreuses reprises pour des espèces proches (par exemple : Feng *et al.*, 2009 ; Mun *et al.*, 2009 ; Yan *et al.*, 2004) ou plus distantes (par exemple : Cenci *et al.*, 2010 ; Timms *et al.*, 2006). Elle peut être étudiée à l'échelle du génome ou de chromosomes entiers (macrosynténie) ou à une échelle plus réduite de quelques Mb, voir quelques kb (microsynténie).

A partir des gènes d'*A. thaliana* et de vigne, présentant des homologies avec les séquences de chicorée, nous avons tenté de distinguer ceux pouvant être considérés comme des orthologues. S'il est facile de définir la notion d'orthologie en théorie, il est plus difficile d'identifier 2 gènes orthologues, d'autant plus que les espèces comparées sont

éloignées. Un des facteurs cruciaux dans l'identification d'orthologues est la réciprocité du résultat de meilleur homologue, mais cela implique de disposer de la totalité du génome des espèces comparées, ce qui n'est pas le cas de la chicorée (Koonin, 2005). Cependant, disposant d'importantes collections d'EST de *C. intybus* et *C. endivia*, ces ressources ont été utilisées pour réaliser des Blast inverses, à partir de certains gènes d'*A. thaliana* et de *V. vinifera*. Environ la moitié des gènes testés ont été identifiés comme les homologues les plus proches des séquences de chicorée étudiées. Néanmoins, malgré l'importance des banques d'EST de *Cichorium sp.* (129 106 SU au total), des séquences absentes de ces banques pourraient présenter une meilleure homologie avec ces gènes. L'autre moitié des gènes a présenté une homologie plus forte pour des séquences différentes de celles associées aux clones BAC. Dans ce cas, cela signifie que les gènes identifiés sur les séquences à proximité du locus *S* ne sont pas des copies uniques. Ces résultats ne sont pas surprenant, notamment lorsque de nombreux résultats avaient été obtenus à partir des Blastx (par exemple pour les positions 23 ou 29, cf. tableau 7.3 p.161).

Les résultats obtenus lors de la comparaison avec le génome de vigne ont mis en évidence une colinéarité relative entre des gènes du chromosome 18 de la vigne et leurs homologues sur la carte physique de la chicorée. En effet, pour chaque clone (Si10, Si42 et Si26) ou groupe de clones (Si37, Si52 et Si27), un bloc comprenant des gènes potentiellement orthologues et homologues a été identifié. Ces différents groupes de gènes sont séparés les uns des autres par plusieurs Mb et distribués sur un intervalle total de plus de 12 Mb. Il existe donc une importante différence entre la taille estimée de la région du locus *S* (4,7 Mb, cf. figure 7.3.4.2 p.171) et la région colinéaire sur le chromosome 18 de la vigne. Cependant, au niveau local, les distances sur lesquelles sont positionnés ces blocs de gènes avoisinent les 200 kb en moyenne, ce qui est comparable aux tailles des séquences des 6 clones étudiés.

Ces observations suggèrent que, pour étudier la conservation de la synténie, il n'est pas nécessaire de sélectionner uniquement les gènes considérés comme orthologues mais plutôt de considérer tous les gènes (homologues et orthologues) appartenant au même bloc chez l'espèce utilisée pour la comparaison (Koonin, 2005). Par conséquent, les gènes précédemment écartés de nos analyses, car présentant un alignement trop faible, devront être également pris en considération.

Des blocs de gènes homologues ont également été mis en évidence chez 4 autres espèces, 2 Astéridées (*M. guttatus* et *S. lycopersicum*) et 2 Rosidées (*P. trichocarpa* et

M. truncatula), en réalisant des Blast des SU de chicorée sur les banques de peptides. Un ordre ABCD représentatif des positions relatives des clones Si10, Si42, du groupe de clone Si37-Si52-Si27 et du clone Si26 sur la carte physique de la chicorée a été défini, chaque lettre étant associée au bloc de gènes homologues pour les différentes espèces étudiées. Les chromosomes, groupes de liaison ou scaffolds présentant le plus grand nombre de blocs ont été schématisés sur la figure 8.1.

Chez *M. truncatula*, la conservation de la synténie a été observée principalement pour 2 blocs de gènes répartis sur le chromosome 1. Un intervalle de plus de 22 Mb sépare le bloc A (région au dessus du locus *S*) des blocs B et C (régions sous le locus *S*). Comme chez la vigne, la distribution des blocs est observée sur une grande portion du chromosome. Chez *M. guttatus* et *P. trichocarpa*, la conservation de la colinéarité des gènes a été observée sur plusieurs groupes de liaison ou scaffold. Dans de nombreux cas, le premier et le second meilleur résultat étaient relativement proches en terme de e-value. Ils correspondaient de façon récurrente à 2 régions, l'une et l'autre étant tour à tour meilleur ou second résultat. Il s'agit probablement de régions dupliquées et colinéaires chez ces espèces. Chez la tomate et *M. guttatus*, les 2 espèces appartenant au clade des Astéridées, les gènes correspondant aux blocs A, B et C se sont avérés regroupés sur le chromosome 4 chez la tomate et le scaffold 2 pour *M. guttatus*, sur des régions respectivement de 5,26 Mb et 2,22 Mb. Chez *P. trichocarpa*, le groupe de liaison 2 réunit des blocs A, B, C et D, le bloc A étant séparés des blocs B, C et D par près de 4 Mb.

Pour ces 3 espèces, la répartition des blocs est observée sur des régions chromosomiques de taille comparable à celle estimée chez la chicorée, leurs génomes semblent donc de bons candidats pour étudier la conservation de la synténie avec le génome de la chicorée. L'ordre ABC a été conservé uniquement chez la tomate, alors que chez les autres espèces (*V. vinifera*, *M. guttatus* et *P. trichocarpa*) on observe une inversion commune, avec une disposition de type BCDA ou BCA des groupes de gènes. *M. guttatus* et la tomate sont des espèces appartenant au même sous-clade (Euasteridée I). Cependant, elles présentent une organisation différente des blocs de gènes. Néanmoins, l'assemblage du génome de *M. guttatus* en chromosomes n'étant pas encore achevé, l'organisation des blocs de gènes est à considérer avec prudence. La chicorée (Euastéridée II) et la tomate, qui appartiennent à des sous-clades différents, semblent présenter une organisation similaire des blocs de gènes étudiés.

La marche chromosomique vers le locus *S* a été orientée à partir du marqueur TTAA

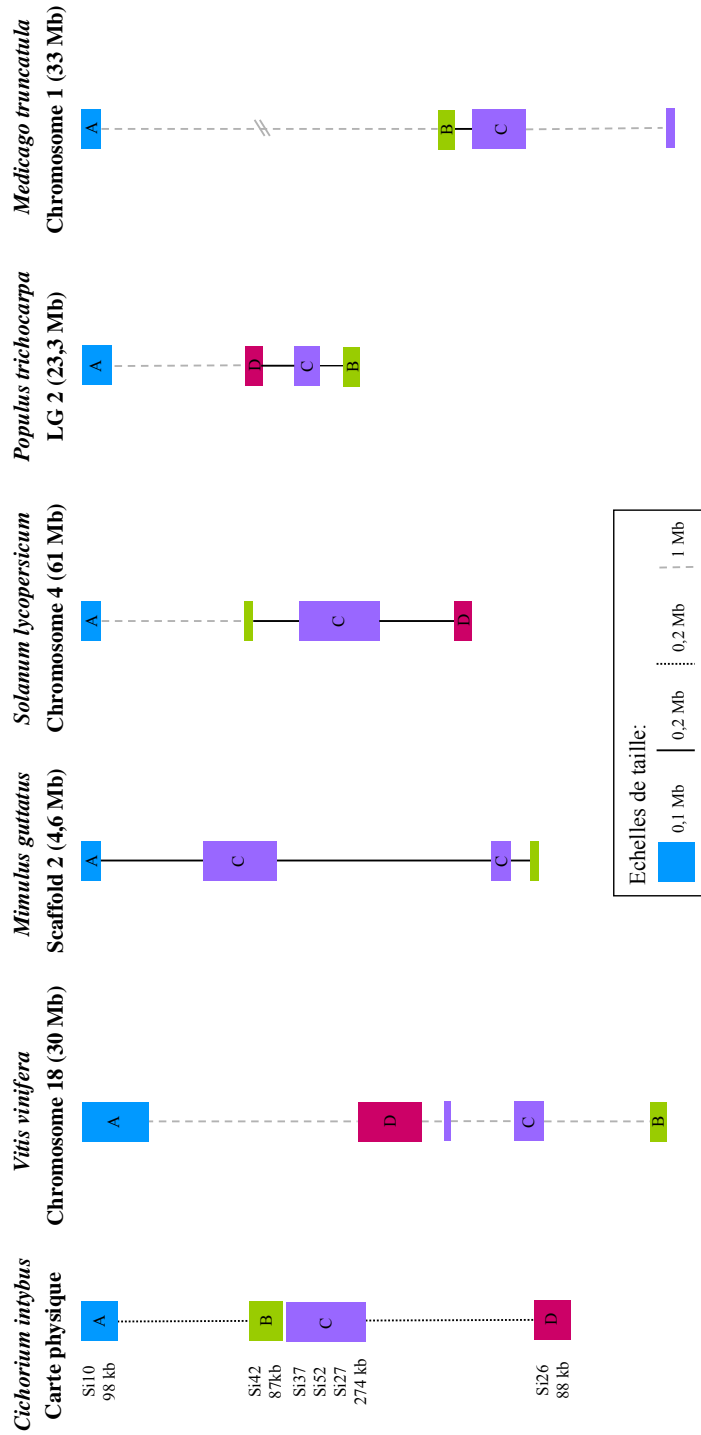


FIGURE 8.1 – Représentation schématique de la position des blocs de gènes (A à D), identifiés sur les clones BAC de chicorée, sur les chromosomes et scaffolds de *V. vinifera*, *M. guttatus*, *S. lycopersicum*, *P. trichocarpa* et *M. truncatula*. La taille des blocs de gènes et les distances les séparant sont représentés selon les échelles données au bas de la figure. Lorsqu'un seul gène est associé à un bloc, il est représenté par un rectangle de couleur ne contenant pas de lettre. Les distances entre les blocs sur la carte physique de la chicorée ont été estimées en considérant une relation Mb/cM homogène sur l'intervalle TAGG-EU12G02.

et va se poursuivre jusqu'à ce que la région d'intérêt soit identifiée. De plus, l'utilisation de bases de données d'EST de chicorée et des bases de peptidiques d'espèces relativement proches (Astéridées) ou plus distantes (Rosidées) a montré que celles-ci représentaient des ressources importantes pour l'identification de gènes copies uniques.

Cela signifie qu'avec l'acquisition de nouvelles séquences de chicorée, l'ensemble de ces ressources permettront d'identifier des points d'ancrages afin de passer d'une banque BAC à l'autre. Enfin, les résultats, obtenus à partir de l'étude de la fraction codante, suggèrent que la synténie peut être conservée entre des séquences de chicorée et celles d'au moins 5 autres espèces (*S. lycopersicum*, *M. guttatus*, *V. vinifera*, *P. trichocarpa* et *M. truncatula*).

Conclusions et perspectives

Conclusions et perspectives

L'objectif de cette thèse était d'initier le clonage positionnel du locus *S* chez la chicorée. Celui-ci avait été positionné à l'extrémité du groupe de liaison 2 et une carte moyenne densité de cette région, réalisée à partir de la population B_a (n = 389), était en cours. Afin d'atteindre l'objectif fixé, nous avons entrepris 1) de construire une carte génétique fine et 2) d'initier la construction d'une carte physique de la région du locus *S*.

Pour réaliser une carte génétique fine, la carte moyenne densité à d'abord été achevée en complétant le génotype de tous les individus de la population B_a avec les marqueurs existants (*cf.* chapitre 2). Cette carte a permis d'identifier 25 plantes recombinantes dans la région du locus *S* et a mis en évidence une différence significative entre les taux de recombinaison des méioses de K28 et K59.

Une stratégie de densification de l'intervalle étudié en marqueurs moléculaires a ensuite été mise en place. Cinq nouveaux marqueurs AFLP ont été identifiés, dont un coségrégant avec le locus *S* et 2 ayant été transformés en marqueurs SCAR codominants. L'approche choisie (BSA assistée par marqueurs, *cf.* chapitre 3) s'est révélée efficace, puisque la densité en marqueurs dans l'intervalle ciblé est comparable à celle attendue pour l'ensemble du génome.

Simultanément, 7 populations, représentant un effectif total de 2824 individus, ont été analysées afin de construire une carte haute résolution de la région du locus *S* (*cf.* chapitre 4). Pour les 7 populations étudiées, le locus *S* a toujours été cartographié entre les marqueurs EU08B02 et TTAA, à l'extrémité du groupe de liaison 2. Le génotypage de ces plantes a donc confirmé le caractère monocus des déterminants de l'autoincompatibilité chez la chicorée et la pertinence du choix de ces populations. Deux cartes consensus ont été construites à partir de l'effectif total, l'une positionnant le locus *S* dans un intervalle de 1,24 cM (méioses mâles), et l'autre, dans un intervalle de 0,6 cM (méioses femelles). Ces cartes, ainsi que les 169 individus recombinants dans la région

Conclusions et perspectives

du locus *S*, constituent un outil puissant pour positionner de nouveaux marqueurs moléculaires, puisqu'au moins 2 événements de recombinaison ont été observés dans un intervalle de moins de 50 kb. Enfin, l'ensemble des données génétiques ont démontré que les différences, observées dans un premier temps entre les taux de recombinaison de K28 et K59, étaient liées au sexe. Cette différence n'est donc pas directement associée au locus *S*. Elle a toutefois un impact sur l'estimation de la progression de la marche chromosomique à partir de la carte génétique, selon que l'on considère les méioses mâles ou les méioses femelles, les premières étant une source d'information plus importante, jusqu'à présent.

Afin d'initier la construction de la carte physique du locus *S*, 2 banques BAC ont été générées et représentent une couverture de 12 fois du génome haploïde de la chicorée (*cf.* chapitre 7). L'une (CinS1S4) a été construite à partir d'un génotype hétérozygote dans la région du locus *S* et l'autre (CinS2S2) à partir d'un génotype homozygote. Trois contigs de clones BAC ont été construits de part et d'autre du locus *S*, pour l'haplotype *S*₂. L'orientation de l'un d'entre eux, le supercontig positionné sous le locus *S*, a été définie à l'aide de la carte génétique. L'identification de recombinants, entre les marqueurs Si30t7 et aacc134 présents sur le supercontig, a permis de montrer la progression de la marche chromosomique : la marche chromosomique vers le locus *S* est donc en bonne voie. En raison des difficultés rencontrées pour organiser les clones d'haplotypes différents en un contig commun, et dans un souci d'efficacité, la marche chromosomique a été entreprise à partir de clones issus de la banque CinS2S2. Cela souligne l'importance de disposer d'au moins une banque BAC issue d'un individu homozygote. Le criblage des banques a été réalisé à partir d'hybridation sur macroarrays. Deux sondes obtenues à partir de marqueurs AFLP, notamment TCAA qui coségrège avec le locus *S*, ont présentés des hybridations aspécifiques. Cette méthode de crible est moins coûteuse mais également moins sensible qu'un criblage par PCR. Toutefois, si la marche chromosomique devait être interrompue, en raison de sondes dont l'hybridation est aspécifique, une nouvelle approche devra être adoptée. Par exemple, la mise en place de pools d'ADN de clones BAC permettrait de cribler les banques par PCR, à l'aide d'amorces très spécifiques de la séquence ciblée.

Le séquençage de 6 clones identifiés à proximité du locus *S* a été réalisé (*cf.* chapitre 7). Ces séquences ont permis, pour la première fois, d'étudier et de caractériser des séquences génomiques de chicorée. Un nombre important de Cg a été obtenu pour chaque séquence (de 7 à 23). Ces séquences ont toutefois permis une première estimation

du rapport Mb/cM au voisinage du locus *S* (environ 0,5 Mb/cM pour les méioses mâles et 2,2 Mb/cM pour les méioses femelles). La composition en éléments répétés a été déterminée : 5,71% d'éléments mobiles, 1,74% de SSR, 2,75% de régions de faible complexité et 10,45% de régions répétées sans homologies. Des régions potentiellement codantes ou exprimées ont été identifiées à partir des banques d'EST du genre *Cichorium* (59 SU) et des banques de peptides d'*A. thaliana* (23 gènes) et de *V. vinifera* (27 gènes). Parmi les gènes potentiels identifiés, au moins 8 d'entre eux correspondraient à des gènes copie unique chez la chicorée. Les difficultés rencontrées avec la réalisation des contigs de clones provenant de différents haplotypes, lors de la marche chromosomique, sont sans doute causées par la présence de séquences répétées et d'éléments transposables au niveau des BES. Le fait de disposer de séquences complètes de clones permet d'identifier des régions potentiellement codantes et copies uniques, pouvant servir de points d'ancrage pour passer d'une banque et d'un haplotype à l'autre. L'étude de la fraction codante a également mis en évidence une conservation de la microsyténie entre les séquences de chicorée et celles d'au moins 5 autres espèces de Dicotylédones (*V. vinifera*, *S. lycopersicum*, *M. guttatus*, *P. trichocarpa* et *M. truncatula*). Ces résultats montrent qu'en l'absence de données de séquences de génomes d'Asteraceae, il est possible d'exploiter les informations de séquences d'espèces éloignées phylogénétiquement, tout en gardant à l'esprit que ces ressources ne contiennent pas nécessairement des gènes homologues aux déterminants de l'AI chez la chicorée.

Au vu de l'ensemble de ces résultats, les objectifs fixés au début de la thèse ont été atteints. La marche chromosomique se poursuit en direction du locus *S*. La stratégie consistant à construire des contigs en utilisant les BES a prouvé son efficacité et constitue la base de la démarche. Toutefois, les données préliminaires acquises récemment ouvrent la voie à de nouvelles approches qui pourraient accélérer la stratégie mise en place. La conservation de la synténie entre la position de gènes au voisinage du locus *S* et leur position sur les chromosomes d'autres espèces devra être approfondie en ce sens. Pour initier cette démarche, la vigne et la tomate semblent les espèces les plus prometteuses. En effet, la vigne est l'espèce qui regroupe le plus grand nombre de gènes associés à un même chromosome. La tomate, quant à elle, est la seule espèce pour laquelle les blocs de gènes synténiques sont à la fois répartis dans un intervalle de quelques Mb et organisés dans le même ordre que chez la chicorée. En recensant, chez la tomate, l'ensemble des gènes présents dans l'intervalle de 4 Mb qui sépare les blocs de gènes A

Conclusions et perspectives

et B (*cf.* figure 8.1 p.183), il devrait être possible d'identifier des gènes potentiellement copies uniques. Les SU de chicorée, homologues de ces gènes, seront alors utilisées pour rechercher du polymorphisme, afin de définir de nouveaux marqueurs moléculaires. Une partie des marqueurs ainsi définis sera potentiellement cartographiée dans la région du locus S . Si ces marqueurs sont positionnés à de plus faibles distances génétiques du locus S que les marqueurs dont on dispose actuellement, ils permettraient d'atteindre plus rapidement le locus S . Cela permettrait de reprendre la marche chromosomique dans la région au dessus du locus S , au-delà du marqueur EU08B02, pour lequel aucun clone n'a pu être validé (*cf.* figure 6.3 p.139).

Enfin, lorsque la séquence du génome du tournesol sera disponible, il sera intéressant de déterminer dans quelle mesure la synténie est conservée avec la région du locus S de la chicorée. Il sera également possible de déterminer si la région du locus S et la région du QTL de l'autofertilité, sur le groupe de liaison 17 du tournesol, correspondent.

En prévision de l'obtention de séquences de la région du locus S , il sera également important de résoudre le problème d'assemblage des séquences génomiques. L'utilisation simultanée de différents logiciels d'assemblages semble être une première piste à explorer pour permettre un meilleur assemblage, en particulier pour réduire le nombre de Cg par clone. L'efficacité de ce type de stratégie a été mise en évidence pour des analyses transcriptomiques (Kumar et Blaxter, 2010) et les premiers tests réalisés chez la chicorée semblent prometteurs. Cependant, si aucun protocole satisfaisant n'est obtenu, des stratégies de séquençage complémentaires devront être envisagées pour ordonner les Cg.

Lorsque la carte physique de l'haplotype S_2 aura suffisamment progressé, les séquences génomiques correspondant aux haplotypes S_1 et S_4 pourront être recherchées. Pour cela, l'utilisation de séquences correspondant à des gènes copies uniques sera privilégiée. En considérant les caractéristiques communes aux différents locus S identifiés jusqu'à présent, la comparaison des séquences des différents haplotypes S présents dans les banques BAC (S_1 , S_2 et S_4) devrait permettre de délimiter le locus S dans une région hypervariable. L'identification des déterminants mâles et femelles nécessitera la mise en œuvre d'outils de prédiction de gènes. Des paramètres adaptés et spécifiques à la chicorée devront être définis pour une utilisation optimale des logiciels de prédiction. Au sein de cette région, les gènes prédits les plus polymorphes représenteront les candidats les plus probables des déterminants de l'autoincompatibilité. L'association du polymorphisme de ces gènes à différents haplotypes et leur spécificité d'expression, dans

le stigmate ou l'anthère, devront également être prises en compte lors de la sélection des candidats. De plus, la présence spécifique de séquences homologues dans la banque d'EST de stigmates constituera un argument supplémentaire dans le choix des candidats femelles à valider.

Enfin, la transformation de plantes par gain de fonction représentera l'étape finale de la validation fonctionnelle de ces gènes. L'haplotype S_2 étant le plus récessif dans les populations étudiées, il offre la possibilité d'obtenir des individus homozygotes S_2S_2 , parmi lesquels il sera nécessaire de sélectionner ceux qui possèdent une bonne aptitude à la transformation génétique. Les gènes candidats correspondant à des haplotypes dominants (par exemple S_1) seront utilisés pour la transformation. Seuls les individus transformés avec les déterminants de l'AI devront alors présenter un phénotype S1, au niveau du pollen ou du stigmate, selon le candidat, mâle ou femelle, que l'on cherche à valider.

A terme, l'accès aux séquences des déterminants mâle et femelle chez la chicorée devrait ouvrir la voie à l'étude de l'AI chez les Asteraceae. Cela permettrait de déterminer si ce mécanisme est commun aux différentes espèces d'Asteraceae autoincompatibles, notamment chez *S. squalidus*, espèce pour laquelle plusieurs haplotypes S ont été caractérisés. Si le déterminisme s'avère commun, la diversité intra- et inter-spécifique au niveau moléculaire ainsi que les différences qui existent entre espèces AI et allogames d'un même genre pourront être explorées.

L'ensemble de ces informations apportera peut être des indices permettant d'expliquer l'apparition de l'autoincompatibilité ou de comprendre quels facteurs sont responsables de sa rupture chez les Asteraceae. L'étude de la variabilité pour le caractère autoincompatible chez la chicorée (présence d'individus pseudocompatibles et autofertiles) permettra d'en définir l'origine. Soit la rupture de l'AI est causée par la perte de fonction des déterminants de l'AI pour certains haplotypes, soit il s'agit de l'action d'un ou plusieurs gènes modulant la réponse AI.

