

Université Lille 1, Sciences et Technologies

UMR1281 Stress Abiotiques et Différenciation des végétaux cultivés

## Thèse

Présentée pour l'obtention du diplôme de

**Docteur en Sciences de l'Université de Lille 1**

Stratégies d'exploitation des fonctions biologiques

**Analyse génétique de l'embryogenèse somatique  
chez la chicorée (*Cichorium intybus* L.) :  
Cartographie des QTL et gènes candidats**

Par **Aline CLABAUT**

Soutenue le 24 Juin 2009, devant la commission d'examen :

Tom Gerats	Professeur, Université de Nijmegen, Pays Bas	Rapporteur
Caroline Hartmann	Professeur, Université de Paris Diderot, France	Rapporteur
Kim Boutilier	Ingénieur de recherche, Wageningen UR, Pays-Bas	Examineur
Jean-Louis Hilbert	Professeur, USTL, France	Examineur
Theo Hendriks	Professeur, USTL, France	Directeur de thèse

## *Remerciements*

Tout d'abord, je tiens à remercier le professeur Jean-Louis Hilbert, directeur du laboratoire SADV, de m'avoir accueillie au sein de son laboratoire pendant ces années d'études et pour les corrections apportées au manuscrit.

Je remercie Tom Gerats, Caroline Hartmann et Kim Boutilier d'avoir accepté de faire partie du jury.

Je tiens à exprimer toute ma gratitude à Theo Hendriks, responsable de ce travail, avec qui j'ai pu échanger un grand nombre d'idées scientifiques et qui a été pour moi un soutien constant dans la réalisation de mon travail. Je tiens également à le remercier pour m'avoir fait confiance mais également pour m'avoir soutenu dans mes démarches et d'être resté aussi présent jusqu'à la fin de la thèse (même les week-end).

Je remercie toute " l'équipe carto " avec Marie-Christine Quillet pour m'avoir initiée à l'art de la cartographie génétique, pour son encadrement en culture *in vitro* et pour sa participation aux corrections du manuscrit. Pour sa joie de vivre, Christelle, avec qui j'ai débuté mes premières manip. Monika, pour m'avoir initiée à l'écologie et son soutien. Et enfin Thierry, pour toutes les discussions scientifiques très enrichissantes mais également sa gentillesse.

Je tiens également à les remercier pour l'ensemble du travail qu'ils ont réalisé pour la construction de la carte génétique K28\*K59 sans laquelle je n'aurais pu mener à bien mon projet.

Merci à Brigitte Huss pour son travail sur le clonage du gène *STM* qui m'a permis de le cartographier sur la carte génétique K28\*K59.

Merci à Nadja et Marie-Christine Dendieviel pour leur aide apporté dans le maintien et la multiplication des plantules *in vitro*.

Merci également aux stagiaires Laurence et Camille pour leur travail accompli dans la caractérisation du potentiel organogènes des explants foliaires.

Je voudrais également remercier l'ensemble du personnel du laboratoire, qui a d'une manière ou d'une autre contribué à mon travail de thèse. Une petite note particulière à Philippe qui m'a permis de faire mes premiers pas dans le labo.

Je tiens à remercier les étudiants, Meriem, Lucy et Ildephonse pour tous ces petits moments de détente qui étaient les bienvenus, mais aussi mes amis Christelle D, Raphaëlle, Sylvain, David, mes amis de Marquise et mes compagnons du TER-GV. Merci à Estelle D. pour les corrections du manuscrit "made in Canada".

Je remercie également le laboratoire LR2B de Boulogne sur Mer où j'effectue actuellement mon poste d'ATER pour m'avoir aménagé du temps libre pour terminer la rédaction du manuscrit de thèse.

J'exprime ma profonde reconnaissance à mes parents, mon frère, ma sœur et ma grand-mère qui m'ont permis d'être ce que je suis et pour leur soutien permanent tout au long de mes études (je sais, je n'ai pas choisi les plus courtes !).

Et enfin, je remercie tout particulièrement, Jonathan, pour sa très grande patience, pour son soutien incontestable et pour tous les bons petits plats qu'il m'a préparés pour tenir le coup tout au long de la thèse.

## Sommaire

<b>Analyse Bibliographique</b>	<b>p11</b>
<b>I. L'embryogenèse somatique chez les végétaux</b>	<b>p12</b>
1. Caractéristiques générales de l'embryogenèse somatique	p12
2. Comment les cellules changent-elles d'état ?	p16
2.1. Le remodelage de la chromatine	p17
2.2. La position cellulaire	p18
2.3. L'action des hormones	p19
3. Les différentes étapes de l'embryogenèse somatique	p20
3.1. Les phases précoces de l'embryogenèse somatique	p20
3.1.1. Induction et réactivation cellulaire	p21
3.1.2. La première division des cellules embryogènes	p22
3.1.3. Les phases tardives	p22
<b>II. Méthodes d'études des gènes impliqués au cours des phases précoces l'embryogenèse somatique</b>	<b>p23</b>
1. Analyse des profils protéiques	p27
2. L'analyse génomique	p29
2.1. Approche par criblage différentiel	p30
2.1.1. Les marqueurs tardifs	p30
2.1.2. Les marqueurs précoces	p32
2.2. Technique de microarray	p35
3. L'approche génétique	p37
3.1. L'analyse par mutagenèse	p37
3.2. Analyse de la variation naturelle	p39
3.2.1. Modèles génétiques 'simples'	p39
3.2.2. Analyses QTL	p40
<b>III. La chicorée : une plante modèle de l'embryogenèse somatique chez les dicotylédones</b>	<b>p43</b>
1. La chicorée : classification botanique	p43
2. Historique de la culture de chicorée	p43
3. L'embryogenèse somatique chez la chicorée	p44
4. Objectifs de la thèse	p44

4.1. Obtention d'une descendance K28*K59 en ségrégation pour l'embryogenèse somatique	p45
4.2. L'évaluation du potentiel embryogène des individus de la descendance K28*K59	p46
4.3. La construction d'une carte génétique F1'a	p46
4.4. La détection de polymorphisme au sein des gènes candidats	p48
4.5. L'analyse QTL	p50
4.5.1. Méthodes statistiques de détection des QTL	p50
a. Méthodes statistiques simples : analyse de variance	p50
b. Méthode de cartographie d'intervalles	p50

## **Chapitre I : Analyse des données phénotypiques des individus pour leur capacité liée à l'embryogenèse somatique et à l'organogenèse**

**p53**

### **I. Introduction**

**p54**

### **II. Matériels et méthodes**

**p58**

#### **1. Matériel végétal**

**p58**

##### **1.1. Génération des plantules de la descendance F1'**

**p58**

##### **1.2. Multiplication des individus par embryogenèse somatique et organogenèse**

**p58**

##### **1.3. Tests et notation du potentiel embryogène**

**p61**

##### **1.4. Tests et notation du potentiel organogène**

**p61**

#### **2. Analyse des données**

**p62**

### **III. Résultats**

**p64**

#### **1. Evaluation du potentiel embryogène des génotypes parentaux**

**p64**

#### **2. Evaluation du potentiel embryogène des individus des sous-descendances F1'a et F1'b**

**p67**

##### **2.1. Comparaison des sous- descendances F1'a et F1'b**

**p68**

#### **3. Analyse des distributions**

**p71**

##### **3.1. Transformation des valeurs des caractères**

**p71**

##### **3.2. Optimisation de la valeur moyenne des caractères étudiés**

**p74**

#### **4. Variabilité génotypique et héritabilité**

**p74**

##### **4.1. Variabilité génotypique**

**p74**

##### **4.2. Héritabilité au sens large ( $h_L^2$ )**

**p75**

5. Evaluation du potentiel organogène à partir des fragments de pétioles	<b>p77</b>
5.1. Evaluation du potentiel organogène des génotypes parentaux	<b>p77</b>
5.2. Evaluation du potentiel organogène des génotypes dans la sous- descendance F1'b	<b>p77</b>
5.3. Calcul de l'héritabilité et effet génotypique	<b>p80</b>
 IV. Discussion	 <b>p82</b>
 <b>Chapitre II : Construction d'une carte génétique F1'a+b pour la détection de QTL</b>	 <b>p86</b>
 I. Introduction	 <b>p87</b>
 II. Matériels et méthodes	 <b>p88</b>
1. Extraction d'ADN à partir de plantes cultivées en serre	<b>p88</b>
2. Détection de polymorphisme au sein des séquences	<b>p89</b>
2.1. Les microsatellites ou SSR	<b>p89</b>
2.1.1. Définition des amorces	<b>p89</b>
2.1.2. Réaction PCR	<b>p89</b>
2.2. Les marqueurs AFLP	<b>p90</b>
2.2.1. Digestion des ADN	<b>p91</b>
2.2.2. Ligature des adaptateurs	<b>p91</b>
2.2.3. Pré-amplification	<b>p91</b>
2.2.4. Amplification sélective	<b>p92</b>
2.3. Les séquences EST	<b>p93</b>
2.3.1. Définition des amorces	<b>p94</b>
2.3.2. Réaction PCR	<b>p100</b>
3. Analyse des produits d'amplification	<b>p100</b>
3.1. Electrophorèse sur gel d'agarose	<b>p100</b>
3.2. Séparation des fragments d'amplification sur électrophorèse capillaire	<b>p101</b>
3.2.1. Les marqueurs SSR et AFLP	<b>p101</b>
3.2.2. Les marqueurs STS (incluant les séquences EST)	<b>p106</b>
3.3. Electrophorèse sur gel d'acrylamide	<b>p106</b>
4. Construction des cartes génétiques de liaison	<b>p106</b>