Conclusions et perspectives

Bibliographie

- Allard, R. W. Formulas and tables to facilitate the calculation of recombination values in heredity. *Hilgardia*, 24(10) :235–271, 1956.
- Allen, A. M., Lexer, C., and Hiscock, S. J. Comparative Analysis of Pistil Transcriptomes Reveals Conserved and Novel Genes Expressed in Dry, Wet, and Semidry Stigmas. *Plant Physiology*, 154(3) :1347–1360, 2010.
- Allen, A. and Hiscock, S. Evolution and phylogeny of self-incompatibility systems in angiosperms. In *Self-Incompatibility in Flowering Plants*, pages 73–101. Springer-Verlag Berlin Heidelberg, 2008.
- Anderson, M. A., Cornish, E. C., Mau, S. L., Williams, E. G., Hoggart, R., Atkinson, A., Bonig, I., Grego, B., Simpson, R., Roche, P. J., Haley, J. D., Penschow, J. D., Niall, H. D., Tregear, G. W., Coghlan, J. P., Crawford, R. J., and Clarke, A. E. Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana glauca*. *Nature*, 321(6065) :38–44, 1986. URL <http://dx.doi.org/10.1038/321038a0>.
- Barker, M. S., Kane, N. C., Matvienko, M., Kozik, A., Michelmore, R. W., Knapp, S. J., and Rieseberg, L. H. Multiple paleopolyploidizations during the evolution of the compositae reveal parallel patterns of duplicate gene retention after millions of years. *Molecular Biology and Evolution*, 25(11) :2445–2455, 2008. URL <http://www.scopus.com/inward/record.url?eid=2-s2.0-54149092976&partnerID=40&md5=a55f317d0dc4fe19d927882379282c3c>.
- Barrett, S. C. H., Jesson, L. K., and Baker, A. M. The evolution and function of stylar polymorphisms in flowering plants. *Annals of Botany*, 85 :253–265, 2000. URL <GotoISI>://000086316900031.
- Barrett, S. and Shore, J. New Insights on Heterostyly : Comparative Biology, Ecology and Genetics. In *Self-Incompatibility in Flowering Plants*, pages 3–32. Springer-Verlag Berlin Heidelberg, 2008.
- Beardsley, P. M. and Olmstead, R. G. Redefining Phrymaceae : the placement of Mimulus, tribe Mimuleae, and Phryma. *American Journal of Botany*, 89(7) :1093–1102, 2002.
- Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., and Taberlet, P. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, 13(11) :3261–3273, 2004. doi : 10.1111/j.1365-294X.2004.02346.x. URL <http://dx.doi.org/10.1111/j.1365-294X.2004.02346.x>.
- Bredemeijer, G. M. M. and Blaas, J. S-specific proteins in styles of self-incompatible *Nicotiana glauca*. *Theoretical and Applied Genetics*, 59(3) :185–190, 1981. URL <GotoISI>://A1981LH73000008.
- Bremer, B., Bremer, K., Chase, M. W., Fay, M. F., Reveal, J. L., Soltis, D. E., Soltis, P. S., Stevens, P. F., Anderberg, A. A., Moore, M. J., Olmstead, R. G., Rudall, P. J., Sytsma, K. J., Tank, D. C., Wurdack, K., Xiang, J. Q. Y., and Zmarzty, S. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants :

- APG III. *Botanical Journal of the Linnean Society*, 161(2) :105–121, 2009. URL <GotoISI> : //000270958200001.
- Brugmans, B., van der Hulst, R. G. M., Visser, R. G. F., Lindhout, P., and van Eck, H. J. A new and versatile method for the successful conversion of AFLPTM markers into simple single locus markers. *Nucleid Acids Research*, 31(10) :e55–, 2003. URL <http://nar.oxfordjournals.org/cgi/content/abstract/31/10/e55>.
- Burke, J. M., Tang, S., Knapp, S. J., and Rieseberg, L. H. Genetic analysis of sunflower domestication. *Genetics*, 161(3) :1257–1267, 2002.
- Cadalen, T., Morchen, M., Blassiau, C., Clabaut, A., Scheer, I., Hilbert, J. L., Hendriks, T., and Quillet, M. C. Development of SSR markers and construction of a consensus genetic map for chicory (*Cichorium intybus* L.). *Molecular Breeding*, 25(4) :699–722, 2010. doi : 10.1007/s11032-009-9369-5. URL <GotoISI> : //000275749500013.
- Cankar, K., van Houwelingen, A., Bosch, D., Sonke, T., Bouwmeester, H., and Beekwilder, J. A chicory cytochrome P450 mono-oxygenase CYP71AV8 for the oxidation of (+)-valencene. *FEBS Letters*, 585(1) :178–182, 2011.
- Casey, N. M., Milbourne, D., Barth, S., Febrer, M., Jenkins, G., Abberton, M. T., Jones, C., and Thorogood, D. The genetic location of the self-incompatibility locus in white clover (*Trifolium repens* L.). *Theoretical and Applied Genetics*, 121(3) :567–576, 2010. doi : 10.1007/s00122-010-1330-9. URL <GotoISI> : //000280090700012.
- Casselmann, A. L., Vrebalov, J., Conner, J. A., Singhal, A., Giovannoni, J., Nasrallah, M. E., and Nasrallah, J. B. Determining the physical limits of the *Brassica* S locus by recombinational analysis. *Plant Cell*, 12(1) :23–33, 2000. URL <GotoISI> : //000085061300002.
- Castaño, C. I., Demeulemeester, M. A. C., and De Proft, M. P. Incompatibility reactions and genotypic identity status of five commercial chicory (*Cichorium intybus* L.) hybrids. *Scientia Horticulturae*, 72(1) :1–9, 1997. doi : 10.1016/S0304-4238(97)00111-8. URL <GotoISI> : //000071694100001.
- Castric, V. and Vekemans, X. Plant self-incompatibility in natural populations : a critical assessment of recent theoretical and empirical advances. *Molecular Ecology*, 13(10) :2873–2889, 2004. URL <GotoISI> : //000223863000001.
- Cavagnaro, P. F., Chung, S. M., Szklarczyk, M., Grzebelus, D., Senalik, D., Atkins, A. E., and Simon, P. W. Characterization of a deep-coverage carrot (*Daucus carota* L.) BAC library and initial analysis of BAC-end sequences. *Molecular Genetics and Genomics*, 281(3) :273–288, 2009. doi : 10.1007/s00438-008-0411-9. URL <GotoISI> : //000263684900004.
- Cenci, A., Combes, M. C., and Lashermes, P. Comparative sequence analyses indicate that *Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleo-hexaploid ancestral genome. *Molecular Genetics and Genomics*, 283(5) :493–501, 2010. doi : 10.1007/s00438-010-0534-7. URL <GotoISI> : //000276661600008.

- Charlesworth, B. and Charlesworth, D. Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, 18 :237–268, 1987.
- Cnops, G., denBoer, B., Gerats, A., VanMontagu, M., and VanLijsebettens, M. Chromosome landing at the *Arabidopsis* TORNAD01 locus using an AFLP-based strategy. *Molecular and General Genetics*, 253(1-2) :32–41, 1996. URL <GotoISI>://A1996VX33900005.
- Coppens d'Eeckenbrugge, G. The progamic phase in *Cichorium intybus* L. Pollen tube growth in the style, incompatibility reaction and gametophytic competition. *Euphytica*, 48 :17–23, 1990.
- Coppens d'Eeckenbrugge, G. and Evrard, B. Quantitative aspects of *Cichorium intybus* L. self-incompatibility .2. Achene distribution among self-pollinized capitula of the same plant. *Revue de l'Agriculture*, 41(1) :59–62, 1988. URL <GotoISI>://A1988N770200005.
- Coppens d'Eeckenbrugge, G., Evrard, B., and Louant, B.-P. Aspects quantitatifs de l'auto-incompatibilité chez la chicorée de witloof (*Cichorium intybus* L.). Relation entre fertilité et autofertilité. *Revue de l'Agriculture*, 40(1) :35–40, 1987.
- Darracq, A., Varre, J. S., and Touzet, P. A scenario of mitochondrial genome evolution in maize based on rearrangement events. *BMC Genomics*, 11, 2010. doi : 23310.1186/1471-2164-11-233. URL <GotoISI>://000277049200002.
- Dauchot, N., Mingeot, D., Purnelle, B., Muys, C., Watillon, B., Boutry, M., and Van Cutsem, P. Construction of 12 EST libraries and characterization of a 12,226 EST dataset for chicory (*Cichorium intybus*) root, leaves and nodules in the context of carbohydrate metabolism investigation. *BMC Plant Biology*, 9 :14, 2009. doi : 10.1186/1471-2229-9-14. URL <GotoISI>://000265062000001.
- de Brito, G. G., Caixeta, E. T., Gallina, A. P., Zambolim, E. M., Zambolim, L., Diola, V., and Loureiro, M. E. Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. *Euphytica*, 173(2) :255–264, 2010. doi : 10.1007/s10681-010-0119-x. URL <GotoISI>://000276479200011.
- de Nettancourt, D. *Incompatibility in angiosperms*. Springer-Verlag, New York, USA, 1977.
- de Nettancourt, D. Incompatibility in angiosperms. *Sexual Plant Reproduction*, 10(4) :185–199, 1997. URL <GotoISI>://A1997XW43800001.
- de Nettancourt, D. *Incompatibility and incongruity in wild and cultivated plants*. Springer, Berlin, 2nd edition, 2001.
- De Simone, M., Morgante, M., Lucchin, M., Parrini, P., and Marocco, A. A first linkage map of *Cichorium intybus* L. using a one-way pseudo-testcross and PCR-derived markers. *Molecular Breeding*, 3(6) :415–425, 1997. URL <GotoISI>://000072253900002.
- De Vienne, D. *Les marqueurs moléculaires en génétique et biotechnologies végétales*. INRA, 2e édition, 1998.

- Doré, C. and Varoquaux, F. La chicorée endive ou witloof. In Quae, editor, *Histoire et amélioration de cinquante plantes cultivées*, pages 207–217. 2006.
- Doré, C., Prigent, J., and Desprez, B. *In situ* gynogenetic haploid plants of chicory (*Cichorium intybus* L.) after intergeneric hybridization with *Cicerbita alpina* Walbr. *Plant Cell Reports*, 15(10) :758–761, 1996. URL <GotoISI>://A1996UT74700008.
- Drouaud, J., Mercier, R., Chelysheva, L., Berard, A., Falque, M., Martin, O., Zanni, V., Brunel, D., and Mezard, C. Sex-specific crossover distributions and variations in interference level along *Arabidopsis thaliana* chromosome 4. *Plos Genetics*, 3(6) :1096–1107, 2007. URL <GotoISI>://000248349300022.
- Eenink, A. H. Compatibility and incompatibility in witloof-chicory (*Cichorium intybus* L.). 2. The incompatibility system. *Euphytica*, 30(1) :77–85, 1981a. doi : 10.1007/BF00033662. URL <http://dx.doi.org/10.1007/BF00033662>.
- Eenink, A. H. Compatibility and incompatibility in witloof-chicory (*Cichorium intybus* L.). 1. The influence of temperature and plant age on pollen germination and seed production. *Euphytica*, 30(1) :71–76, 1981b. URL <http://dx.doi.org/10.1007/BF00033661>.
- Eenink, A. H. Compatibility and incompatibility in witloof-chicory (*Cichorium intybus* L.). 3. Gametic competition after mixed pollinations and double pollinations. *Euphytica*, 31(3) : 773–786, 1982. URL <http://dx.doi.org/10.1007/BF00039217>.
- Eenink, A. H. Compatibility and incompatibility in witloof chicory (*Cichorium intybus* L.). 4. Formation of self-seeds on a self-incompatible and a moderately self-compatible genotype after double and triple pollinations. *Euphytica*, 33(1) :161–167, 1984. URL <http://dx.doi.org/10.1007/BF00022762>.
- Entani, T., Iwano, M., Shiba, H., Che, F. S., Isogai, A., and Takayama, S. Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume* : identification of a pollen-expressed F-box gene with allelic diversity. *Genes to Cells*, 8(3) :203–213, 2003. URL <GotoISI>://000181371500001.
- Feng, Q., Huang, T., Zhao, Q., Zhu, J. J., Lin, Z. X., and Han, B. Analysis of collinear regions of *Oryza* AA and CC genomes. *Journal of Genetics and Genomics*, 36(11) :667–677, 2009. doi : 10.1016/s1673-8527(08)60159-9. URL <GotoISI>://000272420400004.
- Ferrer, M. M. and Good-Avila, S. V. Macrophylogenetic analyses of the gain and loss of self-incompatibility in the Asteraceae. *New Phytologist*, 173(2) :401–414, 2007. URL <GotoISI>://000242872500016.
- Foote, H. C. C., Ride, J. P., Franklinton, V. E., Walker, E. A., Lawrence, M. J., and Franklin, F. C. H. Cloning and Expression of a Distinctive Class of Self-Incompatibility (S) Gene from *Papaver-Rhoeas* L. *Proceedings of the National Academy of Sciences of the United States of America*, 91(6) :2265–2269, 1994. URL <GotoISI>://A1994NC04300062.

- Franklin-Tong, N. and Franklin, F. C. H. Gametophytic self-incompatibility inhibits pollen tube growth using different mechanisms. *Trends in Plant Science*, 8(12) :598–605, 2003. URL <GotoISI>://000187629000009.
- Funk, V. A., Bayer, R. J., Keeley, S., Chan, R., Watson, L., Gemeinholzer, B., Schilling, E., Panero, J. L., Baldwin, B. G., Garcia-Jacas, N., Susanna, A., and Jansen, R. K. Everywhere but Antarctica : using a supertree to understand the diversity and distribution of the Compositae. *Biologische Skriften*, 55 :343–374, 2005.
- Funk, V. A., Susanna, A., Steussy, T. F., and Robinson, H. E. *Systematics, Evolution, and Biogeography of Compositae*. International Association for Plant Taxonomy, Vienna, Austria, 2009.
- Gandhi, S. D., Heesacker, A. F., Freeman, C. A., Argyris, J., Bradford, K., and Knapp, S. J. The self-incompatibility locus (S) and quantitative trait loci for self-pollination and seed dormancy in sunflower. *Theoretical and Applied Genetics*, 111(4) :619–629, 2005. URL <GotoISI>://000231313100001.
- Gaude, T. and Cabrillac, D. Self-incompatibility in flowering plants : The *Brassica* model. *Comptes Rendus de l'Académie des Sciences*, 324(6) :537–542, 2001. URL <GotoISI>://000169785600006.
- Gaude, T., Glemin, S., Cabrillac, D., and Mignot, A. L'autoincompatibilité chez les plantes à fleurs. *Médecine/sciences*, 17(6-7), 2001.
- Ge, R.-C., Zhao, M.-L., Gao, H.-W., Zhao, B.-C., and Li, G.-L. Study on C-banding and karyotype analysis of Puna Chicory. *Acta Agrestia Sinica*, 10(3) :190–193, 2002.
- Gerstel, D. U. Self-incompatibility studies in guayule II. Inheritance. *Genetics*, 35(4) :482–506, 1950.
- Gibbs, P. E. Do homomorphic and heteromorphic self-incompatibility systems have the same sporophytic mechanism. *Plant Systematics and Evolution*, 154(3-4) :285–323, 1986. URL <GotoISI>://A1986F736000010.
- Gibbs, P. E. and Bianchi, M. B. Does late-acting self-incompatibility (LSI) show family clustering? Two more species of Bignoniaceae with LSI : *Dolichandra cynanchoides* and *Tabebuia nodosa*. *Annals of Botany*, 84(4) :449–457, 1999. URL <GotoISI>://000083236200005.
- Giovannoni, J. J., Wing, R. A., Ganal, M. W., and Tanksley, S. D. Isolation of Molecular Markers from Specific Chromosomal Intervals Using DNA Pools from Existing Mapping Populations. *Nucleic Acids Research*, 19(23) :6553–6568, 1991.
- Glemin, S., Bataillon, T., Ronfort, J., Mignot, A., and Olivieri, I. Inbreeding depression in small populations of self-incompatible plants. *Genetics*, 159(3) :1217–1229, 2001. URL <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0035679022&partnerID=40&rel=R5.6.0>.

- Gonthier, L., Bellec, A., Blassiau, C., Prat, E., Helmstetter, N., Rambaud, C., Huss, B., Hendriks, T., Berges, H., and Quillet, M.-C. Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (*Cichorium intybus* L., Asteraceae). *BMC Research Notes*, 3(1) :225, 2010. URL <http://www.biomedcentral.com/content/3/1/225>.
- Goodwillie, C., Kalisz, S., and Eckert, C. The evolutionary enigma of mixed mating systems in plants : Occurrence, theoretical explanations, and empirical evidence. *Annual review of ecology and systematics*, 36 :47–79, 2005.
- Grover, A. and Sharma, P. C. Microsatellite Motifs with Moderate GC Content Are Clustered Around Genes on *Arabidopsis thaliana* Chromosome 2. *In Silico Biology*, 7(2) :201–213, 2007.
- Haque, M. Z. and Godward, M. B. E. Effects of seed-irradiation on meiosis in M1 plants of *Lactuca* and related plants. *Genetica*, 65 :179–186, 1984.
- Hauge, B. M., Hanley, S. M., Cartinhour, S., Cherry, J. M., Goodman, H. M., Koornneef, M., Stam, P., Chang, C., Kempin, S., Medrano, L., and Meyerowitz, E. M. An integrated genetic RFLP map of the *Arabidopsis thaliana* genome. *Plant Journal*, 3(5) :745–754, 1993. URL <GotoISI>://A1993LC75800013.
- Hedrick, P. W. Sex : differences in mutation, recombination, selection, gene flow, and genetic drift. *Evolution*, 61(12) :2750–2771, 2007. doi : 10.1111/j.1558-5646.2007.00250.x. URL <GotoISI>://000251250600003.
- Hiscock, S. J. Self-incompatibility in *Senecio squalidus* L (Asteraceae). *Annals of Botany*, 85 : 181–190, 2000a. URL <GotoISI>://000086316900022.
- Hiscock, S. J. Genetic control of self-incompatibility in *Senecio squalidus* L. (Asteraceae) : a successful colonizing species. *Heredity*, 85(1) :10–19, 2000b. URL <GotoISI>://000088938900002.
- Hiscock, S. J. and Tabah, D. A. The different mechanisms of sporophytic self-incompatibility. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 358 (1434) :1037–1045, 2003. URL <GotoISI>://000183891600011.
- Hiscock, S. J., Hoedemaekers, K., Friedman, W. E., and Dickinson, H. G. The stigma surface and pollen-stigma interactions in *Senecio squalidus* L. (Asteraceae) following cross (compatible) and self (incompatible) pollinations. *International Journal of Plant Sciences*, 163(1) : 1–16, 2002. URL <GotoISI>://000172542400001.
- Hiscock, S. J., McInnis, S. M., Tabah, D. A., Henderson, C. A., and Brennan, A. C. Sporophytic self-incompatibility in *Senecio squalidus* L. (Asteraceae) - the search for S. *Journal of Experimental Botany*, 54(380) :169–174, 2003. URL <GotoISI>://000179865500019.

- Hong, C. P., Plaha, P., Koo, D. H., Yang, T. J., Choi, S. R., Lee, Y. K., Uhm, T., Bang, J. W., Edwards, D., Bancrofts, I., Park, B. S., Lee, J., and Lim, Y. P. A survey of the *Brassica rapa* genome by BAC-End sequence analysis and comparison with *Arabidopsis thaliana*. *Molecules and Cells*, 22(3) :300–307, 2006. URL <GotoISI>://000243370500008.
- Hsu, C. C., Chung, Y. L., Chen, T. C., Lee, Y. L., Kuo, Y. T., Tsai, W. C., Hsiao, Y. Y., Chen, Y. W., Wu, W. L., and Chen, H. H. An overview of the *Phalaenopsis* orchid genome through BAC end sequence analysis. *BMC Plant Biology*, 11, 2011. doi : 310.1186/1471-2229-11-3. URL <GotoISI>://000286619300001.
- Hughes, M. B. and Babcock, E. B. Self-incompatibility in *Crepis foetida* L. subsp. *rhoeadifolia* Bieb Schinz et Keller. *Genetics*, 35(5) :570–588, 1950. URL <GotoISI>://A1950XW84200005.
- Igic, B., Bohs, L., and Kohn, J. Historical inferences from the self-incompatibility locus. *New Phytologist*, 161(1) :97–105, 2004. URL <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-0347388387&partnerID=40&rel=R5.6.0>.
- Imano, M., Sakamoto, K., Suzuki, G., Watanabe, M., Takayama, S., Fukui, K., Hinata, K., and Isogai, A. Visualization of a self-incompatibility gene in *Brassica campestris* L. by multicolor FISH. *Theoretical and Applied Genetics*, 96(6-7) :751–757, 1998. URL <GotoISI>://000074101500005.
- Initiative, T. A. G. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814) :796–815, 2000. URL <http://dx.doi.org/10.1038/35048692>http://www.nature.com/nature/journal/v408/n6814/supinfo/408796a0_S1.html.
- Ioerger, T. R., Clark, A. G., and Kao, T. H. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24) :9732–9735, 1990. URL <http://www.pnas.org/content/87/24/9732.abstract>.
- Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Lascou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pè, M. E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A. F., Weissenbach, J., Quétier, F., and Wincker, P. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449(7161) :463–467, 2007. URL <http://www.scopus.com/inward/record.url?eid=2-s2.0-34848886909&partnerID=40&md5=e89e390a219342d57c1fa9a9edb279ec>.
- Jones, C. J., Edwards, K. J., Castaglione, S., Winfield, M. O., Sala, F., vandeWiel, C., Bredemeijer, G., Vosman, B., Matthes, M., Daly, A., Brettschneider, R., Bettini, P., Buiatti, M., Maestri, E., Malcevski, A., Marmiroli, N., Aert, R., Volckaert, G., Rueda, J., Linacero,

- R., Vazquez, A., and Karp, A. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, 3(5) :381–390, 1997. URL <GotoISI>://A1997YB95300005.
- Jordan, T., Romer, P., Meyer, A., Szczesny, R., Pierre, M., Piffanelli, P., Bendahmane, A., Bonas, U., and Lahaye, T. Physical delimitation of the pepper Bs3 resistance gene specifying recognition of the AvrBs3 protein from *Xanthomonas campestris* pv. vesicatoria. *Theoretical and Applied Genetics*, 113(5) :895–905, 2006.
- Kakeda, K., Ibuki, T., Suzuki, J., Tadano, H., Kurita, Y., Hanai, Y., and Kowyama, Y. Molecular and genetic characterization of the S locus in *Hordeum bulbosum* L., a wild self-incompatible species related to cultivated barley. *Molecular Genetics and Genomics*, pages 1–11, 2008. URL <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-52249123054&partnerID=40&rel=R8.2.0>.
- Kakeda, K., Urabayashi, T., Ohashi, T., Oguro, T., and Kowyama, Y. Agrobacterium-mediated transformation of *Ipomoea trifida*, a diploid relative of sweet potato. *Breeding Science*, 59(1) :95–98, 2009. URL <GotoISI>://000265588000012.
- Kearsey, M. J., Ramsay, L. D., Jennings, D. E., Lydiate, D. J., Bohuon, E. J. R., and Marshall, D. F. Higher recombination frequencies in female compared to male meioses in *Brassica oleracea*. *Theoretical and Applied Genetics*, 92(3-4) :363–367, 1996. URL <GotoISI>://A1996UE86400011.
- Kejnovsky, E., Hobza, R., Cermak, T., Kubat, Z., and Vyskot, B. The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity*, 102(6) :533–541, 2009. doi : 10.1038/hdy.2009.17. URL <GotoISI>://000266199100003.
- Kiers, A. M. *Endive, Chicory and their wild relatives. A systematic and phylogenetic study of Cichorium (Asteraceae)*. Phd thesis, Leiden University, The Netherlands, 2000.
- Kiers, A. M., Mes, T. H. M., van der Meijden, R., and Bachmann, K. A search for diagnostic AFLP markers in *Cichorium* species with emphasis on endive and chicory cultivar groups. *Genome*, 43(3) :470–476, 2000. URL <GotoISI>://000087279800008.
- Klein, D. E., Freitas, L., and Da Cunha, M. Self-incompatibility in a distylous species of Rubiaceae : is there a single incompatibility response of the morphs? *Sexual Plant Reproduction*, 22(3) :121–131, 2009. doi : 10.1007/s00497-009-0097-0. URL <GotoISI>://000269152600002.
- Koonin, E. V. Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics*, 39 :309–338, 2005. doi : 10.1146/annurev.genet.39.073003.114725. URL <GotoISI>://000234685200015.
- Koopman, W. *Zooming in on the lettuce genome : species relationships in Lactuca s.l., inferred from chromosomal and molecular characters*. Phd thesis, Wageningen University, The Netherlands, 2002.

- Kowyama, Y., Kakeda, K., Kondo, K., Imada, T., and Hattori, T. A putative receptor protein kinase gene in *Ipomoea trifida*. *Plant and Cell Physiology*, 37(5) :681–685, 1996. URL <GotoISI>://A1996VA67300017.
- Kowyama, Y., Tsuchiya, T., and Kakeda, K. Sporophytic self-incompatibility in *Ipomoea trifida*, a close relative of sweet potato. *Annals of Botany*, 85 :191–196, 2000. URL <GotoISI>://000086316900023.
- Kumar, S. and Blaxter, M. L. Comparing *de novo* assemblers for 454 transcriptome data. *BMC Genomics*, 11(571), 2010.
- Labonne, J. D. J., Hilliker, A. J., and Shore, J. S. Meiotic recombination in *Turnera* (Turnera-ceae) : extreme sexual difference in rates, but no evidence for recombination suppression associated with the distyly (S) locus. *Heredity*, 98(6) :411–418, 2007. doi : 10.1038/sj.hdy.6800957. URL <GotoISI>://000246721700011.
- Labonne, J. D. J., Vaisman, A., and Shore, J. S. Construction of a first genetic map of distylous *Turnera* and a fine-scale map of the S-locus region. *Genome*, 51(7) :471–478, 2008. doi : 10.1139/g08-031. URL <GotoISI>://000256782400001.
- Lagercrantz, U. and Lydiate, D. J. RFLP mappong in *Brassica nigra* indicates differing recombinaion rates in male and femelle meiosis. *Genome*, 38(2) :255–264, 1995. URL <GotoISI>://A1995QX26100009.
- Lahaye, T., Hartmann, S., Topsis, S., Freialdenhoven, A., Yano, M., and Schulze-Lefert, P. High-resolution genetic and physical mapping of the Rar1 locus in barley. *Theoretical and Applied Genetics*, 97(4) :526–534, 1998. URL <GotoISI>://000076269000004.
- Lai, C. W. J., Yu, Q. Y., Hou, S. B., Skelton, R. L., Jones, M. R., Lewis, K. L. T., Murray, J., Eustice, M., Guan, P. Z., Agbayani, R., Moore, P. H., Ming, R., and Presting, G. G. Analysis of papaya BAC end sequences reveals first insights into the organization of a fruit tree genome. *Molecular Genetics and Genomics*, 276(1) :1–12, 2006. doi : 10.1007/s00438-006-0122-z. URL <GotoISI>://000238521500001.
- Lai, Z., Ma, W. S., Han, B., Liang, L. Z., Zhang, Y. S., Hong, G. F., and Xue, Y. B. An F-box gene linked to the self-incompatibility (S) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. *Plant Molecular Biology*, 50(1) :29–42, 2002. URL <GotoISI>://000176415000003.
- Langridge, P. and Baumann, U. Self-Incompatibility in the Grasses. In Heidelberg, S.-V. B., editor, *Self-Incompatibility in Flowering Plants*, pages 275–287. 2008.
- Lashermes, P., Combes, M. C., Prakash, N. S., Trouslot, P., Lorieux, M., and Charrier, A. Genetic linkage map of *Coffea canephora* : effect of segregation distortion and analysis of recombination rate in male and female meioses. *Genome*, 44(4) :589–596, 2001. URL <GotoISI>://000170177800011.

- Lawrence, M. J. Population genetics of the homomorphic self-incompatibility polymorphisms in flowering plants. *Ann Bot*, 85(suppl_1) :221–226, 2000. URL http://aob.oxfordjournals.org/cgi/content/abstract/85/suppl_1/221.
- Lawson, M. J. and Zhang, L. Q. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biology*, 7(2) :11, 2006. doi : R1410.1186/gb-2006-7-2-r14. URL <GotoISI>://000237298500011.
- Legrand, S., Hendriks, T., Hilbert, J. L., and Quillet, M. C. Characterization of expressed sequence tags obtained by SSH during somatic embryogenesis in *Cichorium intybus* L. *BMC Plant Biology*, 7 :27, 2007. doi : 10.1186/1471-2229-7-27. URL <GotoISI>://000247986900001.
- Legrand, S. *Etude transcripومية au cours des phases précoces de l'embryogénèse somatique chez la chicorée (Cichorium intybus)*. Phd thesis, Lille University, France, Lille, France, 2006.
- Lenormand, T. and Dutheil, J. Recombination difference between sexes : A role for haploid selection. *Plos Biology*, 3(3) :396–403, 2005. URL <GotoISI>://000227984000009.
- Levin, D. A. The evolutionary significance of pseudo-self-fertility. *American Naturalist*, 148 (2) :321–332, 1996.
- Liu, B.-H. *Statistical genomics : linkage, mapping, and QTL analysis*. CRC Press, New York, 1998.
- Llaurens, V., Gonthier, L., and Billiard, S. The sheltered genetic load linked to the S-locus in plants : new insights from theoretical and empirical approaches in sporophytic self-incompatibility. *Genetics*, 183(3) :1105–1118, 2009. doi : 10.1534/genetics.109.102707. URL <GotoISI>://000272295800028.
- Lorieux, M., Perrier, X., Goffinet, B., Lanaud, C., and Deleon, D. G. Maximum-likelihood models for mapping genetic-markers showing segregation distortion. 2. F2 populations. *Theoretical and Applied Genetics*, 90(1) :81–89, 1995. URL <GotoISI>://A1995QF11000011.
- Lorieux, M. Aspects statistiques de la cartographie des marqueurs moléculaires. Technical Report 1-94, CIRAD, janvier 1994.
- Lucchin, M., Varotto, S., Barcaccia, G., and Parrini, P. Chicory and Endive. In Prohens, J., Nuez, F., and Carena, M. J., editors, *Vegetables I*, volume 1 of *Handbook of Plant Breeding*, pages 3–48. Springer New York, 2008. URL http://dx.doi.org/10.1007/978-0-387-30443-4_1.
- Ma, X.-H., Qin, R.-L., and Xing, W.-B. Chromosome observation of some medical plants in Xinjiang. *Acta Phytotaxonomica Sinica*, 22(3) :243–249, 1984.

- Maliepaard, C., Alston, F. H., van Arkel, G., Brown, L. M., Chevreau, E., Dunemann, F., Evans, K. M., Gardiner, S., Guilford, P., van Heusden, A. W., Janse, J., Laurens, F., Lynn, J. R., Manganaris, A. G., den Nijs, A. P. M., Periam, N., Rikkerink, E., Roche, P., Ryder, C., Sansavini, S., Schmidt, H., Tartarini, S., Verhaegh, J. J., Vrielink-van Ginkel, M., and King, G. J. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theoretical and Applied Genetics*, 97(1-2) :60–73, 1998. URL <GotoISI>://000075377300008.
- McClure, B. A. and Franklin-Tong, V. Gametophytic self-incompatibility : understanding the cellular mechanisms involved in "self" pollen tube inhibition. *Planta*, 224(2) :233–245, 2006. URL <GotoISI>://000238523000001.
- McCubbin, A. Heteromorphic self-incompatibility in *Primula* : twenty-first century tools promise to unravel a classic nineteenth century model system. In Heidelberg, S.-V. B., editor, *Self-Incompatibility in Flowering Plants*, pages 289–308. 2008.
- McCubbin, A. G., Wang, X., and Kao, T. H. Identification of self-incompatibility (S-) locus linked pollen cDNA markers in *Petunia inflata*. *Genome*, 43(4) :619–627, 2000. URL <GotoISI>://000088273600005.
- McInnis, S. M., Costa, L. M., Gutierrez-Marcos, J. F., Henderson, C. A., and Hiscock, S. J. Isolation and characterization of a polymorphic stigma-specific class III peroxidase gene from *Senecio squalidus* L. (Asteraceae). *Plant Molecular Biology*, 57(5) :659–677, 2005. URL <GotoISI>://000231220400004.
- Meksem, K., Leister, D., Peleman, J., Zabeau, M., Salamini, F., and Gebhardt, C. A high-resolution map of the vicinity of the R1 locus on the chromosome V of potato based on RFLP and AFLP markers. *Molecular and General Genetics*, 249(1) :74–81, 1995. URL <GotoISI>://A1995TG13700010.
- Menkis, A., Jacobson, D. J., Gustafsson, T., and Johannesson, H. The mating-type chromosome in the filamentous ascomycete *Neurospora tetrasperma* represents a model for early evolution of sex chromosomes. *Plos Genetics*, 4(3) :10, 2008. doi : e100003010.1371/journal.pgen.1000030. URL <GotoISI>://000255407300017.
- Meudt, H. M. and Clarke, A. C. Almost Forgotten or Latest Practice? AFLP applications, analyses and advances. *Trends in Plant Science*, 12(3) :106–117, 2007. URL <http://www.sciencedirect.com/science/article/B6TD1-4N276GY-1/2/9c830eda37db800d82dbcfe775ddc8c2>.
- Michelmore, R. W., Paran, I., and Kesseli, R. V. Identification of markers linked to disease-resistance genes by bulked segregant analysis - a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America*, 88(21) :9828–9832, 1991. URL <GotoISI>://A1991GM78100095.

- Mun, J.-H., Kwon, S.-J., Yang, T.-J., Seol, Y.-J., Jin, M., Kim, J.-A., Lim, M.-H., Kim, J. S., Baek, S., Choi, B.-S., Yu, H.-J., Kim, D.-S., Kim, N., Lim, K.-B., Lee, S.-I., Hahn, J.-H., Lim, Y. P., Bancroft, I., and Park, B.-S. Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biology*, 10(10) :R111, 2009. URL <http://genomebiology.com/2009/10/10/R111>.
- Nasrallah, J. B. and Nasrallah, M. E. Pollen stigma signaling in the sporophytic self-incompatibility response. *Plant Cell*, 5(10) :1325–1335, 1993. URL <GotoISI>://A1993ME88400019.
- Nishio, T. and Hinata, K. Analysis of S-specific proteins in stigma of *Brassica oleracea* L by isoelectric-focusing. *Heredity*, 38(JUN) :391–396, 1977. URL <GotoISI>://A1977DH10900012.
- Noguera, F. J., Capel, J., Alvarez, J. I., and Lozano, R. Development and mapping of a codominant SCAR marker linked to the andromonoecious gene of melon. *Theoretical and Applied Genetics*, 110(4) :714–720, 2005. doi : 10.1007/s00122-004-1897-0. URL <GotoISI>://000227714400014.
- Olmstead, R. G., dePamphilis, C. W., Wolfe, A. D., Young, N. D., Elisons, W. J., and Reeves, P. A. Disintegration of the Scrophulariaceae. *American Journal of Botany*, 88(2) :348–361, 2001.
- Paris, M., Meyer, C.-L., Blassiau, C., Coissac, E., Taberlet, P., and Després, L. Two methods to easily obtain nucleotide sequences from AFLP loci of interest. In Pompanon, F. and Bonin, A., editors, *Data Production and Analysis in Population Genomics*, volume in press of *Methods in Molecular Biology*. Humana Press, New-Jersey, USA, 2011.
- Perry, J. and Ashworth, A. Evolutionary rate of a gene affected by chromosomal position. *Current Biology*, 9(17) :987–989, 1999. URL <GotoISI>://000082518200028.
- Pierre, M., Noel, L., Lahaye, T., Ballvora, A., Veuskens, J., Ganal, M., and Bonas, U. High-resolution genetic mapping of the pepper resistance locus Bs3 governing recognition of the *Xanthomonas campestris* pv *vesicatora* AvrBs3 protein. *Theoretical and Applied Genetics*, 101(1-2) :255–263, 2000. URL <GotoISI>://000088403800037.
- Quillet, M.-C., Blassiau, C., Mörchen, M., Habarugira, I., Huss, B., Gagneul, D., Gonthier, L., Rambaud, C., Cadalen, T., Heuvelmans, P., van de Wal, M., Prat, E., Hilbert, J.-L., Berges, H., and Hendriks, T. Cloning and characterization of nuclear male sterility 1 (*nms1*) in chicory (*Cichorium intybus* L., Asteraceae) . en préparation.
- Rahman, M. H., Tsuchiya, T., Suwabe, K., Kohori, J., Tomita, R. N., Kagaya, Y., Kobayashi, I., Kakeda, K., and Kowayama, Y. Physical size of the S locus region defined by genetic recombination and genome sequencing in *Ipomoea trifida*, Convolvulaceae. *Sexual Plant Reproduction*, 20(2) :63–72, 2007a. URL <GotoISI>://000246599800002.

- Rahman, M., Uchiyama, M., Kuno, M., Hirashima, N., Suwabe, K., Tsuchiya, T., Kagaya, Y., Kobayashi, I., Kakeda, K., and Kowyama, Y. Expression of stigma- and anther-specific genes located in the S locus region of *Ipomoea trifida*. *Sexual Plant Reproduction*, 20(2) :73–85, 2007b. URL <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-34249087767&partnerID=40&rel=R7.0.0>.
- Rambaud, C., Dubois, J., and Vasseur, J. The induction of tetraploidy in chicory (*Cichorium intybus* L. var Magdebourg) by protoplast fusion. *Euphytica*, 62(1) :63–67, 1992. URL <GotoISI>://A1992JX64800008.
- Reamon-Buttner, S. M., Schmidt, T., and Jung, C. AFLPs represent highly repetitive sequences in *Asparagus officinalis* L. *Chromosome Research*, 7(4) :297–304, 1999. URL <GotoISI>://000081063100008.
- Reyna-Lopez, G. E. and Ruiz-Herrera, J. Specificity of DNA methylation changes during fungal dimorphism and its relationship to polyamines. *Current Microbiology*, 48(2) :118–123, 2004. doi : 10.1007/s00284-003-4142-1. URL <GotoISI>://000188470600007.
- Richards, A. J. *Plant breeding systems*. Chapman et Hall, London, UK, second edition, 1997.
- Ride, J. P., Davies, E. M., Franklin, F. C. H., and Marshall, D. F. Analysis of *Arabidopsis* genome sequence reveals a large new gene family in plants. *Plant Molecular Biology*, 39(5) :927–932, 1999. URL <http://dx.doi.org/10.1023/A:1006178511787>.
- Schellenbaum, P., Mohler, V., Wenzel, G., and Walter, B. Variation in DNA methylation patterns of grapevine somaclones (*Vitis vinifera* L.). *BMC Plant Biology*, 8 :10, 2008. doi : 7810.1186/1471-2229-8-78. URL <GotoISI>://000258377800002.
- Schopfer, C. R., Nasrallah, M. E., and Nasrallah, J. B. The male determinant of self-incompatibility in *Brassica*. *Science*, 286(5445) :1697–1700, 1999. URL <GotoISI>://000083912200026.
- Sherman-Broyles, S. and Nasrallah, J. Self-Incompatibility and Evolution of Mating Systems in the Brassicaceae. In Heidelberg, S.-V. B., editor, *Self-Incompatibility in Flowering Plants*, pages 123–147. 2008.
- Silva, N. F. and Goring, D. R. Mechanisms of self-incompatibility in flowering plants. *Cellular and Molecular Life Sciences*, 58(14) :1988–2007, 2001. URL <GotoISI>://000173110400002.
- Stein, J. C., Howlett, B., Boyes, D. C., Nasrallah, M. E., and Nasrallah, J. B. Molecular-cloning of a putative receptor protein-kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *Proceedings of the National Academy of Sciences of the United States of America*, 88(19) :8816–8820, 1991. URL <GotoISI>://A1991GH80500110.
- Stone, J. Sheltered load associated with S-alleles in *Solanum carolinense*. *Heredity*, 92(4) :335–342, 2004. URL <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-1842855265&partnerID=40&rel=R5.6.0>.

- Stout, A. B. Fertility in *Cichorium intybus* : the sporadic occurrence of self-fertile plants among the progeny of self-sterile plants. *American Journal of Botany*, 4(7) :375–395.
- Suzuki, G., Tanaka, S., Yamamoto, M., Tomita, R. N., Koyama, Y., and Mukai, Y. Visualization of the S-locus region in *Ipomoea trifida* : toward positional cloning of self-incompatibility genes. *Chromosome Research*, 12(5) :475–481, 2004. URL <GotoISI>://000222609900006.
- Suzuki, G., Kai, N., Hirose, T., Fukui, K., Nishio, T., Takayama, S., Isogai, A., Watanabe, M., and Hinata, K. Genomic Organization of the S Locus : Identification and Characterization of Genes in SLG/SRK Region of S9 Haplotype of *Brassica campestris* (syn. rapa). *Genetics*, 153(1) :391–400, 1999. URL <http://www.genetics.org/cgi/content/abstract/153/1/391>.
- Tabah, D. A., McInnis, S. M., and Hiscock, S. J. Members of the S-receptor kinase multigene family in *Senecio squalidus* L. (Asteraceae), a species with sporophytic self-incompatibility. *Sexual Plant Reproduction*, 17(3) :131–140, 2004. URL <GotoISI>://000224529100004.
- Tagu, D. and Moussard, C. *Principes des techniques de biologie moléculaire*. INRA editions, 2nd edition, 2003.
- Takayama, S. and Isogai, A. Self-incompatibility in plants. *Annual Review of Plant Biology*, 56 :467–489, 2005. URL <GotoISI>://000230282800019.
- Terol, J., Naranjo, M., Ollitrault, P., and Talon, M. Development of genomic resources for *Citrus clementina* : Characterization of three deep-coverage BAC libraries and analysis of 46,000 BAC end sequences. *BMC Genomics*, 9, 2008. doi : 10.1186/1471-2164-9-423. URL <http://www.scopus.com/inward/record.url?eid=2-s2.0-53349173110&partnerID=40>.
- Thomas, S. G., Huang, S. J., Li, S. T., Staiger, C. J., and Franklin-Tong, V. E. Actin depolymerization is sufficient to induce programmed cell death in self-incompatible pollen. *Journal of Cell Biology*, 174(2) :221–229, 2006. URL <GotoISI>://000239306000006.
- Timms, L., Jimenez, R., Chase, M., Lavelle, D., McHale, L., Kozik, A., Lai, Z., Heesacker, A., Knapp, S., Rieseberg, L., Michelmore, R., and Kesseli, R. Analyses of synteny between *Arabidopsis thaliana* and species in the Asteraceae reveal a complex network of small syntenic segments and major chromosomal rearrangements. *Genetics*, 173(4) :2227–2235, 2006. doi : 10.1534/genetics.105.049205. URL <GotoISI>://000240620100033.
- Tomita, R. N., Fukami, K., Takayama, S., and Koyama, Y. Genetic mapping of AFLP/AMF-derived DNA markers in the vicinity of the self-incompatibility locus in *Ipomoea trifida*. *Sexual Plant Reproduction*, 16(6) :265–272, 2004a. doi : 10.1007/s00497-003-0194-4. URL <GotoISI>://000220231000002.
- Tomita, R. N., Suzuki, G., Yoshida, K., Yano, Y., Tsuchiya, T., Kakeda, K., Mukai, Y., and Koyama, Y. Molecular characterization of a 313-kb genomic region containing the self-incompatibility locus of *Ipomoea trifida*, a diploid relative of sweet potato. *Breeding Science*, 54(2) :165–175, 2004b. URL <GotoISI>://000224886600012.

- Ushijima, K., Sassa, H., Dandekar, A. M., Gradziel, T. M., Tao, R., and Hirano, H. Structural and transcriptional analysis of the self-incompatibility locus of almond : Identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. *Plant Cell*, 15(3) : 771–781, 2003. URL <GotoISI>://000185078000017.
- Uyenoyama, M. Evolution under tight linkage to mating type. *New Phytologist*, 165 :63–70, 2005.
- Van de Peer, Y., Fawcett, J. A., Proost, S., Sterck, L., and Vandepoele, K. The flowering world : a tale of duplications. *Trends in Plant Science*, 14(12) :680–688, 2009. doi : 10.1016/j.tplants.2009.09.001. URL <GotoISI>://000272855700006.
- Van Stallen, N., Vandenbussche, B., Verdoodt, V., and De Proft, M. Construction of a genetic linkage map for witloof (*Cichorium intybus* L. var. foliosum Hegi). *Plant Breeding*, 122(6) : 521–525, 2003. URL <GotoISI>://000187256600012.
- Varotto, S., Pizzoli, L., Lucchin, M., and Parrini, P. The incompatibility system in Italian red chicory (*Cichorium intybus* L). *Plant Breeding*, 114(6) :535–538, 1995. URL <GotoISI>://A1995TY68800013.
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Vandele, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. AFLP - a New Technique for DNA-Fingerprinting. *Nucleic Acids Research*, 23(21) :4407–4414, 1995. URL <GotoISI>://A1995TG42400026.
- Wang, Y., Wang, X., McCubbin, A. G., and Kao, T.-h. Genetic mapping and molecular characterization of the self-incompatibility (S) locus in *Petunia inflata*. *Plant Molecular Biology*, 53(4) :565–580, 2003. URL <http://dx.doi.org/10.1023/B:PLAN.0000019068.00034.09>.
- Watanabe, M., Suzuki, G., and Takayama, S. Milestones Identifying Self-Incompatibility Genes in *Brassica* Species : From Old Stories to New Findings. In Heidelberg, S.-V. B., editor, *Self-Incompatibility in Flowering Plants*, pages 151–172. 2008.
- Wei, P., Feng, H., Piao, Z. Y., Li, C. Y., Liu, Z. Y., Wang, Y. G., Ji, R. Q., Zou, T., and Ji, S. J. Identification of AFLP markers linked to Ms, a genic multiple allele inherited male-sterile gene in Chinese cabbage. *Breeding Science*, 59(4) :333–339, 2009. URL <GotoISI>://000274757300003.
- Wheeler, M. J., Armstrong, S. A., Franklin-Tong, V. E., and Franklin, F. C. H. Genomic organization of the *Papaver rhoeas* self-incompatibility S-1 locus. *Journal of Experimental Botany*, 54(380) :131–139, 2003. doi : 10.1093/jxb/erg006. URL <GotoISI>://000179865500015.
- Wheeler, M. J., Vatovec, S., and Franklin-Tong, V. E. The pollen S-determinant in *Papaver* : comparisons with known plant receptors and protein ligand partners. *Journal of Experimental Botany*, 61(7) :2015–2025, 2010. doi : 10.1093/jxb/erp383. URL <GotoISI>://000277238100010.

- Wheeler, M. J., de Graaf, B. H. J., Hadjiosif, N., Perry, R. M., Poulter, N. S., Osman, K., Vatovec, S., Harper, A., Franklin, F. C. H., and Franklin-Tong, V. E. Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. *Nature*, 459(7249) :992–995, 2009. URL <http://dx.doi.org/10.1038/nature08027>.
- Wicker, T., Narechania, A., Sabot, F., Stein, J., Vu, G., Graner, A., Ware, D., and Stein, N. Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. *BMC Genomics*, 9(1) :518, 2008. URL <http://www.biomedcentral.com/1471-2164/9/518>.
- Wills, D. M. and Burke, J. M. Quantitative Trait Locus Analysis of the Early Domestication of Sunflower. *Genetics*, 176(4) :2589–2599, 2007.
- Wright, S. I., Agrawal, N., and Bureau, T. E. Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Research*, 13(8) : 1897–1903, 2003. doi : 10.1101/gr.1281503. URL <GotoISI>://000184530900012.
- Xiao, Q. J., Wibowo, T. A., Wu, X. L., Michal, J. J., Reeves, J. J., Busboom, J. R., Thorgaard, G. H., and Jiang, Z. H. A simplified QTL mapping approach for screening and mapping of novel AFLP markers associated with beef marbling. *Journal of Biotechnology*, 127(2) : 177–187, 2007. URL <GotoISI>://000243657600001.
- Yan, H. H., Mudge, J., Kim, D. J., Shoemaker, R. C., Cook, D. R., and Young, N. D. Comparative physical mapping reveals features of microsynteny between *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana*. *Genome*, 47(1) :141–155, 2004. doi : 10.1139/g03-106. URL <GotoISI>://000220101100016.
- Yang, B., Thorogood, D., Armstead, I., and Barth, S. How far are we from unravelling self-incompatibility in grasses? *New Phytologist*, 178(4) :740–753, 2008. URL <GotoISI>://000255722300006.
- Yang, Q. Y., Zhang, D. F., Li, Q., Cheng, Z. K., and Xue, Y. B. Heterochromatic and genetic features are consistent with recombination suppression of the self-incompatibility locus in *Antirrhinum*. *Plant Journal*, 51(1) :140–151, 2007. URL <GotoISI>://000247478200011.
- Yoo, E. Y., Kim, S., Kim, Y. H., Lee, C. J., and Kim, B. D. Construction of a deep coverage BAC library from *Capsicum annuum*, 'CM334'. *TAG Theoretical and Applied Genetics*, 107 (3) :540–543, 2003. doi : 10.1007/s00122-003-1279-z. URL <http://dx.doi.org/10.1007/s00122-003-1279-z>.
- Yu, Q. Y., Tong, E., Skelton, R. L., Bowers, J. E., Jones, M. R., Murray, J. E., Hou, S. B., Guan, P. Z., Acob, R. A., Luo, M. C., Moore, P. H., Alam, M., Paterson, A. H., and Ming, R. A physical map of the papaya genome with integrated genetic map and genome sequence. *BMC Genomics*, 10 :12, 2009. doi : 37110.1186/1471-2164-10-371. URL <GotoISI>://000270393000001.
- Zhang, Y. and Xue, Y. Molecular biology of S-RNase-based self-incompatibility. In Heidelberg, S.-V. B., editor, *Self-Incompatibility in Flowering Plants*, pages 193–215. 2008.

- Zhang, Z. Z., Guo, M. L., and Zhang, J. D. Identification of AFLP fragments linked to hydroxysafflor yellow A in *Flos Carthami* and conversion to a SCAR marker for rapid selection. *Molecular Breeding*, 23(2) :229–237, 2009. doi : 10.1007/s11032-008-9228-9. URL <GotoISI>://000262125600005.
- Zhu, W., Ouyang, S., Iovene, M., O'Brien, K., Vuong, H., Jiang, J., and Buell, C. R. Analysis of 90 Mb of the potato genome reveals conservation of gene structures and order with tomato but divergence in repetitive sequence composition. *BMC Genomics*, 9(1) :286, 2008. URL <http://www.biomedcentral.com/1471-2164/9/286>.

Annexes

Annexe A

Matériels et méthodes

A.1 Culture *in vitro*

TABLEAU A.1 – Milieux d’induction de l’embryogénèse somatique (M17S20) et de culture *in vitro* (H5 et H5 solide).