III. Résultats	
1. Analyse des marqueurs cartographiés dans la sous-descendance F1'a	<b>p108</b>
1.1. Les marqueurs microsatellites ou SSR	<b>p108</b>
1.2. Les marqueurs AFLP	<b>p109</b>
1.3. Les séquences ESTs	<b>p111</b>
2. Construction d'une carte génétique F1'a+b pour l'analyse QTL liée à l'embryogenèse somatique et à l'organogenèse à partir des génotypes des sous-descendances F1'a et F1'b	<b>p117</b>
3. Distorsion de ségrégation des marqueurs de la sous-descendance F1'b	<b>p121</b>
IV. Discussion	<b>p126</b>
<b>Chapitre III : Détections de QTL liées à la régénération <i>in vitro</i></b>	<b>p132</b>
I. Introduction	<b>p133</b>
II. Matériels et méthodes	<b>p134</b>
1. Détection des QTL	<b>p136</b>
2. Calcul des interactions	<b>p138</b>
III. Résultats	<b>p138</b>
1. Analyse des QTL liée à la formation des plantules, des structures chlorophylliennes unipolaires et aux évènements totaux	<b>p139</b>
1.1. Analyse des QTL dans la sous-descendance F1'a	<b>p139</b>
1.1.1. Détections des QTL liées à la formation des plantules	<b>p139</b>
1.1.2. Détections des QTL liées à la formation des structures chlorophylliennes unipolaires	<b>p142</b>
1.1.3. Détections des QTL liées à la formation des évènements totaux	<b>p143</b>
1.2. Analyse des QTL dans la sous-descendance F1'b	<b>p144</b>
1.2.1. Détections des QTL liées à la formation des plantules	<b>p148</b>
1.2.2. Détections des QTL liées à la formation des structures chlorophylliennes unipolaires	<b>p149</b>
1.2.3. Détections des QTL liées aux évènements totaux	<b>p149</b>
1.3. Analyse des QTL dans la descendance F'a+b	<b>p153</b>

2. Détections des QTL liées à la formation de cals et de bourgeons à partir d'explants foliaires	p157
3. Effets d'interactions	p158
IV. Discussion	p159
<b>Chapitre IV : Des gènes candidats qui co-localisent avec les QTL</b>	<b>p168</b>
I. Introduction	p169
II. Etude des gènes à fonction connue liés aux QTL détectés	p169
1. Protéines impliquées dans la biogenèse des ribosomes	p172
2. Les gènes en lien avec les divisions cellulaires	p173
3. Les gènes impliqués dans le transport	p175
4. Les protéines impliquées dans les phases tardives de l'embryogenèse (LEA, Late Embryogenesis Abundant)	p177
III. Etude des gènes <i>SHOOTMERISTEMLESS</i> et <i>ARGONAUTE</i>	p177
IV. Conclusion	p186
<b>Chapitre V : Discussion générale et perspectives</b>	<b>p189</b>
<b>Références bibliographiques</b>	<b>p202</b>
<b>Annexes</b>	<b>p235</b>
<b>Annexe 1:</b> Article soumis à <i>Molecular Breeding: A consensus genetic map for chicory</i> ( <i>Cichorium intybus L.</i> )	<b>p236</b>
<b>Annexe 2 :</b> Milieux de culture utilisés pour la multiplication clonale des génotypes de la descendance F1', les tests d'embryogenèse somatique et d'organogenèse	<b>p276</b>
<b>Annexe 3:</b> EST pour lesquels des SNP ( <i>Single Nucleotide Polymorphisme</i> ) a pu être détectés	<b>p277</b>
<b>Annexe 4:</b> EST pour lesquels des amorces ont été définies autour du ou des régions introniques	<b>p279</b>

<b>Annexe 5:</b> EST pour lesquels des alignements des séquences de chicorée avec les séquences laitue <i>sp.</i> ont été réalisés	<b>p285</b>
<b>Annexe 6:</b> EST pour lesquels la prédiction de jonctions « intron-exon » a été réalisée	<b>p286</b>
<b>Annexe 7:</b> EST pour lesquels des amorces ont été définies au niveau des exons	<b>p287</b>
<b>Annexe 8:</b> Liste des EST cartographiés dans la carte génétique F1'a	<b>p289</b>
<b>Annexe 9:</b> Couples d'amorces définies autour des InDel afin de détecter du polymorphisme au sein du gène <i>CiSTM</i>	<b>p291</b>
<b>Annexe 10:</b> Analyse IM (Interval Mapping). Identification des QTL liés à la formation des plantules (PL), des structures chlorophylliennes unipolaires (SH) et des événements totaux (ET) dans les sous-descendances F1'a, F1'b et la descendance F1'a+b	<b>p293</b>
<b>Annexe 11 :</b> Courbes LOD score obtenues après analyse IM des données plantules (PL), structures chlorophylliennes unipolaires (SH) et événements totaux (ET) dans la sous-descendance F1'a	<b>p295</b>
<b>Annexe 12 :</b> Courbes LOD score obtenues après analyse IM des données plantules (PL), structures chlorophylliennes unipolaires (SH) et événements totaux (ET) dans la sous-descendance F1'b	<b>p297</b>
<b>Annexe 13 :</b> Courbes LOD score obtenues après analyse IM des données plantules (PL), structures chlorophylliennes unipolaires (SH) et événements totaux (ET) dans la descendance F1'a+b	<b>p300</b>
<b>Annexe 14:</b> Effets alléliques observés pour les marqueurs liés significativement aux QTL impliqués dans la formation des plantules (PL), des structures chlorophylliennes unipolaires (SH) et des événements totaux (ET) pour chaque paramètre d'analyse (PL <sub>50</sub> , PL <sub>90</sub> , SH <sub>50</sub> , SH <sub>90</sub> , ET <sub>50</sub> , ET <sub>90</sub> )	<b>p303</b>
<b>Annexe 15:</b> Analyse IM (Interval Mapping). Identification des QTL liés à la formation des cals (Cal) et des bourgeons (Bg) dans la sous-descendance F1'b	<b>p305</b>
<b>Annexe 16 :</b> Courbes LOD score obtenues après analyse MQM des données cals (Cal) et bourgeons (Bg) dans la sous-descendance F1'b	<b>p306</b>
<b>Annexe 17 :</b> Effets additifs des QTL majeurs (QTL2, QTL6 et QTL8) liés à la formation des plantules (PL90)	<b>p309</b>
<b>Annexe 18 :</b> Alignements des séquences Argonaute d' <i>Arabidopsis</i> et la séquence chicorée (Ci. CSA1.6010)	<b>p313</b>

# Analyse Bibliographique

## I. L'embryogenèse somatique chez les végétaux

Chez les végétaux, les cellules somatiques ne sont pas différenciées à vie, elles peuvent sous certaines conditions retourner à l'état de totipotence et initier un processus de régénération à partir du développement d'embryons somatiques (l'embryogenèse somatique) et/ou du développement de bourgeons (l'organogenèse). Ce phénomène appelé transdifférenciation ou reprogrammation est connu depuis plusieurs années (Reghavan *et al.*, 1986 ; Thelen *et al.*, 1989, Beresford, 1990). Cependant, certains aspects restent encore inconnus. A l'heure actuelle, plusieurs revues s'interrogent encore sur les mécanismes soulignant ce processus "the initiation phase of somatic embryogenesis: what we know and what we don't" (Fehér *et al.*, 2008); "How cells dedifferentiate: a lesson from plants" (Grafi *et al.*, 2003); "How does a single cell become a whole plant" (Kennedy and Norman, 2005) "open minded cells: how cells can change fate" (Costa *et al.*, 2006a).

L'étude des événements moléculaires et physiologiques contrôlant l'embryogenèse somatique est d'un grand intérêt. En effet, ce processus pourrait fournir des connaissances importantes sur l'acquisition de la totipotence des cellules déjà différenciées mais également permettrait de comprendre comment les cellules changent d'état.

D'un point de vue appliqué, l'embryogenèse somatique permet de disposer de matériel végétal important et à n'importe quelle période de l'année. De plus, c'est une alternative rapide pour les sélectionneurs afin de maintenir les génotypes élites par exemple chez les espèces allogames telles que la chicorée. De même, elle est un moyen de multiplier les espèces à cycle de vie de reproduction très long comme les arbres forestiers.

### 1. Caractéristiques générales de l'embryogenèse somatique

Bien qu'il existe certains cas de régénération spontanée ou *in vivo* des embryons somatiques comme le Kalanchoe, *Bryophyllum* (Yarbrough, 1932) ou les fougères du genre *Malaxis* (Taylor, 1967), chez la majorité des espèces végétales l'embryogenèse somatique est induite *in vitro*. La régénération d'embryons

somatiques à partir de cellules somatiques se déroule selon trois phases principales : acquisition de la compétence, induction et différenciation (Encadré 1).

Tout d'abord, des cellules déjà différenciées suivent un processus conduisant à l'obtention de cellules compétentes (Fehér, 2008). La compétence a été définie comme l'aptitude à recevoir et/ou à répondre à un signal de développement. La compétence est alors considérée comme un état physiologique, cellulaire et moléculaire nécessaire à la préparation de l'étape suivante du processus morphogène.

Lors de la phase d'induction, sous l'influence de stimuli externes, tels que des régulateurs de croissance et/ou différents stress, les cellules compétentes deviennent déterminées à former un embryon, on parle alors de cellules embryogènes. Le terme de cellule embryogène est restreint aux cellules qui ont achevé la transition de l'état de cellule somatique à l'état dans lequel l'application de stimuli externes n'est plus nécessaire pour produire un embryon. Cependant, on trouve des cellules compétentes qui ont commencé à devenir embryogènes mais exigent encore l'application de stimuli externes.

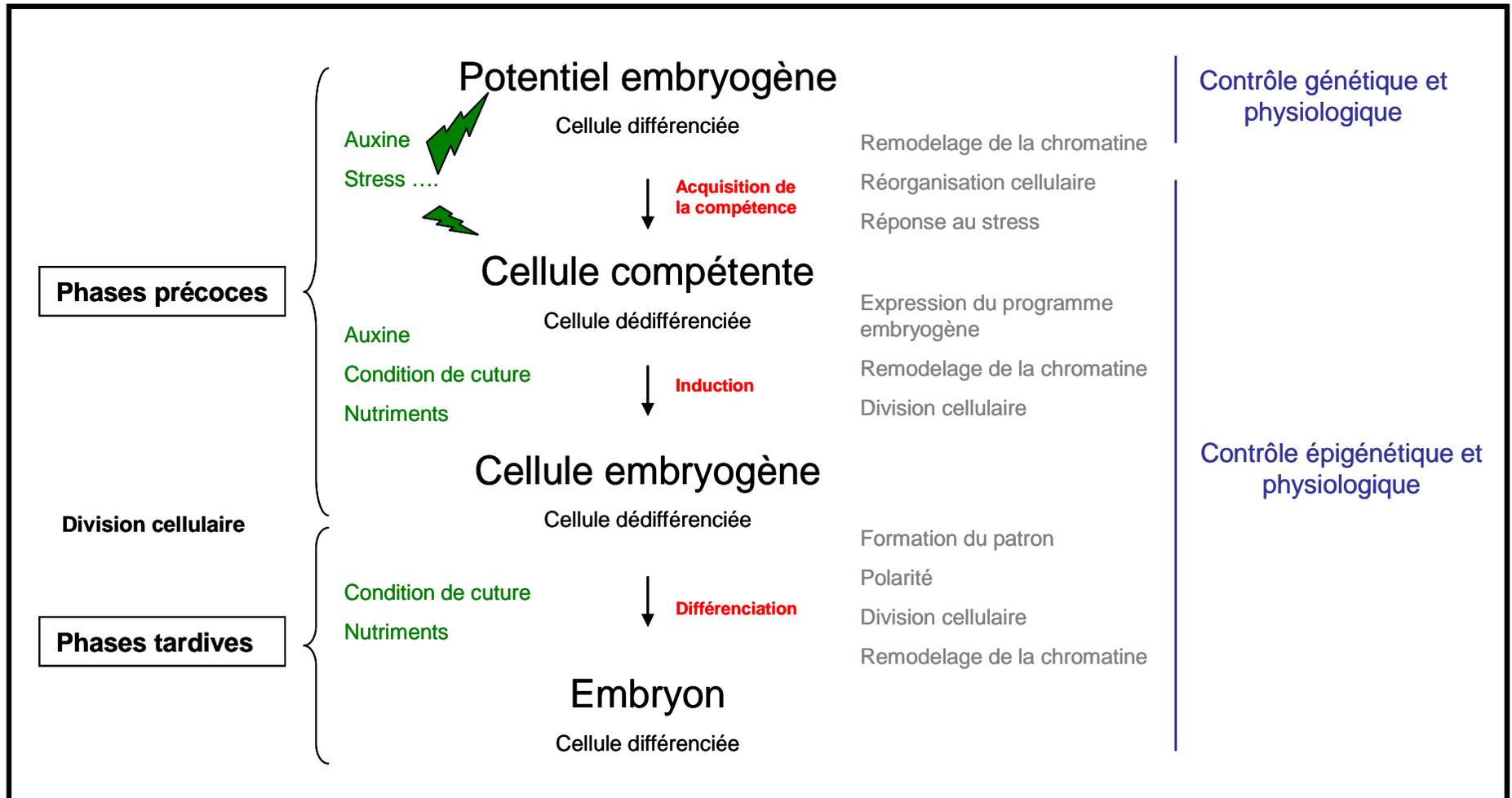
Un nouveau programme de différenciation conduisant à la formation d'un embryon somatique est ensuite initié.

Il est important de rappeler que le génotype et l'explant initial mais également la nature du stimulus externe représentent trois caractères à prendre en compte pour initier un programme d'embryogenèse somatique.

L'obtention d'embryons somatiques peut se faire selon deux modes, le mode direct ou indirect. Lorsque l'embryogenèse est de type direct, la cellule somatique subit une dédifférenciation puis une redifférenciation au sein du tissu dont elle est issue. C'est le cas de la chicorée où les embryons somatiques sont obtenus à partir de différents types d'explants sans passage par une phase de callogénèse (Dubois *et al.*, 1990).

L'embryogenèse somatique de type indirect passe par un état de cal formant une masse de cellules à caractère indifférencié dont certaines vont donner des embryons. C'est le cas des suspensions cellulaires chez la carotte (de Vries *et al.*, 1988)

Par ailleurs, les embryons somatiques peuvent avoir deux origines différentes. Ils sont soit issus d'une seule cellule comme chez la carotte et la chicorée : l'embryogenèse est d'origine unicellulaire (Dubois *et al.*, 1990 ; Toonen *et al.*, 1994). Soit les embryons sont issus d'un ensemble de cellules formant un complexe pro-embryogène comme chez le trèfle (Maheswaran et Williams, 1985) : l'embryogenèse est alors d'origine pluricellulaire. Dans ce cas, les cellules du complexe pro-embryogène sont pluripotentes, puisque chaque cellule va se différencier en un type cellulaire spécifique.



**Encadré 1** : Induction de l'embryogenèse somatique à partir de cellules végétales différenciées

## 2. Comment les cellules changent-elles d'état ?

Il a longtemps été admis que la totipotence était une propriété spécifique des végétaux. En effet, ce processus était considéré comme irréversible chez les animaux supérieurs jusqu'à la deuxième partie du 20<sup>ème</sup> siècle. C'est à ce moment-là que le dogme d'une différenciation irréversible est tombé avec les premières expériences de clonage animal. Le transfert du noyau d'une cellule d'un organisme différencié, dans un œuf énuclé, conduit dans un certain nombre de cas à un individu viable (Briggs and King, 1952). Ces expériences ont également montré que les cellules vont maintenir leur information génétique au cours de la différenciation et qu'elles peuvent reprogrammer ces informations et acquérir une nouvelle identité. En 2006, des chercheurs japonais ont même réussi à reprogrammer des cellules différenciées de la peau en cellules souches pluripotentes induites ou iPS (induced pluripotent stem cell) et ceci grâce à quatre facteurs de transcription. Oct-4<sup>1</sup> et Sox-2 interviennent dans le maintien de l'auto-renouvellement des cellules souches embryogènes indifférenciées et, c-Myc et Klf-4 sont deux facteurs oncogènes (Takahashi *et al.*, 2006). Contrairement aux cellules somatiques, les cellules souches pluripotentes ne sont pas capables de former un individu entier mais ont la possibilité de se différencier en quelques types cellulaires spécifiques. Il est néanmoins avéré que deux des quatre gènes impliqués dans la reprogrammation pouvaient conduire à des cancers, ne permettant pas d'envisager une utilisation en thérapie génique. Très récemment, une équipe allemande a montré que l'introduction du seul facteur Oct-4 était suffisante pour reprogrammer des cellules déjà différenciées en iPS (Kim *et al.*, 2009).

Lors du changement d'état, plusieurs événements caractéristiques peuvent être observés comme le remodelage de la chromatine. Ces événements semblent être influencés par la position cellulaire mais aussi par l'action d'inducteurs comme les hormones.

---

<sup>1</sup> par convention, nous noterons en majuscules italiques le nom des gènes, en majuscules droites le nom des protéines et en minuscules italiques le nom des mutants perte de fonction

## 2.1. Le remodelage de la chromatine

La dédifférenciation cellulaire chez les animaux comme chez les plantes est caractérisée par un changement remarquable du profil d'expression des gènes. Le programme d'expression des gènes est gouverné par une portion du génome qui est transcrite (euchromatine) et une proportion de gènes réprimés (hétérochromatine). Lors de la reprogrammation cellulaire, il est vraisemblable qu'une nouvelle balance doit être établie entre l'euchromatine et l'hétérochromatine.

La liaison entre la dédifférenciation et la réorganisation de la chromatine a été démontrée chez les végétaux. Des changements similaires dans l'organisation de la chromatine ont été observés à partir de protoplastes (Grafi, 2004). Chez les protoplastes, la perte de l'état différencié est accompagnée d'un changement dans la distribution et l'organisation de l'hétérochromatine avec une décondensation de la chromatine et la fragmentation du nucléole.

De même, des expériences de reprogrammation à partir d'hétérokaryons chez l'humain et la souris ont montré que ce processus est caractérisé par un grossissement du noyau et une réorganisation de la chromatine (Terranova *et al.*, 2006).

Suite à ces observations, l'implication, dans le remodelage de la chromatine, d'un nombre important de facteurs et de complexes protéiques a été constatée.

On peut évoquer le rôle des protéines du groupe Polycomb (PcG) qui sont impliquées dans le changement d'état cellulaire aussi bien chez les plantes que chez les animaux. Les protéines PcG sont responsables du maintien d'un état réprimé de gènes cibles. Ces gènes fixés par les protéines du PcG sont majoritairement des facteurs de transcription impliqués dans des voies de régulation majeures contrôlant la différenciation, la prolifération et la morphogenèse. Plusieurs équipes ont analysé le profil de fixation des PcG dans les cellules souches embryonnaires humaines et murines et, ont pu montrer que les gènes cibles des protéines du PcG sont réprimés dans les cellules souches embryonnaires mais deviennent activés au cours de la différenciation (Boyer *et al.*, 2006 ; Lee *et al.*, 2006).