	M17S20	H5	H5 solide
Solution de macroéléments 10 X	100 mL	100 mL	100 mL
Microéléments de Heller	1 mL	1 mL	1 mL
Glutamine	250 mg	-	-
Vitamines de Morel et Wetmore	10 mL	10 mL	10 mL
Inositol	100 mg	-	-
2ip	0.5 mg	-	-
ANA	0,02 mg	-	-
KH ₂ P ₀₄	85 mg	-	-
Solution de Fe-EDTA	5 mL	5 mL	5 mL
Saccharose	20 mg	5 g	5 g
Agar	-	-	6 g
H ₂ O	qsp 1 L	qsp 1 L	qsp 1 L

TABLEAU A.2 – Macroéléments

Solution stock macro éléments 10 X (mg/l)	M17S20	Heller
MgSO ₄ , 7H ₂ O	1850	2500
CaCl ₂ , 2H ₂ O	2200	750
NH ₄ N ₀₃	8250	
KCl	7500	7500
NaN ₀₃		6000
NaH ₂ P ₀₄ , 2H ₂ O		1413

TABLEAU A.3 – Microéléments

(a) Microéléments Heller (1953)		(b) Microéléments de Mura- shigue et Skoog (MS, 1962)	
Composés	mg/mL	Composés	mg/mL
ZnS04,7H20	1	H3B03	620
H3B04 1	1	CoCl2	2,5
MnS04	0,075	CUS04	2,5
CuS04, 5H20	0,03	MnS04	1690
AlCl3	0,03	KI	83
NiCL2, H20	0,01	ZnS04, 7H20	860
KCl	0,01	Na2Mo04, 2H20	25

TABLEAU A.4 – Vitamines de Morel et Wetmore (1951)

Composés	mg/mL
Panthoténate de Ca	1
Thiamine	1
Acide nicotinique	1
Pyridoxine	1
Biotine	0,5

TABLEAU A.5 – Solution de Fe-EDTA.

Composés	g/L
Na2EDTA	7,46
FeS04, 7H20	5,56

A.2 Amorces PCR

TABLEAU A.6 – Séquences et caractéristiques des amorces PCR utilisées dans le cadre de la thèse. Pour chaque amorce, la taille du produit d'amplification est indiquée (pb) ainsi que le T_m et le nombre de cycles PCR (CP).

Nom	pb	$T_m(^{\circ}C)$	CP
TACG1	293	60	35
TACG2	1500	60	35
TGCT	1027	60	35
Si1t7	462	60	35
Si1m13	363	58	35
Si2t7	220	58	35
Si2m13	530	60	35
Si3t7	481	60	35
Si3m13	373	60	35
Si7t7	176	60	35
Si9t7	151	60	35
Si9m13	340	60	35
Si9_intron	249	60	35
Si9_intron2	161	60	35
Si10m13	373	60	35
Si10t7	543	60	35
Eu06b03	315	60	35
Eu08sonde294	294	60	35
eu08b02sonde	704	60	35
TCAA	272	60	35
TCAA 2	113	57	35
nTTAA	201	60	35
nTTAA R2	339	60	35
nTTAA R3	378	57	30
AACC	109	60	35
Eu12_160	160	60	35
Eu12_380	380	60	35
Si16t7	112	60	35
Si16m13	224	60	35
Si17t7	192	60	35
Si17m13	300	60	35
Si20t7	382	60	35
Si20t7 2	302	60	35
Si23m13	317	60	35
Si25t7	474	60	35
Si25t7 R2	307	57	35
Si25m13	472	60	35

suite page suivante

Annexe A. Matériels et méthodes

Nom	pb	Tm(°C)	CP
Si25m13 L2	250	60	35
Si26t7	474	60	35
Si26t7 L2	383	60	35
Si26m13	206	60	35
Si27m13	385	60	35
Si27t7	342	60	35
Si28m13	125	60	35
Si28t7	319	60	35
Si29m13	302	60	35
Si29t7	393	60	35
Si30m13	384	60	35
Si30t7	253	60	35
Si31m13	313	60	35
Si31t7	305	60	35
Si32m13	304	60	35
Si33m13	237	60	35
Si33t7	338	60	35
Si34m13	335	60	35
Si35m13	279	58	35
Si35t7	600	60	30
Si36m13	293	60	35
Si36t7	271	60	35
Si37m13	304	60	35
Si38m13	324	60	35
Si38t7	289	58	35
Si39m13	375	60	35
Si39t7	265	59	35
Si40t7	267	60	35
Si40m13	284	60	35
Si41t7	251	60	35
Si41m13	265	60	35
Si42t7	271	60	35
Si42m13	276	60	35
Si43t7	293	60	35
Si43m13	290	60	35
Si44t7	275	60	35
Si44m13	272	60	35
Si47t7	272	60	35
Si47m13	265	60	35
Si48m13	277	60	35
Si49m13	300	60	35
Si50t7	391	60	35
Si50m13	253	60	35
Si51m13	290	60	35

suite page suivante

A.2. Amorce PCR

Nom	pb	T _m (°C)	CP
Si52t7	259	60	35
Si52m13	259	60	35
Si53m13	279	60	35
Si54t7	283	60	35
Si54m13	292	60	35
Si55t7	257	60	35
Si56t7	300	60	35
Si56m13	274	60	35
Si57t7	260	60	35
Si57m13	252	60	35

Annexe A. Matériels et méthodes

Annexe B

Carte physique

B.2 Annotation des gènes homologues d'*A. thaliana*

TABLEAU B.2 – Annotation des gènes homologues d'*A. thaliana* (cf. tableau 7.3 p.161).

Alias	Gène	Fonction
1At1	AT1G22870.1	Protein kinase family protein with ARM repeat domain
2At1	AT1G10720.1	BSD domain-containing protein
3At4	AT4G34660.1	SH3 domain-containing protein
4At1	AT1G10140.1	Uncharacterised conserved protein UCP031279
5At4	AT4G07410.1	Transducin family protein / WD-40 repeat family protein
6At1	AT1G76880.1	Duplicated homeodomain-like superfamily protein
7At1	AT1G21200.1	sequence-specific DNA binding transcription factors
8At1	AT1G30760.1	FAD-binding Berberine family protein
9At3	AT3G43700.1	BTB-POZ and MATH domain 6
10At2	AT2G34440.1	AGAMOUS-like 29
11At3	AT3G43630.1	Vacuolar iron transporter (VIT) family protein
12At1	AT1G76810.1	eukaryotic translation initiation factor 2 (eIF-2) family protein
13At5	AT5G58740.1	HSP20-like chaperones superfamily protein
14At1	AT1G76920.1	F-box family protein
15At1	AT1G43700.1	VIRE2-interacting protein 1
16At4	AT4G38470.1	ACT-like protein tyrosine kinase family protein
17At4	AT4G29090.1	Ribonuclease H-like superfamily protein
18At1	AT1G76950.1	Regulator of chromosome condensation (RCC1) family with FYVE zinc finger domain
19At5	AT5G42120.1	Concanavalin A-like lectin protein kinase family protein
20At5	AT5G50320.1	Radical SAM domain-containing protein / GCN5-related N-acetyltransferase (GNAT) family protein
21At3	AT3G29785.1	unknown protein
22At3	AT3G21470.1	Pentatricopeptide repeat (PPR-like) superfamily protein
23At3	AT3G09040.1	Pentatricopeptide repeat (PPR) superfamily protein

Annexe B. Carte physique

Annexe C

Construction and characterization of
two BAC libraries representing a
deep-coverage of the genome of
chicory
(*Cichorium intybus* L., Asteraceae)

SHORT REPORT

Open Access

Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (*Cichorium intybus* L., Asteraceae)

Lucy Gonthier^{1†}, Arnaud Bellec^{2†}, Christelle Blassiau¹, Elisa Prat², Nicolas Helmstetter², Caroline Rambaud¹, Brigitte Huss¹, Theo Hendriks¹, H el ene Berg es^{2*}, Marie-Christine Quillet^{1*}

Abstract

Background: The Asteraceae represents an important plant family with respect to the numbers of species present in the wild and used by man. Nonetheless, genomic resources for Asteraceae species are relatively underdeveloped, hampering within species genetic studies as well as comparative genomics studies at the family level. So far, six BAC libraries have been described for the main crops of the family, *i.e.* lettuce and sunflower. Here we present the characterization of BAC libraries of chicory (*Cichorium intybus* L.) constructed from two genotypes differing in traits related to sexual and vegetative reproduction. Resolving the molecular mechanisms underlying traits controlling the reproductive system of chicory is a key determinant for hybrid development, and more generally will provide new insights into these traits, which are poorly investigated so far at the molecular level in Asteraceae.

Findings: Two bacterial artificial chromosome (BAC) libraries, CinS2S2 and CinS1S4, were constructed from *HindIII*-digested high molecular weight DNA of the contrasting genotypes C15 and C30.01, respectively. C15 was hermaphrodite, non-embryogenic, and S_2S_2 for the *S*-locus implicated in self-incompatibility, whereas C30.01 was male sterile, embryogenic, and S_1S_4 . The CinS2S2 and CinS1S4 libraries contain 89,088 and 81,408 clones. Mean insert sizes of the CinS2S2 and CinS1S4 clones are 90 and 120 kb, respectively, and provide together a coverage of 12.3 haploid genome equivalents. Contamination with mitochondrial and chloroplast DNA sequences was evaluated with four mitochondrial and four chloroplast specific probes, and was estimated to be 0.024% and 1.00% for the CinS2S2 library, and 0.028% and 2.35% for the CinS1S4 library. Using two single copy genes putatively implicated in somatic embryogenesis, screening of both libraries resulted in detection of 12 and 13 positive clones for each gene, in accordance with expected numbers.

Conclusions: This indicated that both BAC libraries are valuable tools for molecular studies in chicory, one goal being the positional cloning of the *S*-locus in this Asteraceae species.

* Correspondence: helene.berges@toulouse.inra.fr; marie-christine.quillet@univ-lille1.fr

† Contributed equally

¹Univ Lille Nord de France, F-59000 Lille, France, Stress Abiotiques et Diff erenciati on des V eg etaux Cultiv es (SADV), UMR INRA-USTL 1281, B at. SN2, F-59655 Villeneuve d'Ascq, France

²Centre National des Ressources G enomiques V eg etales (CNRGV), INRA, Chemin de Borde Rouge, BP 52627, F-31326 Castanet Tolosan, France

Full list of author information is available at the end of the article

Introduction

Chicory (*Cichorium intybus* L.) is an Asteraceae (Compositae) species belonging to the Lactuceae tribe of the Lactuoideae subfamily [1,2], as are lettuce and other vegetable crops (e.g. endive and salsify), and widespread weeds (e.g. dandelion). The genus *Cichorium* includes 6 diploid species ($2n = 18$) native from the Old World [3]. The closely related species chicory (*C. intybus* L.) and endive (*C. endivia* L.) have been domesticated and are mainly cultivated in Europe. Wild chicory is perennial [3] but the crop has been selected to be cultivated as a biennial species. Four main cultivar groups are distinguished, depending on the use of their roots in transformed products (root or industrial cultivar group) or the consumption of leaves as fresh or cooked vegetables (witloof, pain de sucre, and radicchio cultivar groups) [4]. Though these crops are not highly ranked on the scale of world's agriculture, they are important at the regional level in Belgium, the North of France, and The Netherlands, where the production of root and witloof chicory is concentrated. Similarly, about 85% of the Italian radicchio production is performed in the north eastern part of that country [4].

Wild forms of *C. intybus* have been reported to be strictly allogamous [3], and diallel cross analyses performed in witloof and radicchio cultivar groups showed that selfing is avoided by a strong sporophytic self-incompatibility system controlled by a single multiallelic *S*-locus [5,6]. As for many allogamous species, breeder's work in the early 1960's showed heterosis effects in progenies obtained from crosses between relatively distant selected genotypes, indicating that F_1 hybrids would be the cultivar type to develop for chicory [7]. Variability for the level of self-fertility was found in different cultivar groups [6,8,9], and in the cultivar group witloof, partly self-compatible genotypes were selected to produce inbred lines used as F_1 hybrids progenitors [10,11]. Selection pressure for self-fertility was not strongly applied in other cultivar groups.

Resolving the molecular mechanisms underlying traits controlling the reproductive system of chicory is a key determinant for chicory hybrid development, and will more generally provide new insights into these traits, which are poorly understood so far at the molecular level in Asteraceae [12-14]. To achieve these goals, both molecular genetic tools and genomic resources are necessary, and a consensus genetic map based on 472 transferable codominant SSR and SNP markers covering the 9 chromosomes of the *C. intybus* haploid genome was recently established [15]. Furthermore, over 80,000 chicory and endive EST sequences are now publicly available on the NCBI dbEST site [16,17] and the CGPDB site (Compositae Genome Project Database, http://cgpdb.ucdavis.edu/asteraceae_assembly/).

BAC libraries are essential tools for detailed characterisation of genomic regions containing genes of interest by the construction of physical maps and positional cloning. Despite their importance in terms of numbers of species present in the wild, and used by man [2], Asteraceae genomic resources are relatively underdeveloped, hampering the development of within species genetic studies, and comparative genomics studies at the family and clade levels [e.g. [18-20]]. So far, six BAC libraries have been described for Asteraceae species, and they are restricted to the main crops of the family, i.e. lettuce [21] and sunflower [22-25]. Here we present the characterization of BAC libraries of *C. intybus* constructed from two genotypes differing in traits related to sexual reproduction, i.e. self-incompatibility and male sterility, and vegetative reproduction, i.e. somatic embryogenesis. To demonstrate the utility of these genomic libraries, we used two cDNA probes corresponding to single copy genes putatively involved in somatic embryogenesis to screen the libraries in order to have access to the gene sequences.

Methods

Plant materials

Two different *C. intybus* genotypes (C15 and C30.01), selected from a Hungarian landrace population (Koospol), were used for the construction of BAC libraries. C15 was obtained after selfing of the K59 genotype, and is hermaphrodite, non-embryogenic [16], and homozygous S_2S_2 for the *S*-locus. C30.01 was obtained from an open pollination of C30, descendant from K59 upon selfing. C30.01 is male sterile, embryogenic, and heterozygous S_1S_4 at the *S*-locus. *S*-locus genotypes were determined by test crosses; dominance relationships between the *S*-alleles being $S_1 = S_4 > S_2$ for the pistil, and $S_1 > S_4 > S_2$ for the pollen. Both genotypes were *in vitro* propagated, either by organogenesis (C15) [26], or by somatic embryogenesis (C30.01) [27]. Ten clones of each genotype were acclimated in a glasshouse at 20°C, with a 16/8 h light/dark cycle, and grown for about three months. After transferring the plants 5 to 7 days in a dark room, young leaves were collected, frozen in liquid nitrogen and stored at -80°C before extracting high molecular weight (HMW) genomic DNA.

BAC libraries construction

HMW DNA of C15 and C30.01 genotypes was prepared as described by Peterson et al. [28] with the following modifications: (1) PVP was lowered to 0.25% w/v in sucrose-based extraction buffer (0.01 M Tris, 0.1 M KCl, 0.01 M EDTA pH 9.4, 500 mM sucrose, 4 mM spermidine, 1 mM spermine tetrahydrochloride, 0.1% w/v ascorbic acid, 0.25% w/v PVP-40, and 0.13% w/v sodium

diethyldithiocarbamate) as described in Chalhoub et al. [29], (2) lysis buffer was 1% w/v sodium lauryl sarcosine, 0.3 mg/ml proteinase K and 0.13% w/v sodium diethyldithiocarbamate dissolved in 0.5 M EDTA pH 9.1, (3) after lysis of the nuclei, agarose plugs were pre-washed 1 h in 0.5 M EDTA pH 9.1 at 50°C, 1 h in 0.05 M EDTA pH 8.0 at 4°C, and then stored at 4°C [30]. Embedded HMW DNA was partially digested with *Hind*III (New England Biolabs, Ipswich, Massachusetts), subjected to two size selection steps by pulsed-field electrophoresis, using a BioRad CHEF Mapper system (BioRad Laboratories, Hercules, California), and ligated to pIndigoBAC-5 *Hind*III-Cloning Ready vector (Epicentre Biotechnologies, Madison, Wisconsin). Pulsed-field migration programs, electrophoresis buffer and ligation desalting conditions were performed for the construction of CinS2S2 library as described in Peterson et al. [28]. These steps of the procedure were modified for the construction of CinS1S4 library, according to Chalhoub et al. [29], in order to increase the insert size (see Results and discussion).

After electroporation of T1 resistant DH10B electrocompetent cells (Invitrogen, Carlsbad, California), transformants were incubated in SOC medium for one hour, and 100 µL of transformants were plated on LB agarose medium containing 12.5 µg/mL chloramphenicol, 80 mg/mL X-gal and 100 mg/mL IPTG for colony counting and insert size estimation. Glycerol stocks (6%, v/v) of the remaining transformants (2.9 mL) were stored at -80°C, for later plating and colony picking. BAC DNA from randomly selected plated colonies of the two libraries was isolated using NucleoSpin fast purification kit (Macherey-Nagel, Düren, Germany). DNA from each BAC was dissolved in 30 µL, and half of this amount was digested with 10 units of *Not*I (New England Biolabs) to excise inserts. The digested BAC DNA was separated in 1% agarose gel by pulsed-field electrophoresis at 12°C in 0.25 × TBE buffer at 6 V/cm with a linear ramp pulse time of 5-15 s for 16 h. Insert sizes were estimated with reference to a Lambda ladder (New England Biolabs).

Colony picking was carried out using a robotic workstation QPix2 XT (Genetix, New Milton, Hampshire, UK) using a white/blue selection. White colonies were arranged in 384-well microtiter plates containing LB medium with chloramphenicol (12.5 µg/mL) supplemented with 6% (v/v) glycerol.

BAC libraries screening

High-density colony filters were prepared from the complete CinS2S2 and CinS1S4 libraries using a robotic workstation QPix2 XT (Genetix). BAC clones were spotted using a 5 × 5 pattern onto 22 × 22 cm Immobilon-Ny+ filters (Millipore Corporate, Billerica,

Massachusetts). On each filter, 27,648 unique clones were spotted in duplicate, and clones were grown at 37°C for 17 h. Filters were then processed as follows: (1) denaturation on Whatman paper soaked with a solution of 0.5 M NaOH and 1.5 M NaCl for 4 min at room temperature, and for 10 min at 100°C, (2) neutralization on Whatman paper soaked with 1 M Tris-HCl pH 7.4, and 1.5 M NaCl for 10 min, incubation in a solution of 0.25 mg/mL proteinase K (Sigma Aldrich, St. Louis, Missouri) for 45 min at 37°C, baking for 45 min at 80°C, and (3) fixation by UV on a Biolink 254 nm crosslinker (Thermo Fischer Scientific, Waltham, Massachusetts) with an energy of 120,000 µJoules. Radiolabelling of probes and hybridization of the filters were performed as described in Sambrook and Russel [31]. Hybridized filters were imaged with a Storm 860 PhosphorImager (GE Healthcare, Little Chalfont, UK), and analyses were performed using the HDFR software (Incogen, Williamsburg, Virginia).

To evaluate the level of contamination by organelle DNA in the libraries, filters were hybridized with four mitochondrial and four chloroplast specific probes generated by PCR using total (*i.e.* nuclear, mitochondrial and chloroplast) DNA from *C. intybus* (genotype K59) as template, and primers described in Table 1. PCR products were purified with the NucleoSpin Extract II kit (Macherey-Nagel). Labelling of the probes with [³²P] dCTP was performed by random priming using the Ready-To-Go DNA Labelling Beads kit (GE Healthcare), and unincorporated nucleotides were then removed using Illustra ProbeQuant G-50 Micro Columns (GE Healthcare).

Some positive BAC clones detected with mitochondrial specific probes were subjected to a fingerprint analysis. BAC DNA was isolated as described in the previous section, and was digested with 100 units of *Eco*RI (New England Biolabs) in a final volume of 60 µL. Thirty five µL of restriction product were separated in 0.7% agarose gel in 1 × TAE buffer at 80 V for 22 h, and fragments were visualised under UV after ethidium bromide staining. Fragment sizes were estimated with reference to the Raoul™ molecular weight marker (MP Biomedicals, Santa Ana, California).

DNA probes corresponding to chicory genes *CiAGP* (arabinogalactan protein), and *CiSTM* (*SHOOT MERISTEMLESS*, a *KNOTTED1-LIKE* transcription factor) were used to ascertain the utility of the libraries to obtain clones containing single copy genes. The partial cDNA sequence of *CiAGP* was obtained from subtractive EST libraries, constructed from leaf explants of two chicory genotypes differing for their embryogenic capacity when cultivated under somatic embryogenesis-inducing conditions [16]. The full length *CiSTM* cDNA was obtained after amplification with degenerated primers,

Table 1 Mitochondrial, chloroplast, and nuclear genes derived probes, and associated primer sequences used for the characterization of the CinS2S2 and CinS1S4 libraries

Probe ^a	Description	Accession numbers ^b	Primer sequences (5' → 3')	Tm (°C)	Size (bp)
<i>atpA</i>	F1-F0 ATP synthase, subunit alpha	X80469, X51422, AF034118, X52838	L: gatcttgtaagcgactg R: agtaatgcctgagtcgagc	66 64	1000
<i>atp9</i>	F1-F0 ATP synthase, subunit 9	X51895, DQ539624	L: aataggggcccggagctgc R: gaaaggccatcattggggc	69 69	184
<i>cob</i>	Apocytochrome b	X98362, EF674047, EF674014, DQ916732	L: gagttatagcagtcctaggg R: ctagtagtaagcaatccgcc	55 58	697
<i>cox2</i>	Cytochrome c oxidase, subunit 2	AJ414385, EF547230, DQ004553, EF488904	L: attcaagacgcagcaacacc R: gtactacctcgtccattgag	65 56	264
<i>matK</i>	Maturase K	AJ633132	L: tggttcaggtcttcgctattgg R: cgtccctttgaagcaagaattg	70 68	398
<i>ndhF</i>	NADH dehydrogenase	AY504736	L: tacttgattgattctattctttg R: caacaagattaagattaataaaag	55 55	581
<i>rbcl</i>	Ribulose 1,5-biphosphate carboxylase/oxygenase, large subunit	L13652	L: ttgccgagataatggcctac R: ccaagatctcggctagagc	64 64	336
<i>trnL-trnF</i>	Intergenic spacer tRNA-Leu (trnL)-tRNA-Phe (trnF) genes	FJ490769	L: ggttcaagtcctctatcccc R: ctaccagctgagctatcccc	65 64	397
<i>CiAGP</i>	Arabinogalactan protein	DT212458	L: aaaccaaccaagactttgaccacg R: ccccttaagttccacaaattac	58 64	418
<i>CiSTM</i>	<i>SHOOT MERSISTEMLESS</i> transcription factor	GU189066	L: gtaggatcatcatgtttgatggggtttggag R: ccctaccttttgcaattca	73 64	344

^a*atpA*, *atp9*, *cob* and *cox2* correspond to mitochondrial sequences, *matK*, *ndhF*, *rbcl* and *trnL-trnF* to chloroplast sequences, and *CiAGP* and *CiSTM* to cDNA sequences.^b Multiple accession numbers for a probe indicate that primer pairs were designed from the consensus sequence of these accessions.

and RACE PCR of cDNA obtained from leaf explants of the embryogenic interspecific (*C. intybus* × *C. endivia*) genotype 474 [32] cultivated under somatic embryogenesis-inducing conditions (unpublished). *CiAGP* and *CiSTM* probes were generated by PCR from cDNA clones by using specific primers (Table 1). Positive clones detected by hybridization were validated individually by PCR amplification using the primer pairs used for probes synthesis, and visualisation of PCR products after agarose gel electrophoresis.

Results and discussion

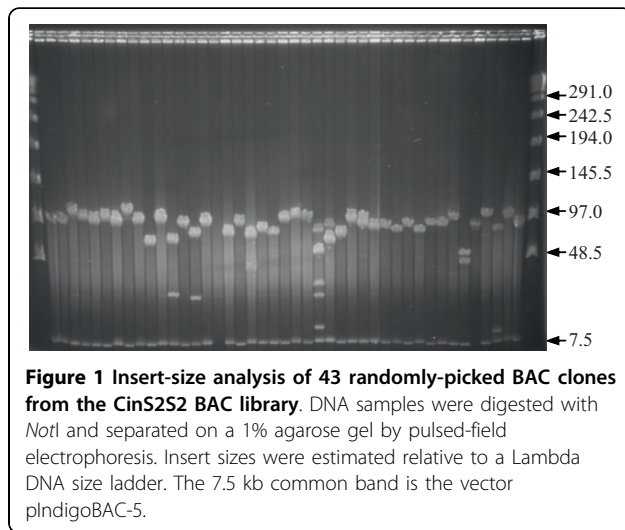
BAC libraries construction, and general features

Two BAC libraries were constructed in pIndigoBAC-5 with *Hind*III from DNA of two *C. intybus* genotypes differing in traits related to sexual and vegetative reproduction. Names of the libraries, CinS2S2 and CinS1S4, refer to the *S*-locus genotypes of the plants used to prepare HMW DNA, being S₂S₂, for C15, and S₁S₄ for C30.01. The main characteristics of the BAC libraries are summarized in Table 2. The CinS2S2 and CinS1S4 libraries consist of 89,088 and 81,408 clones, respectively, arrayed in a total of 444 384-well microtiter plates. A random selection of 178 clones for CinS2S2 and 303 clones for CinS1S4 was analysed after *Not*I digestion, followed by pulsed-field gel electrophoresis (Figure 1). From the clones sampled, the number of clones without insert was 1 (0.56%) and 8 (2.64%) for the CinS2S2 and CinS1S4 libraries, respectively. The average insert sizes of the CinS2S2 and

CinS1S4 libraries were calculated to be about 90 kb and 120 kb, respectively (Figure 2). More than 70% of clones from CinS2S2 and 90% for CinS1S4 libraries showed an insert size greater than 75 kb. The higher mean insert size obtained for CinS1S4 library could be explained by modifications of the first sizing step applied for the construction of CinS2S2 library. During the construction of this library, partial digested DNA was separated using the following parameters: 6 V/cm, 18 h, 120°, 1-40 s linear ramping in 0.5 × TBE. For the CinS1S4 library construction performed later on, two programs were aggregated: 6 V/cm, 18 h, 120°, 1-40 s linear ramping in 0.25 × TBE directly followed, in the same gel without removing or handling the DNA, by 5 V/cm, 8 h, 120°, 4-5 s ramping linear in the same buffer as described by Chalhoub et al. [29].