D'autres études ont montré que l'acquisition de la pluripotentialité implique une réorganisation de sous-domaines chromosomiques spécifiques (Avivi *et al.*, 2004). En effet, la décondensation chromatinienne apparaît être spécifique de certains domaines du génome, permettant ainsi l'activation de certains gènes

précédemment réprimés comme par exemple le gène *VIP1* (*VirE2-interacting protein*). Ce gène, identifié chez *Arabidopsis*, code pour une protéine nucléaire Zip interagissant avec des protéines de virulences bactériennes (VirE2) facilitant le transport d'ADN bactérien jusqu'au noyau des plantes et induisant la formation de tumeurs. Il est impliqué dans le maintien et l'acquisition de la pluripotentialité (Li *et al.*, 2002).

L'étude d'un mutant chez *Arabidopsis* a également permis de montrer l'implication du remodelage de la chromatine lors de l'embryogenèse somatique. Le mutant *pickle* présente un phénotype caractérisé par la génération spontanée d'embryons somatiques au niveau des racines (Ogas *et al.*, 1999). Ce mutant présente un gène muté qui code pour un facteur de remodelage de la chromatine codant pour une histone acétylée. Ce facteur réprime certains gènes liés à l'embryogenèse comme le gène *LEC1* (*LEAFY COTYLEDON1*) dans les embryons somatiques (Li *et al.*, 2002).

L'identification des zones réprimées au cours du développement végétatif et activées au cours de l'embryogenèse somatique pourrait être une source d'informations importante permettant ainsi la recherche des gènes clés au cours de la reprogrammation.

## 2.2. La position cellulaire

L'information apportée par la position d'une cellule semble un facteur important lors de la programmation mais aussi lors de la reprogrammation cellulaire et l'organisation de la chromatine. Par exemple, chez *Arabidopsis*, l'épiderme racinaire est composé de deux types cellulaires (les cellules épidermiques avec ou sans poils absorbants) organisés en alternance. Des modifications dans la position cellulaire sont accompagnées par un changement d'état cellulaire et un changement rapide de l'accessibilité d'une région du génome (40kb) couverte par le locus *GL2* (*GLABRA2*, codant pour un facteur de transcription à homéodomaine) (Costa *et al.*, 2006b). Ce gène joue un rôle dans le développement de trichomes chez *Arabidopsis*. Les auteurs supposent que la présence de protéines PcG ou de complexes protéiques similaires réprimerait la région du locus *GL2*. Ainsi, la régulation négative des protéines PcG pourrait réduire le nombre de gènes cibles réprimés et

augmenterait la plasticité cellulaire en facilitant l'accessibilité du génome à de nouveaux facteurs de transcription.

La communication cellulaire apparaît également être requise pour initier l'embryogénèse somatique (McCabe *et al.*, 1997). Des facteurs solubles, comme les AGP (arabinogalactanes protéines) ont été associés à la différenciation cellulaire pour aboutir à la formation des embryons (Windels, 2005). Il a aussi été montré que l'ajout d'endochitinases au milieu de culture a permis la formation d'embryons somatiques chez un mutant thermosensible (*ts11*) de carotte (De Jong *et al.*, 1992, 1995). Les gènes des endochitinases ne sont pas exprimés par les cellules ayant acquis la compétence cellulaire mais sont exprimés par d'autres types cellulaires, impliquant ainsi une communication entre cellules embryogènes et non embryogènes (Van Hengel *et al.*, 1998).

### 2.3. L'action des hormones

Plusieurs études ont été faites pour identifier les gènes clés qui gouvernent l'embryogénèse somatique en réponse à l'induction du signal. Les agents utilisés pour induire l'embryogénèse somatique *in vitro* sont multiples : les hormones, le stress, la pression osmotique... Après traitement à l'auxine (2,4D, acide 2,4-dichlorophénoxyacétique ; le plus communément utilisé), les cellules cultivées *in vitro* acquièrent leur compétence à former des embryons. L'obtention de ces embryons somatiques en culture *in vitro* a permis d'étudier les gènes qui gouvernent l'embryogénèse somatique par différentes stratégies et ainsi d'identifier des marqueurs spécifiques de ce processus. Nous discuterons plus tard dans ce chapitre des différentes stratégies utilisées.

Deux marqueurs moléculaires de la compétence cellulaire ont pu être identifiés chez la carotte. Un de ces gènes code pour un récepteur de type kinase (SERK ; Somatic Embryo Receptor-like Kinase) dont l'expression est corrélée avec l'acquisition de la compétence (Schmidt *et al.*, 1997). Nous rentrerons dans le détail de ce gène un peu plus loin dans l'analyse bibliographique. Un autre marqueur est également souvent cité comme marqueur de la compétence cellulaire, c'est un antigène pariétal reconnu par un anticorps monoclonal JIM8 (McCabe *et al.*, 1997). Cependant, ce résultat est basé sur un anticorps dirigé contre une culture de cellules

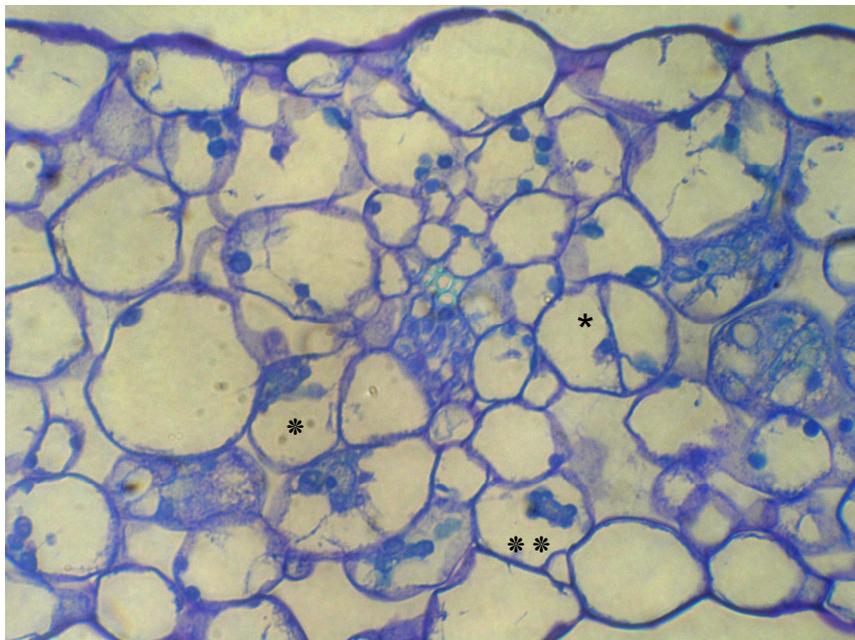
embryogènes uniquement. Le suivi microscopique de cellules par vidéo (cell tracking) n'a pas permis de mettre en évidence une corrélation entre la présence de cet antigène pariétal et la formation des embryons somatiques (Toonen *et al.*, 1996).

Les gènes clés responsables de l'induction n'ont pas encore été identifiés. Bien que l'auxine soit requise lors de l'induction, cette hormone seule est insuffisante pour initier les embryons somatiques. L'auxine semble être nécessaire pour créer un environnement approprié et d'autres facteurs, comme le stress par exemple, déclencheraient le programme embryogène (Fehér, 2006).

### 3. Les différentes étapes de l'embryogenèse somatique

#### 3.1. Les phases précoces de l'embryogenèse somatique

Les phases précoces de l'embryogenèse somatique sont initiées par l'induction du processus et se terminent par l'apparition des premières divisions embryonnaires.



**Figure 2 :** Observations cytologiques d'explants foliaires du génotype K59 (T. Hendriks, données personnelles).

Des cellules en cours de réactivation (\*), réactivées (\*\*\*) et en division (\*) sont observées chez le génotype K59.

### 3.1.1. Induction et réactivation cellulaire

Chez les plantes, l'induction de l'embryogenèse somatique peut être spontanée (*in vivo*) ou provoquée (*in vitro*) dépendante de l'application de facteurs endogènes ou exogènes. Cette étape est importante puisqu'elle est nécessaire à la réactivation des cellules compétentes induisant une réinitiation des divisions cellulaires.

L'action de phytohormones comme l'auxine (2,4 D, acide 2,4-dichlorophénoxyacétique ; ANA, acide naphthalène acétique) est nécessaire à l'obtention de cellules compétentes chez la carotte (Toonen *et al.*, 1994), la luzerne (Dos Santos *et al.*, 1983) et la chicorée (Dubois *et al.*, 1991). Cependant l'auxine n'est pas le seul facteur permettant la réinitiation des divisions cellulaires. D'autres phytohormones comme les cytokinines (Sagare *et al.*, 2000) et l'acide abscissique (Nishiwaki *et al.*, 2000) peuvent induire le passage de cellules épidermiques d'embryons immatures à celui de cellules embryogènes. Chez la chicorée, une cytokinine, le Zip (inositol phosphate), fournie à de faibles doses, est également nécessaire pour l'activation de cellules compétentes.

Par ailleurs, d'autres facteurs comme la qualité (D'Onofrio *et al.*, 1998) et l'intensité (Gaj, 2002) lumineuse mais aussi des facteurs de stress comme la température (Decout *et al.*, 1994) et les conditions osmotiques (Vasseur *et al.*, 1995) peuvent être déterminants.

Après l'induction, les cellules induites passent par de nombreuses modifications cytologiques pour arriver à un état dit de réactivation cellulaire. Chez la chicorée, cette phase de réactivation cellulaire a été particulièrement décrite chez l'hybride « 474 » (*Cichorium intybus* × *Cichorium endivia*). Les premières modifications observées sont la migration du noyau en position latérale et la migration de chloroplastes vers ce même noyau ce qui correspond à une cellule en cours de réactivation. Puis le noyau, entouré d'une couronne de chloroplastes, se déplace vers le centre de la cellule, la cellule est réactivée. Les cellules réactivées présentent un noyau très volumineux et dense, probablement dû à une augmentation de synthèse d'ADN, d'où part un réseau de microtubules en forme d'étoiles. Ces

multiples modifications ont pour rôle de préparer les cellules à entrer en division et ainsi à se re-différencier (Blervacq *et al.*, 1995). Chez la carotte, une fois les cellules entrées en division, aucun stimulus n'est nécessaire à la poursuite du processus (De Jong *et al.*, 1993).

### 3.1.2. La première division des cellules embryogènes

Lors de la reproduction sexuée, la première division du zygote est dans la plupart des cas asymétrique comme chez *Arabidopsis* et donne naissance à une petite cellule apicale qui formera le futur embryon et à une grande cellule basale qui deviendra le suspenseur (Cresti *et al.*, 1992). Au cours de l'embryogenèse somatique, la division asymétrique reste plus fréquemment observée que la division symétrique (Trigiano *et al.*, 1989). Chez la carotte, lors de la formation d'embryons somatiques à partir de cultures cellulaires, la première division peut-être asymétrique ou symétrique (Toonen *et al.*, 1994). Au contraire, la première division des embryons zygotiques est asymétrique. Chez la chicorée, seule la division symétrique est observée lors des premières divisions. De plus, le développement des embryons somatiques se faisant majoritairement autour des éléments vasculaires indique que cette première division est anticline par rapport à l'orientation des éléments vasculaires (Blervacq *et al.*, 1995).

### 3.1.3. Les phases tardives

Le développement de l'embryon somatique est similaire à celui de l'embryon zygotique et passera par des phases semblables de développement (Zimmerman, 1993) : les stades globulaires, cordiformes, torpilles et cotylédonaire. Au cours de ces phases tardives, on pourra observer un protoderme mais également un méristème racinaire aboutissant à la formation d'une radicule et un méristème caulinaire, initiant le développement des cotylédons, caractéristique d'une plantule.

## II. Méthodes d'étude des gènes impliqués au cours des phases de l'embryogenèse somatique

Différentes stratégies ont été utilisées pour détecter les gènes différentiellement exprimés au cours de l'embryogenèse somatique chez les végétaux. Des gènes et des protéines isolés lors de la formation d'embryons somatiques chez différentes espèces sont regroupés dans le tableau 1.

Les premières études ont porté sur l'analyse des profils protéiques obtenus en condition embryogène et non embryogène par électrophorèse afin de mettre en évidence les protéines spécifiques de ce processus. Par la suite, l'analyse de l'expression des transcrits (microarray, differential display) a été réalisée lors des phases de développement des embryons somatiques. Enfin, l'étude de mutants lors de l'embryogenèse zygotique mais également l'identification de locus liés à l'embryogenèse somatique par une analyse QTL ont permis de caractériser le programme génétique du développement des embryons somatiques.

Afin d'illustrer les différentes approches utilisées pour identifier les gènes liés au processus de l'embryogenèse somatique, quelques exemples vont être détaillés à partir de plantes modèles de l'embryogenèse somatique comme la carotte, *Arabidopsis* mais aussi la chicorée, plante modèle pour notre étude.

## Analyse Bibliographique

	Gène/Protéine	Technique de détection	Nature/Fonction	Espèce	Stade d'expression	Références
<b>Gènes de ménage</b>	H3-1/H3-11	Northern blot	histone	luzerne	--	Kapros <i>et al.</i> , 1992
	CEM1	crible différentiel de banque d'ADNc	facteur d'élongation 1 $\alpha$	carotte	stade globulaire	Kawahara <i>et al.</i> , 1992
	CEM6	hybridation soustractive	facteur d'élongation	carotte	stade pré-globulaire et globulaire	Sato <i>et al.</i> , 1995
	top I	crible de banque d'ADNc	topoisomérase I	carotte	phase de prolifération et stade torpille	Balestrazzi <i>et al.</i> , 1996, 2001
	cdc2	mutant	protéine kinase	luzerne	stade précoce	Hirt <i>et al.</i> , 1991
	OsIAA1	crible de banque d'ADNc	cycline	riz	--	Thakur <i>et al.</i> , 2001
	CGS102/CGS103/CGS201	crible de banque d'ADNc	glutamine synthase	carotte	stade précoce	Higashi <i>et al.</i> , 1998
	CHI-GST1	mRNA differential display	glutathione S transférase	chicorée	stade précoce	Galland <i>et al.</i> , 2001
CHI3206	crible de banque d'ADNc	hémoglobine non symbiotique	chicorée	stade précoce	Hendriks <i>et al.</i> , 1998	
<b>Gènes induits par l'auxine</b>	<i>Dchsp-1/Dcarg-1</i>	crible différentiel de banque d'ADNc	protéine de choc thermique	carotte	stade précoce	Kitamiya <i>et al.</i> , 2000
	Nos43	mRNA differential display	protéine thaumatine-like	carotte	cellule embryogène	Yasuda <i>et al.</i> , 2001
	Nos87	mRNA differential display	non identifiée	carotte	cellule embryogène	Yasuda <i>et al.</i> , 2001
	Nos93	mRNA differential display	protéine précurseur de Dc2.15	carotte	cellule embryogène	Yasuda <i>et al.</i> , 2001
<b>Gènes induits par l'acide abscissique</b>	DcECP31/40/63	crible de banque d'ADNc	protéine inductible par ABA	carotte	stade torpille	Mundy <i>et al.</i> , 1990
	C-ABI3	crible de banque d'ADNc	facteur de transcription	carotte	-	Shiota <i>et al.</i> , 1998
	Pavp 1	crible de banque d'ADNc	facteur de transcription	épicéa	suspension embryogène et pic au stade cotylédonaire précoce	Footitt <i>et al.</i> , 2003 ; Fischerova <i>et al.</i> , 2008
<b>Transduction du signal</b>	ASET1/ASET2/ASET3	crible de banque d'ADNc	protéine kinase	luzerne	stade précoce	Giroux <i>et al.</i> , 1997
	<i>DcSERK</i>	crible de banque d'ADNc	Récepteur kinase	carotte	des cellules compétentes au stade globulaire	Schmidt <i>et al.</i> , 1997
	<i>AtSERK 1à5</i>	crible de banque d'ADNc	Récepteur kinase	<i>Arabidopsis</i>	stade précoce	Hetch <i>et al.</i> , 2001; Shah <i>et al.</i> , 2001a; Salaj <i>et al.</i> , 2008
	<i>ZmSERK1 et 2</i>	crible de banque d'ADNc	Récepteur kinase	maïs	stade précoce	Baudino <i>et al.</i> , 2001
	<i>SsSERK</i>	crible de banque d'ADNc	Récepteur kinase	pomme de terre	stade précoce	Sharma <i>et al.</i> , 2008
	swCDPK	Western blot	calcium et indépendante des calmodulines	bois de santal	stade globulaire	Anil <i>et al.</i> , 2000
	<i>MsCa1</i>	<i>mutant</i>	protéine calcium dépendante	luzerne	-	Dudits <i>et al.</i> , 1991
	<i>MsCPK3</i>	crible de banque d'ADNc	protéines kinases similaires aux calmodulines	luzerne	stade précoce	Davletova <i>et al.</i> , 1995
CRK	crible de banque d'ADNc	protéine kinase apparenté à une CDPK	carotte	-	Lindzen and Choi., 1995	