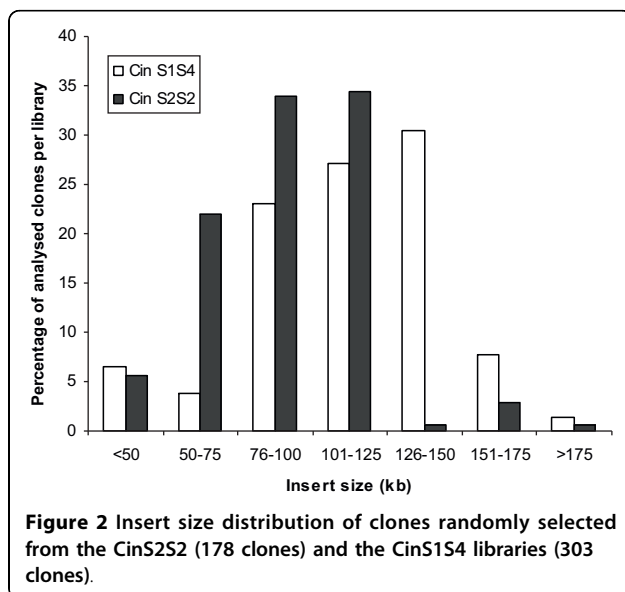
Table 2 Characteristics of CinS2S2 and CinS1S4 BAC libraries

Library	CinS2S2	CinS1S4	Total
No. of clones	89,088	81,408	170,496
No. of 384-well plates	232	212	444
Average insert size (kb)	91.78	118.20	104.40
Empty clones (%)	0.56	2.64	1.55
mtDNA clones (%)	0.02	0.03	0.03
cpDNA clones (%)	1.00	2.35	1.65
Nuclear haploid genome equivalents	5.75	6.53	12.28



Screening with mitochondrial and chloroplast genome probes

The mitochondrial (mt) genome of chicory has been estimated to be 300-400 kb [33], but genome structure and gene order are unknown. Similarly, in other Asteraceae species, mt genomes are poorly documented, and no complete sequences are publicly available. The probe mixture of four chicory mitochondrial genes (*atpA*, *atp9*, *cob*, and *cox2*) led to the detection of 43 and 29 positive clones in the CinS2S2 and CinS1S4 libraries, respectively. A set of 10 positive clones were randomly selected from each library, and subjected to PCR amplification with primers designed for the mt probes production. Five clones for library CinS2S2 and 8 clones for library CinS1S4 gave PCR products of the expected sizes for at least one of the targeted genes (Table 3). Three



clones gave a PCR product only with *cob* primers, and four clones with *atpA*, *atp9* and *cox2* primers. The remaining six clones could be amplified with *atpA* primers (1 clone), *atpA* and *cox2* primers (2 clones), or *atpA* and *atp9* primers (3 clones). To investigate whether the BAC clones containing the *cob* gene were possibly contiguous with those containing other mt genes, the 13 PCR validated BAC clones were subjected to *EcoRI* digestion for fingerprint analysis. Assuming that bands of the same size estimated after electrophoresis represent identical DNA fragments, pair-wise comparisons of fingerprints revealed that BAC clones of which the inserts contained the *cob* gene shared at least 2 fragments (Table 4). Similarly, at least 3 fragments of the same size were found present in fingerprints of BAC clones with inserts containing two or three other mt genes. The smallest insert (clone 09A06) containing the 3 genes *atpA*, *atp9*, and *cox2*, had an estimated size of 80 kb, delimiting the maximum interval containing these 3 genes (Table 4). However, the fingerprint of this clone did not show any fragment in common with fingerprints of *cob*-containing clones. In contrast, fingerprints of 3 other BAC clones with larger inserts containing the genes *atpA*, *atp9*, and *cox2* (92M03 and 31K22), or the genes *atpA* and *atp9* (96A09), shared 2 fragments (9.4 and 11.2 kb, respectively) with fingerprints of *cob*-containing clones (06I01, 26G13, and 25G18) (Table 4), suggesting that the inserts of these BAC clones partly overlapped. Considering all informative fingerprint fragments, the construction of a single contig encompassing the 4 mt genes was only possible when the BAC clone inserts are assumed to originate from a circular mt genome. As only 2 fragments were found in common between the *cob*-containing clones and the other clones, the *cob* gene may be located in a more distant part of the mt genome. Further experiments are necessary to determine the physical distances between these 4 genes. In sunflower, *atpA* and *cob* genes were reported to be only 10 kb apart, in a region associated with cytoplasmic male sterility [34,35]. However, because of frequent genome rearrangements, the organization of mitochondrial genes can vary between related species, and even within a single species [36].

Complete sequencing of chloroplast (cp) genomes in *Helianthus annuus* and *Lactuca sativa* indicated that genome size (151.1-152.7 kb), and gene content and order are the same in these two species, representing distantly related subfamilies in Asteraceae [37]. The four sequences (*matK*, *ndhF*, *rbcL* and *trnL-trnF*), chosen to evaluate cp contaminations, are distributed over a region of approximately 120 kb of the consensus cp genome of Asteraceae, the most distant genes (*ndhF* and *rbcL*) being separated by a distance of about 70 kb. The mixture of four probes led to the detection of 994 and 1919

Table 3 PCR amplifications of mitochondrial and chloroplast genome specific sequences on ten randomly selected clones of the CinS2S2 and CinS1S4 libraries detected after filter hybridization

	Mitochondrial genome specific sequences				Chloroplast genome specific sequences					
	Clone	<i>cox2</i>	<i>atpA</i>	<i>atp9</i>	<i>cob</i>	Clone	<i>matK</i> 2 kb*	<i>trnI-trnf</i> 47 kb*	<i>rbcL</i> 54 kb*	<i>ndhF</i> 124 kb*
CinS2S2 (90 kb)	06I01	-	-	-	+	02L02	+	-	-	-
	09A06	+	+	+	-	04M20	-	-	-	-
	09E24	-	-	-	-	06I22	+	-	-	+
	09I04	-	-	-	-	12E23	-	+	-	-
	13F07	-	-	-	-	26G24	+	+	+	-
	13M05	-	+	-	-	32H03	-	+	+	+
	26G13	-	-	-	+	37C04	+	+	-	-
	27J01	-	-	-	-	41L23	-	+	+	+
	41H20	+	+	+	-	51H20	-	-	+	-
	54B09	-	-	-	-	61O16	-	+	+	-
CinS1S4 (120 kb)	01H06	+	+	-	-	02N01	-	+	+	+
	16P11	+	+	-	-	21F01	+	-	-	+
	22B14	-	-	-	-	45D02	+	+	+	+
	25G18	-	-	-	+	51D03	-	-	+	+
	27O16	-	+	+	-	55M03	+	+	+	+
	31K22	+	+	+	-	62G03	+	+	+	+
	91G03	-	+	+	-	72G02	-	+	+	+
	92M03	+	+	+	-	84D02	+	-	-	+
	96A09	-	+	+	-	91N03	+	+	+	+
	108A12	-	-	-	-	106P01	+	+	+	+

+ indicates the presence of a PCR product of the expected size after agarose gel electrophoresis;
 - indicates no amplification; * Sequence coordinates on *H. annuus* and *L. sativa* chloroplast genome according to [37].

positive clones in the CinS2S2 and CinS1S4 libraries, respectively. PCR validation with specific primers performed on 10 positive clones of each library gave PCR products of expected size for at least one primer pair for 9 clones of the CinS2S2 library, and the 10 clones selected from CinS1S4 library (Table 3). A range of 1-3 (CinS2S2), and 2-4 (CinS1S4) PCR products per clone were detected, and the PCR data confirm the gene order described for the cp genome of *H. annuus* and *L. sativa*

[37]. In addition, five of the 10 clones sampled in the CinS1S4 library produced amplification products with the four primer pairs, suggesting that a large continuous fragment of the cp genome is present in these clones.

Based on the proportion of clones validated by PCR, we estimated that 0.024% and 0.028% of the BAC clones would result from mtDNA contamination, and that 1.00% and 2.35% of the clones contain putative cpDNA in CinS2S2 and CinS1S4 libraries, respectively. These

Table 4 Fingerprint analysis of BAC clones containing mt genes

Library	Clone	mt gene present	Insert size (kb)	Informative fingerprint fragment (kb)						
				4.8 4.9	11.2	5.6	11.7 9.0 8.8	7.0	9.4	3.6 4.4
CinS2S2	09A06	<i>cox2, atpA, atp9</i>	80	-	-	+	+	-	-	-
CinS1S4	96A09	<i>atpA, atp9</i>	150	-	-	-	+	+	+	-
CinS1S4	92M03	<i>cox2, atpA, atp9</i>	150	-	-	+	+	+	+	-
CinS1S4	31K22	<i>cox2, atpA, atp9</i>	95	-	+	+	+	+	-	-
CinS2S2	06I01	<i>cob</i>	55	+	+	-	-	-	+	+
CinS2S2	26G13	<i>cob</i>	60	+	+	-	-	-	-	+
CinS1S4	25G18	<i>cob</i>	25	-	-	-	-	-	+	+

DNA of BAC clones was digested with *NotI* for insert size determination, and with *EcoRI* for fingerprint analysis. Fingerprint fragments ranging from 3 to 12 kb were analysed, and fragments were considered as informative when present in at least 2 clones. Only the results for clones containing at least *atpA* and *atp9*, and sharing a fragment with *cob*-containing clones are shown. Clone 09A06 was included as it contained the smallest insert with the 3 genes *atpA*, *atp9*, and *cox2*. + indicates presence of a fingerprint fragment; - indicates absence of a fingerprint fragment.

results are consistent with: (1) the generally low frequency of mtDNA contamination (0-0.3%) reported in recent plant BAC libraries characterizations [38,39], and (2) the range of commonly observed frequencies (0.3-2.5%) attributed to cpDNA contamination [40,41].

The detected BAC clones containing mtDNA and cpDNA sequences could represent valuable tools as a starting point to investigate the structure of organelle genomes in chicory, particularly the mt genome involved in cytoplasmic male sterility [33,42]. However, the possibility that some of these organelle DNA contaminants actually are the result of horizontal transfer of DNA from mitochondria and/or plastid to the nucleus, as has been observed in most of the plant species studied [43-45], cannot be excluded. Whether large organelle DNA fragments exist in Asteraceae nuclear genomes remains to be determined; chromosome fluorescent *in situ* hybridization experiments [46] using chicory mtDNA- and cpDNA-containing BAC clones would be an approach to investigate this question.

Theoretical and empirical genome coverage

C. intybus has a haploid genome size estimated by flow cytometry to be between 1300 Mb (P. Devaux, personal communication) and 1400 Mb [47]. After subtraction of non-recombinant and putative organelle DNA contaminant clones, considering a haploid genome size of 1400 Mb, and mean insert sizes of 90 kb for CinS2S2 and 120 kb for CinS1S4, the two libraries represent approximately 5.7 and 6.5 haploid genome equivalents, respectively. Together these two libraries provide a coverage of the *C. intybus* genome of about 12.3-fold (Table 2), resulting in a theoretical probability superior to 0.9999 of recovering any targeted genomic sequence [48]. In comparison, the lettuce BAC library described by Frijters et al. [21] covers about 2 genome equivalents, and the 5 combined sunflower BAC libraries published represent about 20 genome equivalents [22-25]. This indicates that the chicory libraries represent a significant genomic resource to investigate Asteraceae genomes.

To validate empirically the genome coverage calculated for both libraries, BAC clones were screened with two cDNA probes (*CiAGP* and *CiSTM*). Mapping data for each of the corresponding genes [49] revealed a unique locus position in the chicory map of Cadalen et al. [15], indicating that they represent single copy sequences. Hybridization allowed the identification of 13 positive clones for *CiAGP*: 2 clones in the CinS2S2 library and 11 clones in the CinS1S4 library. The *CiSTM* probe revealed 5 positive clones in the CinS2S2 library and 7 positive clones in the CinS1S4 library, *i.e.* a total of 12 clones in the combined libraries. The clones detected by hybridization were subjected to PCR amplification with the primer pairs used for the

synthesis of the hybridization probes, and agarose gel electrophoresis of PCR products showed amplicons of the expected size for all of them. As for the *CiAGP* gene, an under- and over-representation was found in the CinS2S2 and CinS1S4 libraries, respectively. Further screening with a larger set of single copy sequence probes are necessary to determine if significant differences between libraries exist as for their content in single copy sequences. Nonetheless, experimental results are in accordance with the theoretical genome coverage estimated for the combined libraries, indicating that they constitute an efficient tool to recover single copy genes or genomic sequences for chicory.

Perspective

One future use of the *C. intybus* BAC libraries will be the positional cloning of the *S*-locus in this species. As in some other plant families, Asteraceae species exhibit a sporophytic self-incompatibility (SSI) system where the pollen phenotype depends on the genotype of the plant that produced the pollen [50]. The pollen and stigma determinants of SSI have been characterised in *Brassica* [51,52], and molecular and genetic studies showed that these genes are different from those involved in the SSI response in Convolvulaceae and Asteraceae, suggesting that different molecular mechanisms of SSI have evolved independently [12,53]. More recently, candidate genes with unknown functions have been identified in *Ipomoeae trifida* (Convolvulaceae) through a positional cloning approach [54]. Similarly, we have chosen for genetic and subsequent physical mapping of the *S*-locus in chicory to identify pollen and stigma determinants of SSI in this Asteraceae species.

Library availability

The libraries described are accessible for scientific research based on collaboration with the laboratories represented by the corresponding authors.

Acknowledgements

The authors thank Bruno Desprez (Florimond Desprez, Cappelle-en-Pévèle, France) for providing seeds of the Koospol population. Part of this work was financed by a 'Bonus Qualité Recherche' of USTL to the UMR 1281. Lucy Gonthier was supported by a doctoral fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche, France. We also like to thank the three anonymous reviewers for their valuable comments on a previous version of the paper.

Author details

¹Univ Lille Nord de France, F-59000 Lille, France, Stress Abiotiques et Différenciation des Végétaux Cultivés (SADV), UMR INRA-USTL 1281, Bât. SN2, F-59655 Villeneuve d'Ascq, France. ²Centre National des Ressources Génétiques Végétales (CNRGV), INRA, Chemin de Borde Rouge, BP 52627, F-31326 Castanet Tolosan, France.

Authors' contributions

The project was conceived and supervised by MCQ and TH. LG, AB, CB, and NH realized the construction of the two BAC libraries under the supervision

of HB in the CNRGV - INRA, Toulouse. EP performed the hybridizations using probes provided by CR and BH, and NH realized the fingerprint experiment. All authors contributed to the analyses and interpretations of the results, took part in the redaction of the manuscript, and approve its final version.

Competing interests

The authors declare that they have no competing interests.

Received: 8 June 2010 Accepted: 11 August 2010

Published: 11 August 2010

References

- Bremer K: *Asteraceae: Cladistics and classification*. Portland, Oregon: Timber Press 1994.
- Funk VA, Bayer RJ, Keeley S, Chan R, Watson L, Gemeinholzer B, Schilling E, Panero JL, Baldwin BG, Garcia-Jacas N, et al: **Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae**. *Biologiske Skrifter* 2005, **55**:343-374.
- Kiers AM, Mes THM, Van der Meijden R, Bachmann K: **Morphologically defined *Cichorium* (Asteraceae) species reflect lineages based on chloroplast and nuclear (ITS) DNA data**. *Syst Bot* 1999, **24**(4):645-659.
- Lucchin M, Varotto S, Barcaccia G, Parrini P: **Chicory and endive**. In *Vegetables I*. Edited by: Prohen J, Nuez F. New-York: Springer; 2008:3-48.
- Eenink AH: **Compatibility and incompatibility in witloof-chicory (*Cichorium intybus* L.)**. 2. The incompatibility system. *Euphytica* 1981, **30**(1):77-85.
- Varotto S, Pizzoli L, Lucchin M, Parrini P: **The incompatibility system in Italian red chicory (*Cichorium intybus* L.)**. *Plant Breed* 1995, **114**(6):535-538.
- Bannerot H, Deconinck B: **Travaux d'amélioration de la chicorée witloof à l'INRA**. *Annales de Gembloux* 1965, **71**:143-158.
- Coppens d'Eeckenbrugge G, Evrard B, Louant B-P: **Aspects quantitatifs de l'auto-incompatibilité chez la chicorée de witloof (*Cichorium intybus* L.)**. Relation entre fertilité et autofertilité. *Revue de l'Agriculture* 1987, **40**(1):35-40.
- Castaño CI, Demeulemeester MAC, De Proft MP: **Incompatibility reactions and genotypic identity status of five commercial chicory (*Cichorium intybus* L.) hybrids**. *Sci Hortic* 1997, **72**(1):1-9.
- Bannerot H, Deconinck B: **L'utilisation des hybrides F1: nouvelle méthode d'amélioration de la chicorée de bruxelles**. *Symposium interne section horticole d'Eucarpia* Gembloux, Belgium 1970, 99-118.
- Doré C, Varoquaux F: **La chicorée endive ou witloof**. *Histoire et amélioration de cinquante plantes cultivées* Quae 2006, 207-217.
- Hiscock SJ, McInnis SM, Tabah DA, Henderson CA, Brennan AC: **Sporophytic self-incompatibility in *Senecio squalidus* L. (Asteraceae) - the search for S**. *J Exp Bot* 2003, **54**(380):169-174.
- McInnis SM, Costa LM, Gutierrez-Marcos JF, Henderson CA, Hiscock SJ: **Isolation and characterization of a polymorphic stigma-specific class III peroxidase gene from *Senecio squalidus* L. (Asteraceae)**. *Plant Mol Biol* 2005, **57**(5):659-677.
- Tabah DA, McInnis SM, Hiscock SJ: **Members of the S-receptor kinase multigene family in *Senecio squalidus* L. (Asteraceae), a species with sporophytic self-incompatibility**. *Sex Plant Reprod* 2004, **17**(3):131-140.
- Cadalen T, Mörchen M, Blassiau C, Clabaut A, Scheer I, Hilbert J-L, Hendriks T, Quillet M-C: **Development of SSR markers and construction of a consensus genetic map for chicory (*Cichorium intybus* L.)**. *Mol Breed* 2010, **25**(4):699-722.
- Légrand S, Hendriks T, Hilbert JL, Quillet MC: **Characterization of expressed sequence tags obtained by SSH during somatic embryogenesis in *Cichorium intybus* L.** *BMC Plant Biol* 2007, **7**:27.
- Dauchot N, Mingeot D, Purnelle B, Muys C, Watillon B, Boutry M, Van Cutsem P: **Construction of 12 EST libraries and characterization of a 12,226 EST dataset for chicory (*Cichorium intybus*) root, leaves and nodules in the context of carbohydrate metabolism investigation**. *BMC Plant Biol* 2009, **9**:14.
- Feng Q, Huang T, Zhao Q, Zhu JJ, Lin ZX, Han B: **Analysis of collinear regions of *Oryza* AA and CC genomes**. *J Genet Genomics* 2009, **36**(11):667-677.
- Yan HH, Mudge J, Kim DJ, Shoemaker RC, Cook DR, Young ND: **Comparative physical mapping reveals features of microsynteny between *Glycine max*, *Medicago truncatula*, and *Arabidopsis thaliana***. *Genome* 2004, **47**(1):141-155.
- Cenci A, Combes MC, Lashermes P: **Comparative sequence analyses indicate that *Coffea* (Asterids) and *Vitis* (Rosids) derive from the same paleo-hexaploid ancestral genome**. *Mol Gen Genet* 2010, **283**(5):493-501.
- Frijters ACJ, Zhang Z, vanDamme M, Wang GL, Ronald PC, Michelmore RW: **Construction of a bacterial artificial chromosome library containing large *EcoRI* and *HindIII* genomic fragments of lettuce**. *Theor Appl Genet* 1997, **94**(3-4):390-399.
- Gentzbittel L, Abbott A, Galaud JP, Georgi L, Fabre F, Liboz T, Alibert G: **A bacterial artificial chromosome (BAC) library for sunflower, and identification of clones containing genes for putative transmembrane receptors**. *Mol Gen Genet* 2002, **266**(6):979-987.
- Ozdemir NR, Horn R, Friedt W: **Construction and characterization of a BAC library for sunflower (*Helianthus annuus* L.)**. *Euphytica* 2004, **138**(2):177-183.
- Bouzidi MF, Franchel J, Tao QZ, Stormo K, Mraz A, Nicolas P, Mouzeyar S: **A sunflower BAC library suitable for PCR screening and physical mapping of targeted genomic regions**. *Theor Appl Genet* 2006, **113**(1):81-89.
- Feng JH, Vick BA, Lee MK, Zhang HB, Jan CC: **Construction of BAC and BIBAC libraries from sunflower and identification of linkage group-specific clones by overgo hybridization**. *Theor Appl Genet* 2006, **113**(1):23-32.
- Decout E, Dubois T, Guedira M, Dubois J, Audran JC, Vasseur J: **Role of temperature as a triggering signal for organogenesis or somatic embryogenesis in wounded leaves of chicory cultured *in vitro***. *J Exp Bot* 1994, **45**(281):1859-1865.
- Bellettre A, Couillerot JP, Blervacq AS, Aubert S, Gout E, Hilbert JL, Vasseur J: **Glycerol effects both carbohydrate metabolism and cytoskeletal rearrangements during the induction of somatic embryogenesis in chicory leaf tissues**. *Plant Physiol Biochem* 2001, **39**(6):503-511.
- Peterson DG, Tomkins JP, Frisch DA, Wing RA, Paterson AH: **Construction of plant bacterial artificial chromosome (BAC) libraries: An illustrated guide**. *Journal of Agricultural Genomics* 2000, **5**.
- Chalhoub B, Belcram H, Caboche M: **Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size**. *Plant Biotechnol J* 2004, **2**(3):181-188.
- Construction and manipulation of large-insert bacterial clone libraries-manual. [<http://hbz7.tamu.edu/homelinks/tool/protocol.htm>].
- Sambrook J, Russel DW: **Hybridization of bacterial DNA on filters**. *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor Laboratory Press NY, Third 2001, 138-142.
- Dubois T, Guedira M, Dubois J, Vasseur J: **Direct somatic embryogenesis in leaves of *Cichorium* - a histological and SEM study of early stages**. *Protoplasma* 1991, **162**(2-3):120-127.
- Rambaud C, Dubois J, Vasseur J: **Male-sterile chicory cybrids obtained by intergeneric protoplast fusion**. *Theor Appl Genet* 1993, **87**(3):347-352.
- Köhler RH, Horn R, Lossl A, Zetsche K: **Cytoplasmic male-sterility in sunflower is correlated with the co-transcription of a new open reading frame with the ATPa gene**. *Mol Gen Genet* 1991, **227**(3):369-376.
- Laver HK, Reynolds SJ, Moneger F, Leaver CJ: **Mitochondrial genome organization and expression associated with cytoplasmic male-sterility in sunflower (*Helianthus annuus*)**. *Plant J* 1991, **1**(2):185-193.
- Kubo T, Newton KJ: **Angiosperm mitochondrial genomes and mutations**. *Mitochondrion* 2008, **8**(1):5-14.
- Timme RE, Kuehl JV, Boore JL, Jansen RK: **A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats**. *Am J Bot* 2007, **94**(3):302-312.
- Gutman W, Pawelkowicz M, Woycicki R, Piszczek E, Przybecki Z: **The construction and characteristics of a BAC library for *Cucumis sativus* L. 'B10'**. *Cell Mol Biol Lett* 2008, **13**(1):74-91.
- Baig MNR, Yu A, Guo WW, Deng XX: **Construction and characterization of two *Citrus* BAC libraries and identification of clones containing the phytoene synthase gene**. *Genome* 2009, **52**(5):484-489.
- Ruffel S, Caranta C, Palloix A, Lefebvre V, Caboche M, Bendahmane A: **Structural analysis of the eukaryotic initiation factor 4E gene controlling potyvirus resistance in pepper: exploitation of a BAC library**. *Gene* 2004, **338**(2):209-216.
- Yasui Y, Mori M, Matsumoto D, Ohnishi O, Campbell CG, Ota T: **Construction of a BAC library for buckwheat genome research - an application to positional cloning of agriculturally valuable traits**. *Genes Genet Syst* 2008, **83**(5):393-401.