## Analyse Bibliographique

<b>Gènes à homéodomains</b>						
sbh1	crible de banque d'ADNc	facteur de transcription	soja	stade coeur à torpille		Ma <i>et al.</i> , 1994
CHB1 à 6	crible de banque d'ADNc	gène homéotique HD-Zip	carotte	stade globulaire à torpille		Burglin <i>et al.</i> , 1989; Kawahara <i>et al.</i> , 1995
DcCB1	crible de banque d'ADNc	protéine de liaison à l'ADN	carotte	stade précoce		Meijer <i>et al.</i> , 1997
WUSCHEL	mutant	facteur de transcription	<i>Arabidopsis</i>	-		Kiyosue <i>et al.</i> , 1998
CUS1	mRNA differential display	facteur de transcription	concombre	-		Zuo <i>et al.</i> , 2002
C-LEC1	crible de banque d'ADNc	facteur de transcription	carotte	-		Filipecki <i>et al.</i> , 1997
AGL15	mRNA differential display	facteur de transcription	luzerne	-		Yazawa <i>et al.</i> , 2004
<i>ZmMADS1/ ZmMADS3</i>	crible de banque d'ADNc	facteur de transcription	maïs	stade précoce		Perry <i>et al.</i> , 1999; Thakare <i>et al.</i> , 2008
CsSEF1	banque soustractive	protéine en doigt de zinc	concombre	stade précoce		Heuer <i>et al.</i> , 2001
						Grabowska <i>et al.</i> , 2009
<b>Protéines extracellulaires</b>						
EP1	crible de banque d'ADNc	chitinase	carotte	-		Van Engelen <i>et al.</i> , 1991,1995
EP3	mutant	chitinase	carotte	-		de Jong <i>et al.</i> , 1992
CG1/CG2/CG3	RT-PCR	β1,3-glucanase	chicorée	-		Helleboid <i>et al.</i> , 1998, 2000
PgChi-1	crible de banque d'ADNc	Endochitinase	épicéa	-		Dong et Dunstan, 1997b
Chia4-Pa	crible de banque d'ADNc	chitinase	épicéa	-		Wiweger <i>et al.</i> , 2003
PgGlu-1	crible de banque d'ADNc	β1,3-glucanase	épicéa	-		Dong et Dunstan, 1997b
DcAGP1	crible de banque d'ADNc	arabinogalactanes	carotte	-		Baldwin <i>et al.</i> , 2001
AGP	anticorps monoclonal	arabinogalactanes	chicorée	-		Chapman <i>et al.</i> , 2000
EP2	mutant	protéine de transfert des lipides	carotte	-		Sterk <i>et al.</i> , 1991
<b>Gènes de maturation</b>						
AtLTP1	crible de banque d'ADNc	protéine de transfert des lipides	<i>Arabidopsis</i>	-		Toonen <i>et al.</i> , 1997
Em	séquençage de protéine	protéine abondante de l'embryogenèse tardive	blé	stade torpille		Cuming <i>et al.</i> , 1979
Dc 2.15	crible de banque d'ADNc	protéine abondante de l'embryogenèse tardive	carotte	stade torpille		Holk <i>et al.</i> , 1996
Dc3/Dc8	crible de banque d'ADNc	protéine abondante de l'embryogenèse tardive	carotte	stade torpille		Cheng <i>et al.</i> , 1996
DcEMB1	mutant	protéine abondante de l'embryogenèse tardive	carotte	stade torpille		Wurtele <i>et al.</i> , 1993
BP8	hybridation hétérologue	protéine abondante de l'embryogenèse tardive	bouleau	stade torpille		Puupponen-Pimia <i>et al.</i> , 1993
Em		protéine abondante de l'embryogenèse tardive	blé	stade torpille		Corre <i>et al.</i> , 1996
LEC1	mutant	lectine	luzerne	stade torpille		Brill <i>et al.</i> , 2001; Meinke, 1992
LEC2	mutant	lectine	luzerne	stade torpille		Stone <i>et al.</i> , 2001, 2008; Braybrook <i>et al.</i> , 2006

## Analyse Bibliographique

<b>Divers</b>	<i>pcGER1</i>	crible de banque d'ADNc	germine	pin	stade pré-globulaire	Neutelings <i>et al.</i> , 1998
	<i>Mshsp18-1/Mshsp18-2</i>	crible de banque d'ADNc	protéine de choc thermique	luzerne	stade précoce	Gyorgyey <i>et al.</i> , 1997
	<i>MtSERF1</i>	microarray	dépend de l'éthylène	tabac	stade globulaire	Mantiri <i>et al.</i> , 2008

**Tableau 1** : Gènes impliqués au cours de l'embryogenèse somatique chez différentes espèces végétales.

## 1. Analyse des profils protéiques

Des études visant à établir une relation entre l'induction et différents stades de développement des embryons somatiques, ont permis d'identifier plusieurs protéines dont certaines sont spécifiques des masses proembryogènes et d'autres sont spécifiques des différents stades de développement des embryons somatiques.

Chez la carotte (*Daucus carota* L.), la séparation par électrophorèse bidimensionnelle de fractions protéiques obtenues à partir de masses proembryogènes et d'embryons somatiques, a permis la détection de deux protéines E1 et E2 faiblement présentes dans les masses proembryogènes cultivées sur un milieu renfermant du 2,4-D (Sung and Okimoto, 1981). Au contraire, ces protéines sont fortement accumulées dans les embryons, après transfert des masses proembryogènes sur un milieu sans auxine et sont absentes des cals non embryogènes. Par la suite, les mêmes auteurs ont identifié deux autres protéines C1 et C2 spécifiques des masses proembryogènes (Sung and Okimoto, 1983). Les auteurs suggèrent alors l'existence d'une régulation coordonnée de l'expression des gènes au cours de l'embryogenèse somatique, se manifestant par l'activation de certains gènes ou la répression de certains autres, suivant le stade du processus. Les séquences protéiques et la fonction des gènes codant ces protéines n'ont pas été identifiées puisqu'elles n'ont pas fait l'objet d'études ultérieures.

Plusieurs équipes ont également observé des différences entre les profils protéiques de cals embryogènes et non embryogènes et ainsi des protéines spécifiques de l'embryogenèse somatique ont pu être mis en évidence chez l'asperge (Dupire *et al.*, 1999), le blé (Fellers *et al.*, 1997), la luzerne (Giroux and Pauls, 1997) et la carotte (Franz *et al.*, 1989) (Tableau 1).

L'analyse par électrophorèse bidimensionnelle en gel de polyacrylamide a permis de mettre en évidence deux protéines de 38 kDa dans des explants foliaires de la chicorée après induction en condition d'embryogenèse somatique. Ces deux protéines s'accumulent seulement au bout du 5<sup>ème</sup> jour de culture et sont absentes des tissus non embryogènes (Hilbert *et al.*, 1992). Plus tard, ces protéines ont été identifiées comme des  $\beta$ -1,3 glucanases (Helleboïd *et al.*, 1998). L'activité de ces enzymes est trois fois supérieure dans le milieu embryogène chez l'hybride 474 contrairement au milieu non embryogène. Ces glucanases seraient impliquées dans

le processus de dégradation de la callose qui entoure les cellules engagées dans la voie embryogène. En effet, les cellules initiant un programme d'embryogenèse somatique s'isolent en déposant progressivement de la callose autour d'elle (Dubois *et al.*, 1990). La callose est ensuite dégradée lors du développement de l'embryon (Grimault *et al.*, 2007). Ce dépôt de callose est analogue à celui observé lors de la microsporogenèse et aurait comme rôle de protéger les microspores (cellules initiant le développement des grains de pollen) en "filtrant" les signaux extérieurs.

Il a été montré, chez la carotte, qu'un milieu conditionné par une culture embryogène avait un effet positif sur la production d'embryons somatiques d'une culture cellulaire (de Vries *et al.*, 1988). Ainsi, l'importance des échanges au cours de la formation des embryons somatiques *in vitro* entre les cellules et leur milieu environnant a favorisé la recherche et l'identification de protéines excrétées dans le milieu de culture.

A l'aide d'anticorps polyclonaux et après criblage d'une banque d'ADNc, des protéines telles que EP1, EP2 et EP3 ont été isolées chez la carotte. La protéine EP1 montre une forte homologie avec une glycoprotéine liée au locus S impliquée dans l'auto-incompatibilité. Elle n'est pas exprimée dans les cellules embryogènes ou les embryons somatiques mais dans les parois des cellules non embryogènes (Sterk *et al.*, 1991). Ainsi, cette protéine ne semble pas liée à l'embryogenèse somatique mais jouerait plutôt un rôle dans l'élongation des cellules (Sterk *et al.*, 1991). EP2 est excrétée uniquement par les cellules embryogènes et par les embryons (Sterk *et al.*, 1991) et son ARNm est détecté dans les masses proembryogènes et au niveau du protoderme des embryons. Cette protéine est une LTP (Lipid Transfert Protein), les auteurs supposent qu'elle est impliquée dans le transport de monères de cutine à travers la matrice extracellulaire (Hendriks *et al.*, 1994). Des cultures cellulaires de carotte ont été transformées avec le gène de luciférase ayant pour promoteur un gène homologue à EP2, *AtLTP1*. Le suivi des cellules par vidéo (cell tracking) a montré que tous les embryons somatiques sont dérivés de masses proembryogènes exprimant le gène *AtLTP1*/luciférase (Toonen *et al.*, 1997). Mais, toutes les masses proembryogènes exprimant ce gène n'ont pas développé d'embryons somatiques indiquant que ce gène n'est pas directement impliqué dans l'embryogenèse somatique.

La protéine EP3, est une endochitinase acide capable de lever le blocage du développement des embryons somatiques chez le mutant *ts11* sensible à la

température (De Jong *et al.*, 1992, 1993). Plusieurs isoenzymes correspondant aux chitinases EP3, ont été identifiées : EP3-1 et EP3-3, produites par des cellules uniques capables de lever le blocage du mutant *ts11*, et EP3-2 et EP3-5, produites principalement par des cals chez la carotte (Kragh *et al.*, 1996). Des expériences d'hybridations *in situ* dans des suspensions cellulaires de carotte n'ont révélé aucune expression du gène EP3 dans les embryons somatiques mais par contre elles ont montré une expression dans les cellules ne formant pas d'embryon somatique (van Hengel *et al.*, 1998). De plus, les ARNm de ces gènes ont été observés dans les tissus maternels et les protéines ont été détectées dans la matrice extracellulaire entourant les embryons zygotiques au stade globulaire (van Hengel *et al.*, 1998). Les auteurs supposent que les chitinases EP3 sont impliquées dans la réinitiation des divisions cellulaires des cellules embryogènes.

Chez la chicorée, d'autres protéines telles que des chitinases, une osmotine (Helleboid *et al.*, 2000), des protéines de transfert des lipides (LTPs) (Blanckaert *et al.*, 2002) et des arabinogalactane protéines (Chapman *et al.*, 2000) ont été caractérisées à partir du milieu embryogène de l'hybride 474.

Des protéines de 32 kDa et 25 kDa ont été identifiées dans le milieu de culture embryogène chez la chicorée. Après investigation de leur séquence protéique et de leurs propriétés biochimiques, les études ont montré qu'au moins deux chitinases et une osmotine sont sécrétées dans le milieu de culture après induction des explants foliaires de l'hybride 474 (Helleboïd *et al.*, 2000). L'accumulation de ces protéines est de 2 à 8 fois supérieure dans le milieu de culture embryogène comparé au non embryogène.

## 2. L'analyse génomique

Plus récemment, des méthodes de génomique fonctionnelle ont été entreprises chez plusieurs espèces pour identifier les gènes impliqués au cours de l'embryogenèse somatique. Des méthodes de transcriptomique, permettant de visualiser l'expression différentielle des gènes ont permis d'établir un profil d'expression de ces gènes pendant le développement des embryons somatiques.

### 2.1. Approche par criblage différentiel

Depuis la première étude de l'expression des gènes différentiellement exprimés durant l'embryogenèse somatique chez la carotte, plusieurs gènes préférentiellement exprimés durant le développement des embryons ont été isolés. Cette approche emploie des techniques basées sur la comparaison des populations d'ARNm entre des conditions ou des stades de culture différents. Cette démarche permet d'étudier les modifications de l'expression des gènes au niveau transcriptionnel. L'hybridation soustractive ou le criblage différentiel de banques ADNc ont été largement utilisés pour identifier les gènes préférentiellement exprimés au cours du processus de l'embryogenèse somatique. Plus récemment, la technique de "Differential Display" a été développée afin de mettre en évidence de tels gènes (Liang and Pardee, 1992 ; Liang, 2002).

### 2.1.1. Les marqueurs tardifs

Dans un premier temps, la plupart des gènes qui ont été identifiés sont spécifiquement exprimés lors des stades tardifs de l'embryogenèse somatique c'est-à-dire lors du développement de l'embryon somatique.

Plus de 20 gènes ont ainsi été clonés à partir de cultures cellulaires (Zimmerman, 1993). Ces gènes sont fortement exprimés lors des phases tardives (cœur et torpille) de l'embryogenèse. Plusieurs de ces gènes appartiennent à la classe des LEA (Late Embryogenesis Abundant). Les gènes LEA codent pour des protéines très hydrophiles, dont le rôle serait de protéger l'embryon lors de la dessiccation et d'empêcher la germination précoce de la graine. La synthèse de cette protéine est induite en réponse à un stress hydrique ou suite à un traitement à l'acide abscissique (ABA).

Le gène DC8 (*Daucus carota* 8) produit une protéine hydrophile de 66 kDa et son expression est embryo-spécifique (Franz *et al.*, 1989). Les transcrits du gène DC8 sont observés uniquement au stade cœur dans les embryons somatiques, de même que dans les embryons zygotiques. De la même manière, les transcrits du gène EMB-1 s'accumulent spécifiquement dans les deux types d'embryons somatiques et zygotiques. L'accumulation débute faiblement au stade globulaire pour atteindre son maximum au stade torpille, reflétant une expression différentielle dans

le temps (Wurtele *et al.*, 1993). La répartition des transcrits du gène EMB-1 est homogène dans les embryons globulaires alors qu'elle se polarise dans les tissus méristématiques lors du développement de l'embryon. Le gène DC3, quant à lui, s'exprime déjà dans les masses pro-embryogènes, ainsi que dans les embryons somatiques à différents stades de développement (Wilde *et al.*, 1988).

Des clones appartenant à la classe des gènes LEA ont également été identifiés chez d'autres espèces. Chez le bouleau (*Betula pendula*), le gène BP8 présente un profil d'expression faible par rapport à son homologue, le gène DC8 de la carotte, et ne permet pas de l'utiliser comme marqueur de l'embryogenèse somatique. Chez l'épicéa (*Picea glauca*), plusieurs clones *pgEMB* présentant une homologie avec des gènes LEA ont été isolés d'une banque d'ADNc préparée à partir d'embryons somatiques cotylédonaire (Dong and Dunstan, 1997a ; 1999).

Le gène *C-ABI3*, homologue du gène *ABI3* (*abscisic acid insensitive 3*) d'*Arabidopsis thaliana*, codant un facteur de transcription impliqué dans la voie de signalisation de l'ABA, a été isolé chez la carotte (Shiota *et al.*, 1998). Son expression est détectée spécifiquement dans les cellules embryogènes, dans les embryons somatiques et lors du développement de la graine. Elle n'est ni détectée dans les cellules non embryogènes, ni dans les plantules.

Les conditions nécessaires au développement d'embryons somatiques sont le plus souvent génératrices de stress (coupures, température ou pression osmotique élevées.....) et l'isolement de gènes codant pour des petites protéines de choc thermiques (HSPs) au cours du processus est fréquemment rapporté. Ces protéines ont un rôle de protection contre les dommages occasionnés par différents stress, particulièrement contre les chocs thermiques. En plus de ce rôle de défense, les HSPs interviennent dans la prolifération cellulaire et la différenciation.

Chez l'épicéa, trois ADNc (*PgEMB* 22, 27 et 29) dont les produits sont des HSP ont été caractérisés au cours de l'embryogenèse somatique (Dong and Dunstan, 1996b). Les transcrits sont détectés dans les tissus embryogènes et lors de la maturation de l'embryon somatique. Ils s'accumulent fortement lors des stades tardifs de l'embryogenèse somatique. De plus, l'expression des gènes *PgEMB* 22 et

29 est induite par une température élevée, un traitement à l'ABA ou au polyéthylène glycol.

De même, chez la luzerne (*Medicago sativa* L.), les gènes *Mshsp 18-1* et *Mshsp 18-2* codant pour des HSPs de faible poids moléculaire sont exprimés dans des cals initiés à partir de pétioles sous différentes conditions de stress (Györgyey *et al.*, 1991). Au cours des stades globulaires et cordiformes, les cultures en conditions normales, présentent une forte accumulation de ces protéines. Ces HSPs auraient un rôle de molécule chaperonne auprès des protéines nouvellement synthétisées au cours de la différenciation des cellules embryogènes.

### 2.1.2. Les marqueurs précoces

Les marqueurs précoces isolés sont spécifiques des étapes de dédifférenciation cellulaire, acquisition de la compétence embryogène, et des premières divisions des cellules embryogènes.

Plusieurs kinases ont été identifiées au cours de l'embryogenèse somatique soulignant l'implication de mécanismes de transduction de signaux afin d'activer des effecteurs responsables des modifications cellulaires observées lors de la formation des embryons somatiques. Ces protéines kinases sont activées par phosphorylation et sont impliquées dans la régulation d'éléments de la transduction du signal (Jimenez *et al.*, 2001 ; Shah *et al.*, 2001a).

Chez la luzerne (*Medicago sativa* L.), trois clones associés aux stades précoces de l'embryogenèse somatique ont été isolés par criblage différentiel d'une banque d'ADNc établie à partir de masses proembryogènes (Giroux and Pauls, 1997). Les gènes correspondants aux trois clones *ASET1*, *ASET2* et *ASET3* (*alfalfa somatic embryogenesis-specific transcript*) s'expriment dans les génotypes embryogènes mais pas dans les embryons matures de ces mêmes génotypes ni dans les génotypes non embryogènes. Ces séquences codent des protéines portant un domaine intra-membranaire et un site de phosphorylation.

Un gène marqueur de la transition d'une cellule somatique à une cellule embryogène a été identifié par mRNA differential display chez la carotte (Schmidt *et al.*, 1997). Ce gène *DcSERK* (*somatic embryo receptor-like kinase*) codant pour un récepteur de kinase, présente une expression qui coïncide avec la formation des

cellules compétentes. En hybridation *in situ*, son expression est observée dans quelques cellules isolées, dans les amas cellulaires après mise en culture des hypocotyles ainsi que dans les embryons globulaires. Elle n'est pas détectée lors des stades tardifs de l'embryogenèse somatique ni dans les cellules non embryogènes. Une expression transitoire a également été observée dans l'embryogenèse zygotique jusqu'au stade globulaire.

Les événements moléculaires ayant lieu lors de la transition d'une cellule somatique à une cellule embryogène sont très peu décrits. Le gène *SERK* est un des rares marqueurs associés à ce stade, permettant de distinguer une cellule compétente d'une cellule non compétente. En effet, au sein d'une culture cellulaire, la faible représentation de cellules évoluant vers la formation d'un embryon est un obstacle pour l'identification de tels gènes, car seuls les gènes les plus exprimés sont mis en évidence. Chez la carotte, il a été estimé que seulement 1 à 2% des cellules des masses proembryogènes évoluent en cellules embryogènes (de Vries *et al.*, 1998). De même, le suivi des cellules par vidéo (cell tracking) montre que 100 cellules sur 10 000 présentant une forme sphérique avec un cytoplasme dense vont se différencier en embryons somatiques (Toonen *et al.*, 1994).

Les gènes *SERK* ont été clonés et caractérisés chez de nombreuses espèces comme la pomme de terre (*Solanum tuberosum*), le concombre (*Cucumis sativus*), *Arabidopsis thaliana*, la luzerne (*Medicago sativa*) et le maïs (*Zea mays*).

Chez *Arabidopsis*, cinq membres de la famille *SERK* ont été identifiés (*AtSERK1*, *AtSERK2*, *AtSERK3*, *AtSERK4* et *AtSERK5*). *AtSERK1* est exprimé dans le nucelle, les mégaspoires, le sac embryonnaire mais aussi durant l'embryogenèse somatique. De plus, la surexpression du gène *AtSERK1* dans les cals augmente de 3 à 4 fois leur compétence embryogène. Selon les auteurs, l'expression de *AtSERK1* coïncide avec les phases précoces de l'embryogenèse somatique mais aussi la précède (Salaj *et al.*, 2008). En effet, ils ont observé l'expression de *AtSERK1* :: GUS dans des cellules uniques, indiquant ainsi que ces cellules sont les précurseurs de l'embryogenèse somatique.