42. Dubreucq A, Berthe B, Asset JF, Boulidard L, Budar F, Vasseur J, Rambaud C: **Analyses of mitochondrial DNA structure and expression in three cytoplasmic male-sterile chicories originating from somatic hybridisation between fertile chicory and CMS sunflower protoplasts.** *Theor Appl Genet* 1999, **99**(7-8):1094-1105.
43. Richly E, Leister D: NUMTs in sequenced eukaryotic genomes. *Mol Biol Evol* 2004, **21**(6):1081-1084.
44. Richly E, Leister D: NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Mol Biol Evol* 2004, **21**(10):1972-1980.
45. Leister D: Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet* 2005, **21**(12):655-663.
46. Stupar RM, Lilly JW, Town CD, Cheng Z, Kaul S, Buell CR, Jiang JM: **Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats.** *Proc Natl Acad Sci USA* 2001, **98**(9):5099-5103.
47. Koopman WJM: **Zooming in on the lettuce genome: species relationships in *Lactuca* s.l., inferred from chromosomal and molecular characters.** *PhD Thesis Wageningen University, Nederlands* 2002.
48. Clarke L, Carbon J: **A colony bank containing synthetic Col EI hybrid plasmids representative of the entire *E. coli* genome.** *Cell* 1976, **9**(1):91-99.
49. Clabaut A: **Analyse génétique de l'embryogenèse somatique chez la chicorée (*Cichorium intybus* L.): cartographie des QTL et gènes candidats.** *PhD Thesis Lille, France: University of Lille* 2009, **1**.
50. de Nettancourt D: **Incompatibility in angiosperms.** New York, USA: Springer-Verlag 1977.
51. Stein JC, Howlett B, Boyes DC, Nasrallah ME, Nasrallah JB: **Molecular cloning of a putative receptor protein-kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*.** *Proc Natl Acad Sci USA* 1991, **88**(19):8816-8820.
52. Schopfer CR, Nasrallah ME, Nasrallah JB: **The male determinant of self-incompatibility in *Brassica*.** *Science* 1999, **286**(5445):1697-1700.
53. Kowiyama Y, Tsuchiya T, Kakeda K: **Sporophytic self-incompatibility in *Ipomoea trifida*, a close relative of sweet potato.** *Ann Bot* 2000, **85**:191-196.
54. Rahman MH, Uchiyama M, Kuno M, Hirashima N, Suwabe K, Tsuchiya T, Kagaya Y, Kobayashi I, Kakeda K, Kowiyama Y: **Expression of stigma- and anther-specific genes located in the S locus region of *Ipomoea trifida*.** *Sex Plant Reprod* 2007, **20**(2):73-85.

doi:10.1186/1756-0500-3-225

Cite this article as: Gonthier et al.: Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (*Cichorium intybus* L., Asteraceae). *BMC Research Notes* 2010 **3**:225.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



Annexe D

Synthesis and translations

D.1 Introduction

1.4 Differences and similarities of the homomorphic SI systems

Different genetic determinisms are responsible for different SI mechanisms, for example :

- ligand-receptor system with ligand secreted on the stigma surface (Papaveraceae),
- ligand-receptor system with ligand secreted by tapetum cells of the anther (Brassicaceae),
- system with ribonuclease activity and F-box proteins (Solanaceae, Rosaceae, Scrophulariaceae).

However, there are also common characteristics between these mechanisms :

- high haplotypic diversity,
- low amino-acid sequence identity between determinants of the different haplotypes
- location of the determinants in a region highly divergent between haplotypes (sequence and structure diversity),
- location in region of the genome with a reduced recombination frequency,
- a preferential expression of the male and female determinant in the anther and the pistil, respectively.

For SI species belonging to a family for which the mechanism has not yet been identified, we expect to find determinants in a genomic region corresponding to these characteristics. We also notice that the different determinants of mechanisms can : 1) belong to gene families coding proteins present in different species, either with known functions (for example SRK or S-RNase) or unknown functions (for example PrpS) or, 2) have no structural equivalent in other species (for example PrpS).

Several approaches (biochemical, candidate gene, or positional cloning) have been used so far to identify the determinants of self-incompatibility. From a historic point of view, the biochemical approach was the first one to have been used. For the species for which we have no genetic information, it is effectively a possible strategy, but it presupposes that the protein is abundant in the pistil or at the level of the pollen, which is not always the case. Another option is the gene candidate approach. However, in view of the variety of already observed mechanisms, this will not automatically provide the desired result. Finally, if the necessary tools (genetic maps, progenies, genomic resources) are available, an approach by positional

cloning can be envisaged.

2 Asteraceae sporophytic self-incompatibility : chicory as a model ?

The Asteraceae family is the most important of the Angiosperms (Funk *et al.*, 2009), and more than 60% of the species belonging to this family may be self-incompatible (Ferrer et Good-Avila, 2007). However, there is a lack of documentation at the molecular level and so far none of the SI mechanisms involved have been identified.

Chicory belongs to the Lactuceae tribe, and the *Cichorium* genus (2n = 18) comprises 6 different species : *C. intybus*, *C. endivia*, *C. pumilum*, *C. calvum*, *C. spinosum*, and *C. bottae* (Kiers *et al.*, 2000).

According to Kiers (2000), 3 of these species are perennial and self-incompatible (*C. intybus*, *C. spinosum* and *C. bottae*), the others being annuals and autogamous. The reproduction systems of chicory (preferentially allogamous) and *C. endivia* (autogamous), which are the only cultivated species of the genus, has been described several times (Castaño *et al.*, 1997 ; Coppens d'Eeckenbrugge, 1990 ; Coppens d'Eeckenbrugge et Evrard, 1988 ; Eenink, 1981a,b, 1982, 1984 ; Kiers, 2000 ; Lucchin *et al.*, 2008 ; Varotto *et al.*, 1995).

Chicory is cultivated as a biennial. Spanning from 4 to 6 weeks, flowering occurs the second year, after vernalization, and is synchronized for all the flowers in the same capitulum. Each plant produces several hundred capitula. These characteristics allow 1) to estimate precisely the seed-set per capitulum, and 2) to realise a large number of repetitions of crosses on each individual.

Germination of pollen can be observed in a few minutes after pollination when the cross is compatible. After incompatible crosses, rejection of pollen takes place on the stigma surface : there is no germination, or, if it happens, growth of the pollen tube is quickly arrested (Coppens d'Eeckenbrugge, 1990). Some studies have shown that selfing is avoided by a sporophytic self-incompatibility system (Eenink, 1981a ; Varotto *et al.*, 1995). However, in the different varieties of cultivated chicory, the levels of compatibility may range from self-incompatible to self-fertile (Castaño *et al.*, 1997 ; Coppens d'Eeckenbrugge *et al.*, 1987). As a cultivated species, some studies have produced genetic resources of chicory :

- 2 genetic maps which had been built with dominant molecular markers (De Simone *et al.*, 1997 ; Van Stallen *et al.*, 2003)
- 1 genetic map which had been built with codominant molecular markers (Cadalen *et al.*, 2010)
- 3 different EST libraries which are composed to more than 50 000 EST sequences (Dauchot *et al.*, 2009 ; Legrand *et al.*, 2007, and Compositae Genome Project Database),

Recently, 2 new EST libraries of chicory were obtained. The first was constructed in our laboratory from stigma RNA, and the second from roots RNA and was made available to us by Jules Beekwilder from the Wageningen University (Cankar *et al.*, 2011). These 2 additional libraries allowed to reach a total of 198 657 EST for *C. intybus* (*cf.* table 1.12 p.65). Finally, EST libraries obtained from related species of chicory are available on the Compositae Genome Project Database website.

A limitation in the study of the Asteraceae's self-incompatibility system is probably the

feeble interest in this character for the genetic improvement of the 2 most cultivated species (lettuce and sunflower) belonging to this family. In fact, cultivated lettuce, as well as almost all of the related species used as genetic resources, are autogamous species. As for sunflower, genotypes used in the current breeding programs, are self-fertile. Additionally, despite its importance, there is a lack of existing genomic resources for this family. Recently, sequencing of the sunflower genome was initiated (Plant & Animal Genomes XVIII Conference), but so far no genome of a Asteraceae species has been sequenced. So, even when the Asteraceae SSI was investigated in *S. squalidus*, and on an ad hoc basis in sunflower and some other species, no global genomic approach seems to have been developed to identify the SI determinants in this family.

In this context, chicory may be considered as a model to undertake the identification of the SI determinants of Asteraceae for several reasons.

This crop is important in various European countries, and France is one of the first three producing countries of this species with Belgium and Netherlands. The knowledge of the SI determinants would allow the optimization of breeding programs in the cultigroups in which SI was maintained.

There is an important variability of the system of reproduction present in the cultivated varieties and wild populations of *C. intybus*, as well as within the genus *Cichorium*. This group of species also represents an interesting model for the study of the processes associated to the natural/artificial selection that have led to self-fertility.

From a practical point of view, the large number of capitula produced by a single plant and the synchronization of flowering of the flowers in a single capitulum, allow a high precision in the estimation of the compatibility behaviour of every individual to be analyzed. Furthermore, for the construction of the consensus genetic map for chicory (Cadalen et al 2010), a large population segregating for the SI trait was generated by crossing 2 strongly self-incompatible individuals. This population was highly instrumental in the development of a system to determine S-locus phenotypes, and allowed to map the S-locus to one of the ends of linkage group 2.

The objective of this thesis was to start the positional cloning of the S-locus in chicory. This strategy consists of cloning the major genes of interest by means of their genetic and physical location in the genome. This approach is potentially longer to implement than the biochemical or candidate genes approaches, but will allow in the end the identification of one or several of the desired genes. However, the speed of progress of this approach depends on the physical distance to be covered, and thus on the relation between genetic and physical distances associated with the region of interest.

Therefore, it was necessary to construct 1) a fine genetic map of the S-locus, followed by 2) a physical map of this region. The construction of a moderate density map had previously been initiated, allowing to position the S-locus on the consensus map and to establish in part the dominance relationships between S-haplotypes. These results were completed at the beginning of the thesis and are the subject of the chapter 2. They form the basis on which the works presented thereafter were realized.

In the chapter following this introduction, we will present the plant material and the various methods used during the thesis (chapter 1). Then, in part I, the different steps leading to the fine mapping of the S-locus will be described (chapters 2, 3 and 4). The construction of BAC libraries and the establishment of the physical map are presented in part II (chapters 6 and

7). Each of these parts will be discussed independently. Finally, in the conclusion a synthesis of the results is presented, as well as the perspectives of the work realized.

D.2 Part I : Fine mapping of the S-locus

Chapter 2 : Preliminary results : Moderate density map of the S-locus

A first population of 389 plants (population B_a) had been obtained after crossing 2 heterozygous individuals, K59 which is hermaphrodite and self-incompatible, and K28 which is male sterile (ms). A genetic map of the whole genome has been produced with 174 of them (Cadalen *et al.*, 2010). SSR markers were grouped and ordered in 9 linkages groups, corresponding to the haploid number of chromosomes in chicory.

Part of this K28 x K59 progeny, i.e. 20 hermaphrodites and 10 ms plants, was used to realize crosses according to diallel or factorial designs, depending on their sexual phenotype. Two classes could be distinguished unambiguously for seed-set averages of all crosses (Figure 2.2) : incompatible crosses (mean seed set < 0,1) and compatible crosses (mean seed-set >0,4). Five compatibility groups could be distinguished and 4 different S-alleles were identified with S1S2 for K59 and S3S4 for K28 (Table 2.1a). Different dominance relationships were observed in pollen and stigma (Figure 2.3).

Once the 4 S-alleles had been identified, the S-locus phenotype of 321 individuals of the B_a population was determined by crosses with selected tester plants (Table 2.1b). Because part of the individuals in the population B_a were ms, the tester plants were only used as pollen donors. In fact, the selection of 2 S-locus genotypes, S1S3 and S2S4, allowed us to identify the S-phenotypes of all plants with a minimum number of crosses to be realized. With these results, the S-locus could be mapped close to one end of linkage group 2. In parallel, the locus of the male sterility (NMS1) was mapped on linkage group 5, and was shown to have no influence on the incompatibility response. Subsequently, all the individuals of the B_a population (n = 389) were genotyped for molecular markers comprised between tacg293 and sw02H09. The S-locus was mapped in a 3,34 cM interval flanked by the markers tacg293 and EU03D06, (Figure 2.4). Thereafter, the AFLP marker tacg293 was transformed into a codominant SCAR marker, named TACG (*cf.* Matériels et méthodes p. 59).

A graphical genotype was constructed for the 389 individuals of the B_a population (Box 1). This representation revealed a difference between male and female recombination frequency in the S-locus region for 25 recombinant individuals : 3 recombination events for K28 (female) meiosis and 23 recombination events for K59 (male) meiosis in the TACG-EU03D06 interval. Study of the whole genome showed that the ratio M/F (male meiosis recombination frequency/female meiosis recombination frequency) was not homogeneous along linkage groups (Figure 2.5) and that the deviation was most important in the TACG-sw02H09 interval (M/F = 9,33). This observation has been taken into consideration during high density and high resolution mapping.

Chapter 3 : High density map

Considering maps obtained from the B_a population, the S-locus was comprised in an interval of 0,77 cM for K28 meiosis and 5,91 cM for K59 meiosis. To increase the density of molecular markers and to reduce the interval around the S-locus, a marker-assisted BSA strategy was employed (Giovannoni *et al.*, 1991 ; Jordan *et al.*, 2006 ; Michelmore *et al.*, 1991).

To target only markers in the vicinity to the S-locus, between the TACG and EU03D06 markers, 8 DNA bulks were generated (Figure 3.2) :

- 2 homozygous S1S1 (P1 : n = 8 ; P2 : n = 8)
- 2 homozygous S2S2 (P3 : n = 15 ; P4 : n = 15)
- 2 heterozygous with one recombination event between EU03D06 and sw02H09, during K59 meiosis :
 - S2S3 and S2S4 (P5 : n = 13),
 - S1S3 and S1S4 (P6 : n = 9),
- 2 heterozygous with no recombination event between EU03D06 and sw02H09 :
 - S1S3 and S1S4 (P7 : n = 13),
 - S2S3 and S2S4 (P8 : n = 9).

An AFLP approach was chosen to produce new markers. The 256 primer combinations tested, generated 31 000 bands, 2 350 thereof revealed a polymorphism between K59 and K28, and 14 were identified as potentially linked to the S-locus, between TACG and EU03D06 (Figure 3.3). One of the 14 was the previously mapped tacg293, the positions of the others were determined by graphical genotyping of the 25 recombinant individuals of the B_a population (Table 3.1 and Figure 3.4). Five of the 13 new markers were validated as really comprised in the TACG-EU03D06 interval. As they were specific for the S1 or S2 allele, they were mapped only on the K59 map (Figure 3.5), resulting in a reduction of the interval spanning the S-locus from 5,91 to 1,29 cM. In the mean time, 2 new SSR markers had been mapped in the S-locus region, resulting in a further reduction of the marker-flanked S-locus interval in the K59 meiosis map to 1,02 cM (Figure 3.5). Four of the 5 AFLP markers closely linked to the S-locus were transformed into SCAR markers and 2 showed codominant polymorphism (GGAT and TTAA, Table 3.3 and Figure 3.5).

Chapter 4 : High resolution mapping

To define how many individuals were required to construct a precise map to order markers and to estimate genetic distances, the population size was estimated according to the information content (i_r) and the α value (Figures 4.1 and 4.2, Allard, 1956 ; Liu, 1998). We have chosen to use a population with at least 2 800 individuals. In theory, this population size gave us the capacity to order makers separated by 0,27 cM with, $\alpha = 0,05$.

To reach a population size of 2 800, 2 496 individuals, obtained by combining progenies from 7 different crosses, were added to the B_a population, resulting in a total of 2 885 individuals (Figures 4.3 and 4.5, and see Table 1.1 p.47). The S-locus genotypes of these individuals were obtained in 3 steps (Figures 4.4 and 4.5).

First, all plants were genotyped for the markers TACG and EU03D06 : 2 291 individuals were non-recombinant in this interval and thus allowed the S-genotype to be inferred, 144 were recombinants, and 61 plants were excluded of further analysis because of contamination (during

Annexe D. Synthesis and translations

crosses, the seed samples, or DNA handling). Secondly, the 144 recombinants were genotyped for all the molecular markers comprised in the TACG-EU03D06 interval, and the S-genotype could be inferred for 92 individuals non-recombinant in the EU08B02-TTAA interval. Finally, the remaining 52 plants, recombinants in this interval, were phenotyped by test-crosses and determining seed-set values. The S-genotype of only 8 individuals could not be determined because of a lack of bolting ($n = 4$), dominance relationships between S-alleles ($n = 3$), and one individual that turned out to be compatible with the 2 tester genotypes.

Before the construction of a consensus map for the different populations, the effects of environment (year of seed production), genotype, and sex on the recombination frequencies were tested. Only the difference between male and female meiosis recombination frequency was significant (Tables 4.2a to 4.2f). This confirmed that the difference in M/F ratio observed between K28 and K59 was sex dependent.

Segregation distortion was observed in 2 of the 7 populations used, but turned out not to be specifically associated with a S-haplotype (Table 4.3). The genetic maps of the parents of the different populations, as well as a male and a female consensus map, were constructed (Figures 4.6 and 4.7). The S-locus was comprised in an interval of 0,60 cM for the female meiosis and 1,24 cM for the male meiosis between the markers EU08B02 and TTAA. The M/F ratio calculated from the 2 consensus maps was equal to 3,6. Differences between male and female meiosis recombination frequencies calculated for 3 intervals of the genetic map around the S-locus suggested that they are not caused by the proximity of the S-locus, as the M/F ratio was highest (4,6) in the most distant TTAA-EU03D06 interval (Table 4.4).

Discussion

The application of the AFLP technique on DNA bulks has allowed to increase marker density in the S-locus region. The number of markers obtained (1,8 AFLP/cM) turned out to be similar as the marker frequency for the entire genome (2,5 AFLP/cM). The results of similar studies (presented in table 5.1) showed a marker density in the region of interest that was at least two times more feeble than that estimated for the entire genome. This indicates that the employed strategy in chicory had been highly efficient. Among the 4 AFLP markers transformed into SCAR markers, two turned out to be codominant. Though the microsequencing technique resulted in obtaining the sequence of the 4 AFLP markers, these sequences did not permit to find polymorphisms linked to the S-locus for all of them. This illustrates the difficulties in transforming (dominant) AFLP markers into codominant markers.

The high resolution maps confirmed the monogenic character of self-incompatibility in chicory. Only a single individual turned out to be compatible with the two pollen donors used for test crosses. The descendants of this plant will offer the possibility to study the genetic determinism underlying this phenotype. This may lead to a first explanation for the presence of the different degrees of compatibility in chicory.

The populations used for the construction of this map have also confirmed that the difference in recombination frequencies observed between K28 and K59 was associated with sex and not with the S-locus. This phenomenon of heterochiasmy has been observed in many species, both plants and animals, but the different proposed hypotheses do not explain the results obtained for chicory.

Segregation distortion was also observed in association with a single S haplotype (S3), but

only in 2 populations. Furthermore, no segregation distortion was observed in homozygous S2S2 individuals, suggesting that an expected S-locus-linked sheltered load is not expressed in these individuals. Either the sheltered load does not exist, or the S2 haplotype is one of the most recessive haplotypes in chicory. In the latter case, this haplotype may be found frequently in the homozygous state, and may have led to purging of the linked sheltered load.

D.3 Part II : Physical map of the S-locus

Chapter 6 : Towards the positional cloning of the S-locus

Two BAC libraries of chicory genomic DNA were produced in collaboration with the Centre National de Ressources Génomiques Végétales (Table 6.1, Gonthier *et al.*, 2010, Annexe C). Two different genotypes have been used ; one heterozygous S1S4 (library CinS1S4), and the other homozygous S2S2 (library CinS2S2). The average sizes of the inserts have been estimated at 120 kb for CinS1S4 and 90 kb for CinS2S2. These libraries, covering 12 times the chicory haploid genome, allowed us to start chromosome walking following a systematic strategy (Figure 6.1) :

- screening of BAC libraries with probes (macro-array filter hybridization),
- PCR validation of the clones obtained by hybridization with primers of the corresponding probe,
- definition of primers on the BAC end sequences (BES) of the validated clones and use of them to build contigs (Table 6.2),
- mapping of polymorphic BES on the genetic map and orientation of the corresponding contig,
- using of the nearest BES to the S-locus as a new probe for library screening.

Initially, both BAC libraries were screened with probes defined from the molecular markers of the genetic map (Tables 6.3 and 6.4). However, because of difficulties in building a single contig using clones obtained with the same probe but for different haplotypes (Table 6.5), we have chosen to concentrate our efforts on the homozygous CinS2S2 library. Screening with 4 of the 7 probes (TACG, TTAA, AACC, and EU12G02) defined from the molecular markers gave 1 to 8 validated clones. For one probe (EU08B02), no clone was detected, and the probes for tgct160 and TCAA turned out to be aspecific (too many clones hybridized, see Figure 6.2), so the corresponding clones were excluded from the pursuit of chromosome walking. Only one clone (Si10) obtained using the TACG probe and corresponding to the region above the S-locus was validated. Three contigs could be constructed for the region below the S-locus, corresponding to clones obtained with the TTAA, AACC and EU12G02 probes, respectively (Figure 6.3). For two of the latter contigs, two polymorphic BES were mapped ; one which allowed the orientation of the AACC contig with respect to the S-locus, and the other that cosegregated with the TTAA marker (Figure 6.3). Thus, both the clone Si10 and the contig corresponding to the TTAA probe, both localized closest to the S-locus, remained to be oriented and 4 new probes were produced from BES positioned at the 2 ends of each of them. Zero to 11 clones were validated for the different probes used to screen the CinS2S2 library. Screening with the probes corresponding to the BES of Si10 resulted in obtaining only a single validated BAC clone that partially overlapped with Si10. In contrast, the BES probes of the TTAA contig yielded 2 and

11 additional BAC clones that extended this contig on both sides. Furthermore, the contig build from the 11 clones obtained for one of the probes corresponding to a BES of the TTAA contig (Si37m13) allowed to link the AACC and TTAA contigs (Figure 6.3). Thanks to the previously achieved orientation of the AACC contig, the supercontig comprising the AACC, TTAA, and TTAA-BES contigs, could be oriented with respect to the S-locus. This will allow the pursuit of chromosome walking towards the S-locus by using a probe corresponding to the BES of clone Si42, that is closest to it.

Chapter 7 : Characterization of genomic region in the vicinity of the S-locus

Six BAC clones representing the contigs around the S-locus were sequenced and analyzed ; 1 corresponding to the TACG probe (Si10 clone), 4 representing the supercontig (Si42, Si37, Si52 and Si27 clones) and 1 corresponding to the EU12G02 probe (Si26 clone). Moreover, Si37, Si52 and Si27 clones were overlapping. In addition, a BAC clone corresponding to the male-sterility locus (NMS1) was also sequenced (60A10 clone) and used for comparison with the genomic region around the S-locus.

An average of 14 contigs was obtained for the sequenced clones (Table 7.1 ; to distinguish these contigs from the contigs constructed by using BES of BAC clones, they have been renamed Cg). This number was clearly higher than obtained for BAC clones from other species that had been sequenced at the same time at the CNRGV (bell pepper, wheat, and sugar cane). A possible explanation of the difficulty in assembling the Cg was the occurrence of SSR in half of the Cg ends. Application of two different software utilities (Newbler and CAP3) did not change the Cg numbers for the clones. The total number of nucleotides sequenced for each BAC clone was comparable to their size estimated by pulse field electrophoresis, and considering the 6 clones of the S-locus region, a total of 546 kb has been sequenced.

Each BAC sequence was validated by verifying the presence of the corresponding probe sequence (Blastn). Next BES of the BAC clones in the same contig were used to order the Cg as best as possible (Figure 7.1). However, a large number of Cg could not be ordered, and to simplify further sequence analysis, all Cg of a BAC clone were concatenated to form a single sequence (Figure 7.4).

The 3 overlapping clones Si37, Si52 and Si27 allowed a first estimation of the ratio between physical and genetic distances (Mb/cM) between the 2 molecular markers TTAA and AACC. Considering the genetic maps for male and female meiosis, this resulted in a ratio of 0,5 Mb/cM and 2,2 Mb/cM, respectively. If these ratios were invariable for the region separating TTAA from the S-locus, only 4 to 8 additional consecutive BAC clones would be required to reach the S-locus.

Repeated element composition of the sequences was studied by using the RepeatMasker software and Blastx on *A. thaliana* peptide library. The total of 546 kb sequenced contained 5,71% of interspersed elements, 1,74% of SSR (1 SSR/3,6 kb) and 2,75% of low complexity sequences (Table 7.2). In addition, Blastn of clone sequences on themselves revealed 10,45% of repeated sequences without homology. Overall, similar results were obtained for the 60A10 BAC clone that corresponded to a different genomic region.

To study coding sequences near the S-locus, 2 different kinds of analysis have been condu-

ced. The first consisted of a Blastn search with BAC clone sequences on the EST libraries of several Asteraceae species, i.e. chicory, *C. endivia*, *Lactuca sativa*, *Taraxacum sp.* and *Helianthus annuus* (e-value < 10⁻³⁰). The libraries of *Cichorium sp.* turned out to be the most informative. In fact, sequences obtained for the others species were redundant to those already identified for the *Cichorium sp.*, and were therefore omitted from further analysis. A total of 112 *C. intybus* SU (Unique Sequence) and 6 *C. endivia* SU were identified that showed homology with BAC clone sequences. Thereof only 59 SU (55 from *C. intybus*, and 4 from *C. endivia*) were considered as true homologous sequences after supplemental selection criteria were applied (alignment of more than 85% of the SU sequence with genomic sequence, Figures 7.2, 7.3, and Figure 7.4).

Secondly, a Blastx search with BAC clone sequences on peptide libraries of *A. thaliana* and *V. vinifera* was realized (e-value < 10⁻⁵). This resulted in the identification of 62 genes in *A. thaliana* library and 65 in *V. vinifera* library (Figure 7.5). Approximately 55% of these genes aligned less than 50% of their length with chicory sequences (Figure 7.3). When only genes that aligned for more than 70% of their length with chicory sequences, or aligning with more than 50% of their length with chicory sequences when the region was identified as SU, were considered (Figure 7.2), the number of homologous genes was reduced to half (23 *A. thaliana* genes and 27 *V. vinifera* genes, Figure 7.5 and Table 7.3). Thereof 15 *A. thaliana* genes and 16 *V. vinifera* genes were positioned in regions that aligned with SU. Of the 59 SU, 33 (56%) showed homology with *A. thaliana* and/or *V. vinifera* genes, and 25 (76%) thereof were homologous to the same gene as found upon aligning with the genomic sequence.

Blastx of the SU have also been done on peptide libraries of *M. guttatus*, *S. lycopersicum*, *P. trichocarpa* and *M. truncatula*, but no additional genes were identified. This suggests that searches for the gene composition of chicory sequences can be limited to the use of *A. thaliana* and *V. vinifera* peptide libraries.

To try to distinguish orthologous genes among all identified homologous genes, an additional criterion was applied : the ratio between the first and the second hit e-value (Δ e-value, Figure 7.2). When Δ e-value < 10⁻⁵, the gene of *A. thaliana* or *V. vinifera* was considered as the best homologous of the chicory gene, and called potentially orthologous (pO gene), otherwise it has been considered to belong to the same gene family (H gene). Consequently, 20 of the 27 *V. vinifera* genes and 14 of the 23 *A. thaliana* genes were classified as pO genes. Unfortunately, without access to the whole chicory genome sequence, it was impossible to identify the RBH (Reciprocal Best Hit) of *A. thaliana* and *V. vinifera*, and to validate these genes as orthologous. Nonetheless, tBlastn searches performed for 9 *A. thaliana* and 14 *V. vinifera* genes in the *Cichorium sp.* EST libraries resulted in the identification of 5 *A. thaliana* genes and 8 *V. vinifera* genes as most likely orthologous genes (pO* genes, Table 7.3).

The Blast search results were compared to gene prediction software (GenScan and Fgenesh) results. On average, 12 genes per clone sequence were identified on the basis of homology with *A. thaliana* or *V. vinifera* genes, but only 3 genes were predicted to be present at the same position as that of SU or of *A. thaliana* and *V. vinifera* genes that aligned with the genomic sequence. This suggests that the tested gene prediction software may have to be optimized for application on chicory genomic sequences.

A certain degree of conservation of microsynteny was observed between genes identified on genomic sequences of chicory in the S-locus region and their homologues in five Dicotyledon species (*V. vinifera*, *S. lycopersicum*, *M. guttatus*, *P. trichocarpa* and *M. truncatula*).

Annexe D. Synthesis and translations

For *V. vinifera*, 65% of the pO genes were localized on chromosome 18 of this species (Figure 7.7). Groups from 2 to 4 genes were distributed on intervals spanning from 22 to 1900 kb along this chromosome. Each block corresponded to homologous genes found present on one of the BAC clones in the S-locus region of chicory. The order of blocks corresponded to the order of the clones positioned below the S-locus but was inverted relative to the block corresponding to the Si10 clone. Furthermore, distances between the blocks were much larger than between the corresponding chicory sequences.

For the other species, genes used to study conservation of synteny were obtained from Blastx of SU on their respective peptide libraries. For each species, at least 1 chromosome, scaffold, or linkage group was identified that represented at least 5 of the 6 clones' sequences (Figure 7.8). In all the cases, blocks of genes homologous to genes identified in one BAC clone were found present, each spanning at most several hundreds of kb. In particular tomato seems to be very interesting, in view of the similarities in the scale and the order of the gene distribution on chromosome 4 of this species, in comparison with the physical map of the S-locus region on linkage group 2 of chicory.

Discussion

During the construction of the physical map, at least 2 recombination events were observed in an interval of less than 50 kb. This illustrates the kind of information which can be obtained after genotyping 2824 individuals for the S-locus region, and confirms the importance of the number of plants used for the construction of a high resolution map.

A Mb/cM ratio was calculated for the first time (between 0,5 and 2,2 Mb/cM), and allowed an estimation of the distance to overcome towards the S-locus (between 0,5 and 0,8 Mb). This result, in combination with the position of the S-locus at the linkage group extremity, suggests that this locus is not localised near a centromere. The physical distance delimiting the S-locus may thus correspond to a few hundred kb, similar as in *I. trifida* or *B. campestris*.

The difficulties encountered during the assembly of the genomic sequences can probably be explained by the high fraction of microsatellites and different types of repeated sequences, some of which are homologous to annotated mobile elements. However, these characteristics were observed both for the S-locus region and the clone harbouring the male sterility locus NMS1, indicating that they are not specific for the former. Furthermore, they may also explain, at least in part, the difficulties encountered during the construction of contigs corresponding to the different S-locus haplotypes.

These difficulties also underpinned the importance of the availability of a BAC library generated from the DNA of a genotype that was homozygous for the S-locus region. Moreover, in view of the complexity of the studied region, the strategy of contig construction has been proven to be well adapted. The stringent criteria for clone selection allowed obtaining reliable results and orienting the chromosome walking in the direction of the S-locus.

Despite the phylogenetic distance between chicory and *A. thaliana* and *V. vinifera*, the analysis of the coding sequences' fraction proved to be very informative. In contrast, the application of gene prediction software was less evident, and the parameters have to be optimized in order to obtain reliable results for the chicory genome.

The study of the conservation of synteny between chicory and 6 dicotyledonous species revealed the limits inherent to the lack of information on the entire genome. Therefore it was

not possible to unambiguously identify orthologous genes. Nonetheless, a relative co-linearity was observed between genes identified in chicory BAC clones and their homologues in 5 of the 6 species studied (*V. vinifera*, *S. lycopersicum*, *M. guttatus*, *P. trichocarpa* and *M. truncatula*). Furthermore, among the genes present in blocks that corresponded to the clones, some had been classified as potential orthologues and others as homologues. It thus seems more pertinent to consider the information contributed by genes present in the same block rather than to consider only orthologous genes. Finally, among the species studied for synteny, tomato turned out to be particularly interesting; its chromosome 4 presented the same organisation of the blocks of genes as in the chicory S-locus region, and that were separated by comparable physical distances (Figure 8.1).

D.4 Conclusions and perspectives

The objective of this thesis was to start the positional cloning of the chicory S-locus. It was previously positioned at one end of linkage group 2 and a moderate density map of this region for the population B_a (n = 389), was in progress. To reach the fixed objective, we first established a high density/high resolution map, and secondly started the construction of a physical map of the S-locus region. To realize fine mapping, the moderate dense genetic map was completed by determining the genotype of all the individuals of the population B_a with the existing markers (*cf.* chapter 2). This map allowed to identify 25 recombinant plants in the region of the S-locus and showed a significant difference between the rates of recombination during female and male meiosis in K28 and K59, respectively.

A strategy to increase the density of molecular markers in the interval studied was implemented. Five new AFLP markers were identified, of which one was cosegregating with the S-locus. Two of the AFLP markers could be transformed into codominant SCAR markers. The chosen approach (marker-assisted BSA, *cf.* chapter 3) turned out to have been efficient, because the markers density in the targeted region was comparable to the expected density calculated for the whole genome.

Simultaneously, 7 populations, representing a total of 2 824 individuals, were analyzed to build a high-resolution map of the S-locus region (*cf.* chapter 4). For the 7 studied populations, the S-locus was always mapped between EU08B02 and TTAA markers, at one end of linkage group 2. Genotyping of these plants thus confirmed the monolocus character of the chicory self-incompatibility trait and the relevance of the choice of these populations. Two consensus maps were built from the 2 824 individuals, one positioning the S-locus in an interval of 1,24 cM (male meiosis), and the other one, in an interval of 0,6 cM (female meiosis). These maps, as well as the 169 recombinant individuals identified in the S-locus region, constitute a powerful tool to map new molecular markers, because a minimum of 2 recombination events were observed in an interval of less than 50 kb.

Finally, the genetic data demonstrated that the differences in the rates of recombination-between male and female meiosis, first observed between K28 and K59, were linked to sex. This difference is not associated directly with the S-locus. It has however an effect on the progress of the chromosome walking estimated from the genetic map, depending on whether the male or female meiosis is considered, the first one contributing more information so far.

For the construction of the physical map of the S-locus, 2 BAC libraries were generated,

Annexe D. Synthesis and translations

covering together 12 times the chicory haploid genome (*cf.* chapter 7). The CinS1S4 library was built from a heterozygous genotype in the S-locus region and the CinS2S2 library was made from a homozygous genotype. Three contigs of BAC clones were built on both sides of the S-locus, for the S2-haplotype. The orientation of one of them, the supercontig positioned under the S-locus, was defined according to the position of markers on the genetic map, and allowed to determine the direction of chromosome walking towards the S-locus.

Because of the difficulties to make contigs from clones representing different haplotypes, the chromosome walking was followed using clones stemming from the CinS2S2 library. This underlined the importance of having at least one BAC library built from a homozygous individual. The screening of the libraries was realized by hybridization on macroarrays. Two probes obtained from AFLP markers, in particular TCAA which cosegregated with the S-locus, presented aspecific hybridization patterns. This method of screening is less expensive but also less sensitive than a screening by PCR. However, if the chromosome walking is interrupted, because of probes showing aspecific hybridization, a new approach must be adopted. For example, the implementation of DNA pools of BAC clones would allow the screening of libraries by PCR, with primers very specific of the targeted sequences.

Six BAC clones identified near the S-locus were sequenced (*cf.* chapter 7). These sequences allowed, for the first time, to study and to characterize genomic sequences of chicory. A high number of Cg was obtained for each clone (from 7 to 23). However, these sequences allowed a first estimation of the Mb/cM ratio in the neighborhood of the S-locus (approximately 0,5 Mb/cM for the male meiosis and 2,2 Mb/cM for the female meiosis). The composition in repeated elements was determined : 5,71% of interspersed elements, 1,74% of SSR, 2,75% of low complexity regions, and 10,45% of repeated regions without homologies. Potentially coding or expressed regions were identified from the EST libraries of *Cichorium sp.* (59 SU) and from peptide libraries of *A. thaliana* (23 genes) and *V. vinifera* (27 genes). Among the potential genes identified, at least 8 of them could correspond to single copy genes in the chicory genome. The difficulties to make contigs from BAC clones originating from different haplotypes during the chromosome walking were probably caused by the presence of repeated sequences and mobile elements in the BES. The availability of complete BAC clone sequences would allow the identification of coding and single copy regions, which could be used as anchor points to pass from one library and one haplotype to another.

The study of the coding fraction also highlighted a relative preservation of the microsynteny between the chicory sequences and those of at least 5 other Dicotyledonous species (*V. vinifera*, *S. lycopersicum*, *M. guttatus*, *P. trichocarpa* and *M. truncatula*). These results show that in the absence of data on Asteraceae genome sequences, it is possible to use the information of sequences of phylogenetically distant species. However, it should be kept in mind that these resources do not necessarily contain genes equivalent to the chicory SI determinants.

In view of these results, the objectives fixed at the beginning of the thesis were reached. The strategy consisting of building contigs by using the BES proved its efficiency and the chromosome walking towards the S-locus will continue. However, the preliminary data obtained recently open the way to new approaches which could accelerate the implemented strategy. In fact, a further in depth study of the conservation of the synteny between genes positioned in the neighborhood of the S-locus and their position on the chromosomes of the other species may be very helpful.

To initiate this approach, vine and tomato seem the most promising species. Indeed, *V. vi-*

nifera is the species which groups together the largest number of genes associated to the same chromosome. As for tomato, it is the only species for which the blocks of syntenic genes are distributed in an interval of only a few Mb, as well as organized in the same order as in chicory. In tomato, by listing all the present genes comprised in the 4 Mb interval which separates the blocks of genes A and B (*cf.* figure 8.1 p.183), it should be possible to potentially identify single copies genes. Chicory's SU, homologous to these genes, can then be used to look for polymorphism to define new molecular markers. A part of markers so defined may potentially be mapped in the region of the S-locus. If these markers are positioned at smaller genetic distances of the S-locus than the markers which are already mapped, they would allow the S-locus to be reached more quickly. In addition, it may allow to start chromosome walking again in the region above the locus S, beyond the EU08B02 marker, for which no clone has been validated so far (*cf.* figure 6.3 p.139).

Finally, when the sequence of the genome of sunflower will be available, it will be interesting to determine the extent to which the synteny with the S-locus region of chicory is conserved, and whether the region of the S-locus corresponds to the region of the self-fertility QTL on linkage group 17 of sunflower (Burke *et al.*, 2002 ; Gandhi *et al.*, 2005 ; Wills et Burke, 2007).

In anticipation of obtaining sequences of the S-locus region, it will be also important to resolve the problem of the assembly of genomic sequences. The simultaneous use of different assembly software packages may be a first step in arriving at better assembly results, in particular to reduce the number of Cg per BAC clone. The efficiency of this type of strategy was demonstrated for transcriptomic analyses (Kumar et Blaxter, 2010), and the first tests realized on chicory genomic sequences seem promising. However, if no satisfactory protocol is obtained, additional strategies of sequencing must be envisaged to order Cg.

When the construction of the physical map of the S2-haplotype will have sufficiently progressed, the analysis of the genomic sequences corresponding to the S1- and S4-haplotypes can be envisaged. For that purpose, the use of sequences corresponding to single copy genes will be privileged. In view of the characteristics in common between the different S-loci identified until now, the comparison of the sequences of the different S-haplotypes present in the BAC libraries (S1, S2 and S4) should allow delimiting the S-locus in a hyper-variable region. The female and male determinant identification will require the implementation of tools of gene predictions. Adapted and specific parameters of chicory must be defined for an optimal use of the prediction software. Within this region, the most polymorphic predicted genes will represent the most likely candidates of the self-incompatibility determinants. The association of the polymorphism of these genes with different haplotypes and their specificity of expression, in the stigma or the anthers, will also have to be considered during the selection of the candidates. Furthermore, the specific presence of equivalent sequences in the EST library of stigmas will constitute an additional argument in the choice of the female candidates to validate.

Finally, gain-of-function transformations will represent the final step of the functional validation of these genes. The S2-haplotype being the most recessive in the studied populations, it offers the possibility to obtain S2S2 homozygous individuals, among whom it will be necessary to select those that possess a good genetic transformation capacity. The candidate genes corresponding to dominant haplotypes, for example S1, will be used to for transformation. Only individuals transformed with the SI determinants will then present a S1 phenotype at the pollen or stigma level, depending on whether it concerns the male or female candidate to be validated, respectively.

Annexe D. Synthesis and translations

The access to the sequences of female and male determinants in chicory should open the way to the study SI in other Asteraceae. It would allow to determine if this mechanism is common to the different species of self-incompatible Asteraceae, in particular to *S. squalidus*, species for which several S-haplotypes were characterized. If the determinism proves to be common, the intra- and inter-specific diversity at the molecular level, as well as the differences which exist between SI and allogamous species of the same genus can be investigated.

All this information will provide indications that may explain the appearance of the self-incompatibility or to understand which factors are responsible for its breakdown in the Asteraceae family. The study of the variability for the self-incompatible character in chicory (presence of pseudo-compatible and self-fertile individuals) will allow defining its origin. The breakdown of SI is caused either by the loss of function of the SI determinants of some haplotypes, or it is due to the action of one or several genes modulating the SI response.

Annexe E

Captions translation

Introduction

Figure 1 : Schematic representation of pollen-pistil interactions occurring during self-incompatibility (SI) mechanisms. When pollen and pistil have the same phenotype, a recognition mechanism prevents fecundation. Contrarily, the fecundation is allowed when pollen and pistil have different phenotypes.

Table 1 : Synthesis of the characteristics of the different SI systems.

Figure 2 : Repartition of the different SI systems among Angiosperms.

Figure 3 : Schematic representation of the S-locus structure in homomorphic SI system. The S-locus includes at least 2 genes, one encoding the female determinant, the other encoding the male determinant. Each allelic combination represents a different haplotype.

Figure 4 : Heteromorphic SI. L, M and C represent long, medium and short morphs, respectively. The S allele is dominant and determines the C morph. In the case of tristylly, ss individuals have either an L morph, when their genotype is mm at the M-locus, or an M morph, if they have at least an M allele. Anthers and stigmas are represented as white ovals and black triangles, respectively. Compatible crosses are indicated with dashed arrows.

Figure 5 : Gametophytic SI. (a) : Gametophytic SI involving one locus (S-locus). The SI reaction occurs when the S-allele expressed by the pollen corresponds to one of the two alleles expressed by the pistil. (b) : Gametophytic SI involving two loci (S- and Z-locus). The SI reaction occurs when the two alleles S and Z expressed by the pollen correspond to 2 of the 4 alleles expressed by the pistil.

Figure 6 : Putative model to explain the interaction between S-RNases and SLF/SLB proteins in Solanaceae, Rosaceae and Scrophulariaceae. (croissance = growth ; arrêt croissance = inhibition of growth)

Figure 7 : Reaction cascade following the interaction between PrpS and PrsS proteins in Papaveraceae.

Figure 8 : Sporophytic SI. The pollen phenotype is determined by the genotype of the plant producing the pollen. In this example, the S1 haplotype is dominant over S2 and S3, and the S2 haplotype is dominant over S3.

Figure 9 : Structural heteromorphism of S-haplotypes in *Brassica rapa*.

Figure 10 : Molecular model representing the SI response in Brassicaceae, with S1 dominant over S2, or codominance between S1 and S2. The mark ‘?’ represents proteins (e.g. MLPK, THL1, THL2, ARC1) involved in the reaction cascade activated by SI.

Figure 11 : Chicory pictures. A) Whole plant (scale 1/13), B) longitudinal cut of a capitulum (scale 1/1), C) detail of an isolated flower (scale 2).

Chapter 1 : Materials and methods

Figure 1.1 : Different steps of the positional cloning procedure and plant materials used for the different steps.

Carte génétique moyenne densité = moderate density map

Carte génétique fine = fine mapping

Carte haute densité = high density map

Carte haute résolution = high resolution map

Banques BAC = BAC libraries

Marche chromosomique vers le locus S = chromosome walking towards the S-locus

Table 1.1 : Characteristics of the population used for the positional cloning of the S-locus in chicory. The B population was obtained from the cross between K28 (male sterile) and K59 (hermaphrodite) genotypes. The C population was obtained from selfing of K59 (C_a), and selfed progenies of K59 (C_b)*. The populations D, E, F, and G were obtained from crosses of B individuals, populations E, F, G having the same female parent (B272). The populations H and M were produced from crosses between K59, used as male parent (H) or female parent (M), with two groups of B individuals (B_H and B_M).

* Even if the K59 genotype is highly SI, the harvest of all seeds in the capitula of 10 isolated clones of K59 led to obtaining 36 plants (C_a). Similarly, the selfing of the 36 C_a individuals resulted in obtaining a total of 122 C_b plants.

Figure 1.2 : Genealogy of the different chicory populations used.

Figure 1.3 : *In vitro* culture. Schematic representation of the different steps necessary to maintain and multiply (somatic embryogenesis) SI genotypes.

Figure 1.4 : Greenhouse culture. Plantlets growing in pot plates.

Table 1.2 : PCR mix for the amplification of STS markers.

Table 1.3 : Mixes for the separation of PCR products by capillary electrophoresis, under denaturing conditions (POP7 polymer) or non-denaturing conditions (CAP polymer).

Table 1.4 : PCR mix for the detection of HRM polymorphism.

Figure 1.5 : Schematic representation of the different steps of the AFLP procedure.

Table 1.5 : Enzymatic digestion mix

Table 1.6 : Ligation mix

Table 1.7 : Pre-amplification mix

Table 1.8 : Selective amplification mix

Table 1.9 : List of primers EcoR1-A+2 and MseI-C+2 used for selective amplifications.

Figure 1.6 : Schematic representation of the different steps in the simplification of the AFLP profile of marker aacc134 (minisequencing). 1 : the 4th selective base is C. 2 : the 5th selective base is C. 3 : the 6th selective base is A.

Table 1.10 : Degenerated primers used for the identification of the 4th, 5th and 6th selective bases, during the minisequencing procedure.

Table 1.11 : Sequencing polyacrylamide gel composition

Table 1.12 : EST databases of the different Asteraceae species used for sequence comparisons. EST : total number of EST sequences (before assembling). SU : number of unique sequences (singletons and contigs) after EST assembling. Contig : number of contigs among SU. Annoté : percentage of SU presenting a homology with *A. thaliana* peptide database.

Table 1.13 : Peptides databases used for sequence comparisons.

Table 1.14 : Expression of i_r (= information carried by an individual) depending on the type of marker considered. He : heterozygote, Ho : homozygote, CC : full progeny classification, BC : back-cross, DD : dominant-dominant configuration. * In the case of dominant markers, the individual is Ho for recessive alleles.

Table 1.15 : Examples of expected frequencies (m) and corresponding derivatives ($\delta m/\delta r$) for the different genotypes/phenotypes expected in a progeny. (a) Case of 2 markers with CC configuration (parental genotypes : A1A2 B1B2 and A1A3 B1B3). (b) Case of 2 markers with configuration DD in repulsion phase (parental genotypes : AaBb). To obtain the expressions corresponding to a configuration DD in repulsion phase, 1-r is substituted by r.

Table 1.16 : Scheme of crosses used to determine the S-phenotype of recombinant individuals around the S-locus. + indicate compatible crosses ; - indicate incompatible crosses.

Part I : Fine mapping of the S-locus

Chapter 2 : Preliminary results : Moderate density map of the S-locus

Figure 2.1 : Basis of the S-locus positional cloning approach. Synthesis of the different tasks performed with the B_a population. H : hermaphrodite individual, MS : male sterile individual.

Croisement = cross

Cartographie = mapping

Figure 2.2 : Seed set frequencies. Frequencies were calculated from a total of 662 capitula obtained from diallel and factorial crosses (30 individuals). Seed set values were associated with the compatible or incompatible behaviour of the crosses.

Figure 2.3 : Dominance relationships between the different S-alleles at the pollen and stigma level. Arrows indicate dominance and double lines indicate codominance.

Table 2.1 : Results of diallel and factorial crosses. (a) Compatibility groups and dominance relationships between S-alleles. Crosses were performed between individuals of the B_a popu-

lation, and with the K28 and K59 genotypes. As K28 is male sterile, the pollen group I was established from hermaphrodite S3S4 individuals of the D population (obtained from a cross between two individuals of the B_a population). Grey squares correspond to compatible crosses, striped squares indicate reciprocal crosses with different results. (b) Tester genotypes choice and crosses' design used for the S-locus phenotyping of B_a individuals.

Figure 2.4 : Moderate density map of linkage group 2 obtained from the B_a population (left, n = 174), and detail of the tacg293-sw2H09 interval (right, n = 389).

Box 1 : Graphical genotyping : description and example.

Graphical genotyping is a schematic representation of a chromosomal region : each line corresponds to a locus and a column to a chromosome. A color code is used to distinguish the parental chromosomes and to represent the genotypes of the descendants. Only the order of the loci is represented, without considering the genetic distances. Graphical genotyping allows a precise visualization of the recombination events and to associate them with the meiosis in one or the other parent. It allows also the detection of errors during genotyping : these may generate artificial double cross-overs that are unlikely to occur in intervals in the order of 1 cM. For each marker used for the construction of the genetic map, the nomenclature of alleles follows the nomenclature used in the JoinMap mapping software. When the 2 parents (K28 and K59) are heterozygous for a locus :

- with 4 different alleles, the alleles are named a,b, c and d ;
- with 3 different alleles, the alleles are named e, f, and g ;
- with 2 different alleles, the alleles are named h and k.

When only one parent is heterozygous, here K59, the alleles of the locus are named l and m.

Figure 2.5 : Comparison of recombination frequencies between K28 and K59 meiosis, for the 9 linkage groups of the chicory genetic map (LG1 to LG9, B_a population, n = 174). For each linkage group, only codominant markers in K28 and K59, and separated by about 10 cM, were selected. The number of recombinants per interval was compared between K28 female meiosis (F) and K59 male meiosis (M). When a significant difference was observed (χ^2 test), the ratio M/F was calculated (M/F > 1, grey boxes; M/F < 1, striped boxes). The total length of linkage groups, in cM, is indicated above each linkage group. The length of the common part of male and female map is indicated between parentheses.

Chapter 3 : High density map

Figure 3.1 : Moderate density map of the S-locus region (B_a population, n = 389) obtained from female meiosis (K28) and male meiosis (K59). Cumulated genetic distances (cM) are indicated on the left of linkage groups, and the marker's names are indicated on the right.

Figure 3.2 : Bulk strategy. Graphical genotypes representation of individuals selected to constitute the bulks, and expected results depending on the linkage configuration of markers with S1 or S2 haplotypes. + indicates the presence of a marker. Each colour corresponds to a parental chromosome. For individuals S1S1 and S2S2, striped part of the chromosomes indicates that the genotype is unknown. Expected results when markers are linked to S1 or S2 haplotypes are indicated in the two first lines (marker linked to S1 in the interval TACG-

EU03D06; marker linked to S2 in the interval TACG-EU03D06). * : Examples of negative results are indicated in the two following lines (marker not linked to S1 or S2; marker distant from S1).

Table 3.1 : Characteristics of AFLP markers identified with the BSA strategy, and information related to their mapping in the vicinity of the S-locus.

Figure 3.3 : Successive results obtained during the search of new AFLP markers by BSA.

Figure 3.4 : Graphical genotypes representing map position of markers identified by BSA in the TACG-EU03D06 interval. The graphical genotype of parental genotypes (K28 and K59) and the different configuration observed among the 25 recombinant individuals of the B_a population are represented. Cosegregated markers are boxed. * The position of marker tcaa300, above or under the S-locus is undetermined because of the limited genetic information provided by this marker.

Figure 3.5 : Female (K28) and male (K59) high density maps of the S-locus region (B_a population, n = 389). AFLP markers are indicated in lower-case letters (e.g. ggat125) and the corresponding SCAR markers are indicated in capitals (e.g. GGAT). Microsatellite markers are indicated in capital letters. Cumulated genetic distances (cM) are indicated on the left of linkage groups.

Table 3.2 : Results of the minisequencing procedure. The 4th, 5th and 6th supplementary selective bases are indicated for each AFLP marker, as well as the 3' extremity of the MseI primers designed.

Figure 3.6 : Example of minisequencing result obtained for the AFLP marker ggat125. The AFLP pattern before minisequencing is presented on the left picture. The right picture shows the simplification of the AFLP pattern after minisequencing of 3 supplementary bases. In this case, the minisequencing procedure led to the detection of the ggat128 fragment, specific of the S1S1 genotype. The two first lines shows AFLP patterns of S1S1 genotypes, the two last lines correspond to S2S2 genotypes.

Table 3.3 : Polymorphism information of AFLP markers and of their equivalent SCAR markers.

Chapter 4 : High resolution mapping

Figure 4.1 : Quantity of information contributed by individual (i_r) as a function of r. The i_r value is expressed proportionally to a full F2 classification (100% F2 CC). BC is for back-cross, 'DD couplage' indicates an F2 configuration for 2 dominant markers in coupling phase, and 'DD repulsion' indicates an F2 configuration for 2 dominant markers in repulsion phase.

Figure 4.2 : Minimum number of individuals required to estimate the confidence interval $r \pm z_{\frac{\alpha}{2}} S_r$ (with $z_{\frac{\alpha}{2}} S_r < \frac{r}{2}$) in the case of a full F2 classification (CC) and in the case of dominant markers in coupling phase (DD couplage), for $\alpha = 0,01$ and $\alpha = 0,05$. The marks (●) indicate the r values for n = 2800, depending of α and of the marker's configuration.

Figure 4.3 : Graphical genotypes of the TACG-EU03D06 interval for the parental genotypes of the 7 populations used for high resolution mapping of the S-locus. Underlined molecular markers indicate markers used for the selection of the populations, the other marker were

obtained later on. White boxes in graphical genotypes indicate non-informative markers for the considered genotype. In progenies with lm x lm configuration for markers (*), mm and lm genotypes can not be distinguished, because of the dominance of the m allele.

Table 4.1 : Minimum values of r (non-overlapping confidence intervals) for a population of 2824 individuals, depending of α and of marker's configuration.

Figure 4.4 : S-locus genotyping and phenotyping strategy in three steps. Step 1 : genotyping of 2496 individuals for TACG and EU03D06 markers. Step 2 : genotyping of the 144 recombinant individuals for markers TTAA, aaca83, aacc134 and EU12G02. Step 3 : Crosses with tester genotypes to determine the S-locus phenotype of the 52 individuals with recombination in the EU08B02-TTAA interval. Parental and progenies chromosomes are representing as graphical genotypes, each box corresponding to a molecular marker, each colour corresponding to a parental chromosome. Hors type = aberrant.

Figure 4.5 : Overview of the seven populations analysed. The scheme indicates the parental genotypes of populations (female x male), the years of seed production, the number of seed (graines) obtained, the final number of individuals analysed, and the number of recombinant individuals in the interval TACG-EU03D06. The final number of individuals (nombre d'individus après germination et génotypage) does not take into account the 61 aberrant individuals indicated in figure 4.4

Table 4.2 : Homogeneity test results (χ^2) for the comparison of the proportion of recombinant and non-recombinants individuals in the TACG-EU03D06 interval, under different parameters :

- the year of seed production (broad sense environment), (a) for male meiosis in K59 and (b) for female meiosis in K28 ;
- (c) the male and female meiosis (sex effect) in the K59 genotype ;
- (d) the genotype of female parent and (e) the genotype of male parent ;
- (f) sex effect for the total number of progenies.

* When at least one of the expected numbers of individuals was less than 5, the Yates correction was applied to the χ^2 calculation.

Table 4.3 : Results of segregation conformity tests (χ^2) for the alleles of TACG and EU03D06 markers, and the S-locus. Significant results ($P < 0,05$) are indicated in bold. * The segregation distortions observed in these populations indicate an under-representation of the S3 haplotype and marker's alleles linked to this haplotype.

Figure 4.6 : Genetic maps of the TACG-EU03D06 interval (top end of linkage group 2), for the different parental genotypes analysed. Cumulated genetic distances (cM) are indicated on the left of linkage groups. Maps built from female and male meiosis are indicated in pink and blue, respectively. * The distance between the S-locus and flanking markers is an approximated value, due to missing data.

Figure 4.7 : Consensus genetic maps of the TACG-EU03D06 interval (top end of linkage group 2), obtained from total female and male meiosis. Full lines on linkage groups indicate informative markers for the 2824 meioses. Distances between markers (cM \pm confidence interval, $\alpha = 0,05$) are indicated on the left of linkage groups. * The distance between the S-locus and flanking markers is an approximated value, due to missing data. Dashed lines indicate the positions of informative markers, only in certain populations. Distances calculated

between these markers and flanking markers are indicated in cM, with in parentheses the ratio recombinants/informative meiosis.

Table 4.4 : Sex effect (χ^2 test) on the proportion of recombinant and non recombinant individuals, for different intervals flanking the S-locus.

Discussion

Table 5.1 : Examples of results obtained for the BSA approach, using AFLP, based on phenotype (BSA) or marker assisted (BSA+), for different plants and different traits. Indicated are : the size of the genetic map (1), size of the genome (2), the total number of fragments generated (3), the percentage of polymorphic markers between parents (4), the number of markers identified as being linked to the locus of interest (5), the interval in which the linked markers were identified (6), the mean distribution of the polymorphic markers in the entire genome (7) and in the region of interest (8), the estimated ratio physical distance/genetic distance for the entire genome (9). References (Réf.) : a) Lahaye *et al.* (1998) et Wicker *et al.* (2008), b) Meksem *et al.* (1995) et Zhu *et al.* (2008), c) Jordan *et al.* (2006), Pierre *et al.* (2000) et Yoo *et al.* (2003), d) Cnops *et al.* (1996), Hauge *et al.* (1993) et Tagu et Moussard (2003), e) Cette étude, f) Zhang *et al.* (2009), g) Noguera *et al.* (2005), h) Pierre *et al.* (2000), i) Wei *et al.* (2009), j) de Brito *et al.* (2010), k) Quillet *et al.* (in prep).

Table 5.2 : Chromosomal location of the S-locus and the relation between physical and genetic distance in the proximity of the S-locus in different species. (*) relation physical distance/genetic distance close to the S-locus, (**) relation physical distance/genetic distance estimated for the entire genome. References : a) Wang *et al.* (2003), b) Yang *et al.* (2007), c) Casey *et al.* (2010), d) Casselman *et al.* (2000) et Imano *et al.* (1998), e) Suzuki *et al.* (2004), f) Wheeler *et al.* (2003).

Part II : Physical map of the S-locus

Chapter 6 : Towards the positional cloning of the S-locus

Table 6.1 : Characteristics of the chicory BAC libraries according to Gonthier *et al.* (2010, Annexe C).

Table 6.2 : Example of the construction of a contig from the clones obtained by the TTAA probe. The presence (1) or the absence (0) of amplification products of the BES (Si39t7 to Si37m13) has permitted to position the clones in relation to the others. The contig has been constructed bearing in mind that the BES clones should imperatively be positioned at the ends of that clone, once the contig has been assembled.

Figure 6.1 : Different steps of chromosome walking toward the S-locus.

Table 6.3 : Results of the hybridization (Hyb) and of the PCR validations (Val) on the CinS2S2 library. The clone names correspond to the library plate number and to the position of that clone on that plate. The abbreviations (Abr) used are those usually employed to designate the clones. Concerning the probes defined from the BES (Si37m13 and Si39m13), only the newly identified clones have been indicated. The presence (P) or the absence of polymorphisms (np)

has been indicated for each BES. The identification of a gene has also been signalled. When a polymorph BES could be placed on the genetic map, it has been highlighted in yellow. The _ symbol corresponds to an absence of data.

Table 6.4 : Results of the hybridization (Hyb) and of the PCR validations (Val) on the CinS1S4 library. Haplotype (Hapl) with which the clone is associated, and the amplified allele (All) corresponding to the marker used as a probe, are indicated. The clone names correspond to the library plate number and to the position of that clone on that plate. The abbreviations (Abr) used are those usually employed to designate the clones. The presence (P) or the absence of polymorphisms (np) has been indicated for each BES. The identification of a gene has also been signalled. The _ symbol corresponds to an absence of data.

Table 6.5 : Comparison of the contigs obtained for the clones associated to the a and c alleles of the TACG marker, for the CinS1S4 library. The order of the BES markers (indicated on the 1st line of each table) in relation to the others is not the same for each contig, and thus has not allowed the construction of a single contig. A few markers have been connected from one contig to the other, in order to underline these differences.

Figure 6.2 : Macroarrays. Example of an aspecific hybridization on the filter on the left hand side (TCAA probe) and of a specific hybridization on the filter on the right hand side (TTAA probe), with 2 copies of each of the five hybridized clones circled in red.

Figure 6.3 : Chromosome walking toward the S-locus. The markers of the genetic map have been positioned in relation to the others, depending on the distances obtained for the consensus map of the male meiosis (*cf.* Chapitre 4 p.114). All the clones obtained with one probe and organised as contigs are represented with the same colour. The markers of the genetic map as well as the BES that have been used as probes have been connected to the corresponding identified clones (continuous lines). The polymorphic BES which were mapped have been positioned on the genetic map (dotted lines). The successive construction of contigs has permitted to obtain a supercontig physically connecting the TTAA and the aacc134 markers. The entirely sequenced clones (*cf.* chapitre 7) are also indicated (*).

Chapter 7 : Characterization of genomic region in the vicinity of the S-locus

Table 7.1 : Characteristics of the inserts obtained from the CinS2S2 library. For each clone, the linkage group and the probe with which it has been identified are indicated (LG_sonde). Si37, Si52, and Si27 (gray) clones are overlapping. Averages and totals had been calculated for the 6 clones positioned on the LG2 around the S-locus. (a) Insert size estimated on electrophoresis gel and (b) total length after assembling. (c) number of Cg of at least 500 pb ; (d) Total considering the overlapping regions of the Si37, Si52, and Si27 clones only once. The _ symbol indicates the absence of data.

Figure 7.1 : Organisation of the Si10, Si42, Si37, Si52, Si27 et Si26 clones' Cg. Only BES and amplified markers on the sequenced clones are indicated (*cf.* chapitre 6 p.132). Each Cg of the 6 clones is represented by a rectangle proportional to its length. Markers of the genetic map represented at the top are connected to the corresponding clone by an arrow. BES and markers are connected to their position on a Cg by a blue line. *The Si10 clone not belonging

to a contig, the positions of amplified markers on this clone are arbitrary.

Table 7.2 : Characterization of sequences in the neighborhood of the S-locus for repeated elements. The mobile elements, microsatellites and low complexity regions, as well as the GC content were determined with RepeatMasker and Blastx searches on the peptide library of *A. thaliana*. The duplicated regions were identified with Blastn search of the 6 clone sequences on themselves. (a) The coverage corresponds to the total number of pb corresponding to the different elements present in the 573,7 kb representing the 6 clones. (b) This percentage corresponds only to the 60A10 clone's genomic regions that were identified on the 6 clones of the S-locus region. (c) Presented results were obtained from the whole *A. thaliana* genome (119 Mb, TAIR10 version).

Figure 7.2 : Selection criteria used for the different sorts of Blast : Blastn, Blastx, and tBlastn. At : *A. thaliana*, Vv : *V. vinifera*, Δ e-value : ratio of the e-values of the 2 best Blast hits.

Figure 7.3 : Frequencies of the parts of sequences representing SU and peptidique sequences deduced from the *A. thaliana* and *V. vinifera* genes that aligned with BAC clone sequences. SU : DNA sequence alignment, At et Vv : alignment of *A. thaliana* and *V. vinifera* peptides on the deduced peptides from the BAC clone sequences.

Figure 7.4 p.155 : Distribution of the different types of elements composing the genomic sequences of the 6 sequenced clones. A kb scale is associated to each clone. From top to bottom are represented the positions of : Cg ordered according to BES (white) and Cg randomly ordered (gray), regions masked by ReapeatMasker (red), repeated regions in common between the 6 clones (orange) and regions in common with the 60A10 clone (yellow), SU (purple), and homologues genes of *A. thaliana* (green) and *V.vinifera* (blue). Notation used for genes correspond to the abbreviation presented in the table 7.3. Continuation p.156 and p.157.

Figure 7.4 p.156 : Continuation – Distribution of the different types of elements composing the genomic sequence of the 6 sequenced clones.

Figure 7.4 p.157 : Continuation – Distribution of the different types of elements composing the genomic sequence of the 6 sequenced clones. For the Si27 clone, gene prediction examples (in pink : from the lightest to the darkest shade : prediction with Fgenesh, prediction with GenScan, both with *A. thaliana* as the reference organism, followed by predictions with Fgenesh and *V. vinifera* as the reference organism).

Figure 7.5 : Distribution of the genes identified by Blastx : (a) on chromosome 1 to 5 of *A. thaliana* or the chloroplastic (C) and mitochondrial (M) genomes, and (b) on the chromosomes 1 to 19 of *V. vinifera*, according to the different applied criteria of selection.

Table 7.3 : Identified genes on the genomic sequences positioned near the S-locus (Blastx realised on peptides library of *A. thaliana* and *V. vinifera*). Relative gene position on the different clones (Pos), the number of identified SU on the same position (SUtot), and the number of SU showing a homology with the same gene as the genomic sequence (SU), are indicated for each gene. Chr : chromosome on which each gene is present. Hits : number of results obtained per position. Δ : Δ e-value, ratio of the e-values of the first and second hit, except for (a) : e-value of the second hit. Alias : abbreviation used in figure 7.4. Cat : category in which each gene was classified in relation to the chicory sequence : homologous gene (H) or potentially orthologous gene (pO). A tBlastn search was performed for some pO genes : the

best homology was obtained either for the identified SU on the same position (pO^{*}), or for a different SU (pO⁻). Precisions on *A. thaliana* genes' functions can be found in the annexe (p.223). (b) : different positions corresponding to a same gene for *A. thaliana* or *V. vinifera*.

Figure 7.6 : Phylogenetic tree of the Dicotyledonous species from which sequence data have been used, and the position of chicory, lettuce and sunflower. When sequencing was completed, the type of assembling is indicated : * indicates the number of chromosomes or linkage groups obtained, # the number of scaffolds. The polyploidization events recently inferred (●) and of hexaploidization of the more recent common ancestor (▲) are also indicated (after Van de Peer *et al.* (2009) et Barker *et al.* (2008)).

Figure 7.7 : Distribution of the genes pO and H(*) of chromosome 18 of *V. vinifera* and comparison with their distribution on the clones of the S-locus region. Each number represents one gene (*cf.* position in the table 7.3). A color code was defined for each clone or overlapping group of clone, only the genes present on the chromosome 18 being highlighted on the physical map of the chicory. The Si27, Si37 and Si52 clones were gathered to form only one sequence. The regions of clones blue highlighted correspond to the Cg that were positioned and oriented.

Figure 7.8 : Distribution of the pO and H genes identified from the chicory's SU on chromosomes, linkage groups, or scaffolds of *M. guttatus*, *S. lycopersicum*, *P. trichocarpa* and *M. truncatula*. Chicory's clones are aligned one above the other (physical map on the left). Si27, Si37 and Si52 clones are united to form only a single sequence. On the physical map, each number corresponds to one gene (*cf.* position on the table 7.3) and only the SU associated genes are highlighted, a color code being specified for each clone or group of overlapping clones. (a) pO gene, (b) H gene, (c) gene corresponding to the second best Blast hit. The clone regions highlighted in blue correspond to the Cg that were positioned and oriented.

Discussion

Table 8.1 : Genomic composition in repeated elements for different species. * Estimation of the average distance between SSR (kb). GSS : Genome Survey Sequence, g : genome, chr : chromosome.

Figure 8.1 : Schematic representation of the position of gene blocks (A to D), identified on the chicory BAC clones, on the chromosomes and scaffolds of *V. vinifera*, *M. guttatus*, *S. lycopersicum*, *P. trichocarpa* et *M. truncatula*. Size of the blocks and distances between them are represented according to the scales given at the bottom of the figure. When only a single gene is associated to a block, it is represented by a colored rectangle without letter. Distances between the blocks on the physical map of chicory were estimated by considering a homogenous Mb/cM ratio in the TACG-EU12G02 interval.

Abstract

Self incompatibility (SI) is one of the most important strategies to prevent selfing in hermaphrodite plants. This mechanism is generally under the control of a single multi-allelic locus, the S-locus. Pollen and pistil determinants involved in the incompatibility reaction have been identified in some Angiosperm families. In the Asteraceae family, determinants of SI are still unknown. Chicory (*Cichorium intybus* L.), which belongs to this family, is a sporophytic self incompatible species. To identify determinants of chicory SI, a positional cloning strategy has been implemented.

The chicory S-locus has been assigned to one end of linkage group 2. To increase the density of molecular markers in the S-locus region and to obtain a high density map, a marker assisted BSA (Bulk Segregant Analysis) approach was developed. Five new AFLP markers were obtained, including one co-segregating with the S-locus. Furthermore, 2 markers turned out to be codominant after they were transformed in SCAR markers. In parallel, a high-resolution map was produced. For that goal, a large population of individuals ($n = 2.824$) segregating for different S-alleles was generated. Due to a difference of recombination frequency between male and female meiosis (heterochiasmy), 2 consensus maps were constructed. The S-locus was positioned in an interval of 0,6 cM or 1,24 cM considering female or male meiosis, respectively.

At the same time, 2 BAC libraries of chicory were produced in collaboration with the Centre National de Ressources Génomiques Végétales. These libraries, covering 12 times the chicory haploid genome, allowed us to start chromosome walking. Three contigs were built around the S-locus. The orientation with respect to the S-locus of one of them could be determined. Moreover, 6 BAC clones representing these contigs were sequenced and analyzed. In this region, corresponding to 546 kb, we relate the physical distance and the genetic distance to a ratio of 0,5 Mb/cM for male meiosis and 2,2 Mb/cM for female meiosis. Repeated element composition of the sequences comprised 5.71% of interspersed elements, 1.74% of SSR, 2.75% of low complexity sequences, and 10.45% of repeated sequences without homology. About thirty genes were identified upon comparisons with *Cichorium sp.* EST libraries and *Arabidopsis thaliana* and *Vitis vinifera* peptides libraries. A certain degree of conservation of microsynteny was observed between these genes and their homologues in five Dicotylédon species (*Vitis vinifera*, *Mimulus guttatus*, *Solanum lycopersicum*, *Populus trichocarpa* and *Medicago truncatula*). Together, these data open the way for new approaches that could accelerate chromosome walking towards the S-locus in chicory.

Keywords : Sporophytic self-incompatibility , positional cloning, S-locus, *Cichorium intybus* L., Asteraceae.

Résumé

L'autoincompatibilité (AI) est la stratégie la plus répandue pour empêcher l'autofécondation chez les plantes hermaphrodites. Les mécanismes associés sont généralement sous le contrôle d'un locus multiallélique, le locus S. Les déterminants pollen et pistil de la réaction d'incompatibilité ont été identifiés dans quelques familles d'Angiospermes. Dans la famille des Asteraceae, les déterminants de l'AI sont encore inconnus. La chicorée (*Cichorium intybus* L.), qui appartient à cette famille, présente un système d'autoincompatible sporophytique. Afin d'identifier les déterminants responsables de l'AI de cette espèce, une stratégie de clonage positionnel du locus S a été mise en place.

Le locus S de la chicorée avait préalablement été cartographié à l'extrémité du groupe de liaison 2. A partir de cette carte, une approche BSA (Bulk Segregant Analysis) assistée par marqueurs a été développée afin de densifier la région du locus S en marqueurs moléculaires et d'obtenir une carte haute densité. Cinq nouveaux marqueurs AFLP ont été identifiés, dont un coségrégant avec le locus S. De plus, 2 marqueurs se sont avérés codominants, une fois transformés en marqueurs SCAR. En parallèle, une carte haute résolution de la région du locus S a été établie. Pour cela, une importante population d'individus (n = 2824), en ségrégation pour différents allèles S, a été générée. En raison d'une différence de fréquence de recombinaison entre les méioses mâles et femelles, 2 cartes consensus ont été construites. Le locus S a été positionné dans un intervalle de 0,6 cM pour les méioses femelles et de 1,24 cM pour les méioses mâles.

Dans le même temps, 2 banques BAC de chicorée ont été construites en collaboration avec le Centre National de Ressources Génomiques Végétales. Ces banques, couvrant 12 fois le génome haploïde de la chicorée, ont permis d'initier la marche chromosomique vers le locus S. Trois contigs de clones BAC ont été construits de part et d'autre du locus S. L'orientation de l'un d'entre eux par rapport à la position du locus S a été définie. De plus, 6 clones BAC représentatifs des contigs, et couvrant un total de 546 kb, ont été séquencés et analysés. Dans cette région, le rapport distance physique/distance génétique a été estimé à 0,5 Mb/cM pour les méioses mâles et à 2,2 Mb/cM pour les méioses femelles. La composition des séquences en éléments répétés comprend 5,71% d'éléments mobiles, 1,74% de SSR, 2,75% de régions de faible complexité et 10,45% de régions répétées sans homologies. Une trentaine de gènes ont été identifiées à partir des banques d'EST de *Cichorium sp.* et de peptides d'*A. thaliana* et de *V. vinifera*. Une relative conservation de la microsyténie a été observée entre ces gènes et leurs homologues chez 5 espèces Dicotylédones (*Vitis vinifera*, *Mimulus guttatus*, *Solanum lycopersicum*, *Populus trichocarpa* et *Medicago truncatula*). L'ensemble de ces données ouvre la voie à de nouvelles approches qui pourraient accélérer la marche chromosomique vers le locus S de la chicorée.

Mots clefs : Autoincompatibilité sporophytique, clonage positionnel, locus S, *Cichorium intybus* L., Asteraceae.