Chez le maïs, deux gènes *ZmSERK1* et *ZmSERK2* ont été isolés en utilisant des amorces dégénérées (Baudino *et al.*, 2001). Le gène *ZmSERK1* présente une expression préférentielle dans les tissus reproducteurs mâles et femelles avec une forte expression dans les microspores alors que *ZmSERK2* présente une expression uniforme dans l'ensemble des tissus. Ces deux gènes ne sont pas spécifiques de

l'embryogenèse somatique puisqu'ils sont exprimés aussi bien dans les cellules embryogènes que non embryogènes.

De même, le gène SERK a été cloné et identifié chez la pomme de terre (*StSERK1*) et montre une forte homologie avec *MtSERK*, *CuSERK*, *AtSERK* et *DcSERK*. L'expression de ce gène augmente lors de l'induction de l'embryogenèse somatique après traitement à l'auxine des cultures. Ensuite, aucune modulation de l'expression n'est observée lors des phases plus tardives du développement des embryons somatiques (Kumar *et al.*, 2008).

Chez la chicorée, des travaux basés sur la recherche de gènes candidats pour identifier les gènes impliqués au cours de l'embryogenèse somatique ont été réalisés. Cette approche a reposé sur la définition d'amorces dégénérées correspondant à des régions très conservées des protéines produites par ces gènes. Ainsi, deux ADNc (*SKL09* et *SKL22*) présentant de fortes homologies avec les gènes SERK de carotte et d'*Arabidopsis* ont été isolés. L'analyse par RT-PCR du niveau de transcrite montre que l'expression de ces deux gènes ne semble pas liée à l'embryogenèse somatique (Duban, 2004). La comparaison du niveau de transcrite de ces deux gènes avec *DcSERK* et *AtSERK1* suggère que ces gènes ne sont pas leurs équivalents fonctionnels. Cependant, l'expression du gène *SKL22* est élevée dans les plantules germées après 4 jours en absence de lumière et cette différence de niveau d'expression disparaît après huit jours de germination. Ce profil d'expression peut être rapproché de celui de *AtSERK3/BAK1* (*BRI1 Associated receptor-like Kinase 1*) chez *Arabidopsis* codant le récepteur aux brassinostéroïdes. En effet, les transcrits de ce dernier sont plus abondants dans les plantules germées 3 jours à l'obscurité qu'à la lumière et cette différence s'estompe après 5 jours de germination (Li and Chory, 1997).

Chez la chicorée, l'analyse des populations d'ARNm présents dans les tissus placés en condition d'embryogenèse somatique a permis d'isoler plusieurs gènes.

La technique par RT-PCR a permis d'isoler trois ADNc partiels de  $\beta$ -1,3 glucanases (CG1, CG2 et CG3). Seule l'expression du gène CG1 est corrélée de manière spécifique avec la formation des embryons somatiques. Les ARNm correspondant à ce gène sont présents dans les tissus foliaires chez l'hybride '474' embryogène uniquement (Helleboid *et al.*, 2000).

Après criblage différentiel d'une banque d'ADNc, un ADNc pleine longueur présentant de fortes homologies avec une hémoglobine non symbiotique a été isolé. Le gène correspondant est exprimé préférentiellement au cours de la phase d'induction et aurait ainsi selon les auteurs un rôle dans le processus de dédifférenciation ou dans la réacquisition de la compétence embryogène (Hendriks *et al.*, 1998).

Chez la chicorée, la technique de différentiel display a permis l'identification d'un gène codant une glutathione S-transferase (CHI-GST1) (Galland *et al.*, 2001). L'analyse par northern blot indique une accumulation des transcrits dans les feuilles des génotypes embryogènes. La combinaison de la technique de Differential Display/RT-PCR avec un criblage d'une banque d'ADNc et la technique RACE-PCR a permis d'isoler deux ADNc *CHI-GTP1* et *CHI-GTP2* codant pour une protéine de liaison au GTP de type rab5. Les transcrits du gène *CHI-GTP1* sont présents uniquement chez le génotype embryogène '474'. Ce gène semble impliqué dans le transport vésiculaire au cours du processus d'embryogenèse somatique (Randoux *et al.*, 2002).

### 2.2. Technique de microarray

L'essor récent des méthodes d'analyse de l'expression génique par séquençage d'EST à haut débit ou grâce à l'utilisation de la technologie microarray a permis d'envisager une étude exhaustive et simultanée de l'expression des gènes au cours du développement embryonnaire. En permettant l'analyse de l'expression d'un très grand nombre de gènes, ces études de transcriptomique ont fourni les premières images globales du processus cellulaire impliqué à différentes étapes de l'embryogenèse somatique.

Chez le soja (*Glycine max*), l'embryogenèse somatique a été étudiée par le biais d'analyses microarray (Thibaud-Nissen *et al.*, 2003). Pour cette analyse, les auteurs ont utilisé les embryons somatiques formés à partir du côté supérieur des cotylédons (côté adaxial) après mise en culture sur un milieu riche en 2,4 D et comme témoin négatif du processus de développement, la face inférieure (côté abaxial) de cet organe (favorisant le développement de cal uniquement) ne réagissant pas aux conditions d'induction. Les ARN des embryons somatiques et des

cals ont ensuite été extraits autorisant une analyse par microarray. En utilisant 9 280 ADNc, les auteurs ont pu comparer les transcrits présents au niveau adaxial des cotylédons à partir duquel se développe les embryons somatiques et le côté abaxial callogène et ainsi identifier 495 gènes montrant une modulation de l'expression en réponse au traitement de 2,4-D durant le développement des embryons somatiques. Le regroupement des gènes par profils d'expression similaires sur la cinétique d'étude a permis d'établir une chronologie des événements moléculaires qui ont lieu au cours du processus. Les résultats ont indiqué que l'apparition des embryons somatiques est précédée par une étape de dédifférenciation des cellules situées du côté adaxial des cotylédons. De plus, l'abondance des gènes caractéristiques du stress oxydatif et de la division cellulaire semble indiquer un réarrangement au niveau cellulaire. Les données suggèrent également que la formation des embryons somatiques est accompagnée par la synthèse d'acide gibbérellique et la transcription des protéines de stockage.

D'autres banques ADNc de type SSH (Suppression Subtractive Hybridisation) (Diatchenko *et al.*, 1996) comparant les transcrits d'embryons au stade cœur et cotylédonaire et enrichies en gènes préférentiellement exprimés au stade cœur ont été créées chez le tournesol (*Helianthus annuus*) (Ben, 2005). Ces travaux ont eu pour but de caractériser les transcrits spécifiques induits au cours des étapes d'élaboration et d'acquisition du patron embryonnaire. Les résultats ont indiqués une régulation transcriptionnelle et post-transcriptionnelle fine et spécifique à chacun des stades de développement des embryons et diverses voies de transduction du signal.

Plus récemment, les ARNm extraits des explants foliaires de *Cichorium intybus* d'un génotype embryogène, K59 et non embryogène, C15 (issu de l'autofécondation de K59) en culture en condition d'embryogenèse somatique ont été utilisés afin de générer une banque embryogène et non embryogène par la technique SSH. Les résultats montrent que plusieurs gènes préférentiellement exprimés chez le génotype embryogène semblent coder des protéines impliquées dans la division cellulaire et la synthèse protéique. Chez le génotype non embryogène, les protéines codées semblent plutôt liées au mécanisme de défense (Legrand *et al.*, 2007). Les détails de ces travaux seront discutés un peu plus tard dans ce chapitre.

### 3. L'approche génétique

Les analyses génomiques que nous venons de détailler ont permis d'identifier la plupart des gènes potentiellement impliqués lors de l'embryogenèse somatique, encore appelés gènes candidats. La caractérisation de ces gènes au niveau génomique et protéique a permis d'envisager des études génétiques afin d'élucider le rôle de ces gènes.

#### 3.1. L'analyse par mutagenèse

A l'heure actuelle, aucun mutant spécifique de l'embryogenèse somatique n'est connu. Au contraire de nombreux mutants zygotiques du développement embryonnaire ont pu être caractérisés fournissant ainsi la majorité des connaissances actuelles sur le programme génétique et ceci principalement chez *Arabidopsis thaliana*.

Le mutant le plus précoce au cours du développement embryonnaire décrit à ce jour est appelé *gnom/emb30 (GN)* (Mayer *et al.*, 1993 ; Shevell *et al.*, 1994). L'inactivation du gène *Gnom/Emb30* dans les embryons d'*Arabidopsis* entraîne une première division anormale (plutôt symétrique) ayant pour cause une élongation selon l'axe apico-basal réduite. Le gène *GNOM* code pour une protéine montrant une similarité avec un facteur d'échange des nucléotides guanines impliqué dans le transport vésiculaire (Bush *et al.*, 1996). D'autres mutants du développement ont été caractérisés et ont permis de comprendre l'établissement de l'axe apico-basal et radial mais aussi le développement du suspenseur chez les embryons zygotiques.

Plusieurs de ces mutants zygotiques d'*Arabidopsis thaliana* ont permis d'étudier certains aspects de l'embryogenèse somatique (Mordhorst *et al.*, 1997 ; 1998). Les embryons somatiques obtenus à partir des mutants *hpt (häuptling)*, *cop2 (constitutive photomorphogenic2)* et *amp1 (altered meristem program1)* présentent un développement anormal de leur méristème apical. De plus, des cellules embryogènes peuvent être obtenues à partir d'autres mutants comme *stm (Shoot Meristemless)* et *wus (Wuschel)* ne présentant pas de méristème apical. L'étude de ces mutants a ainsi révélé qu'un méristème apical fonctionnel n'est pas nécessairement requis lors de l'embryogenèse somatique.

Par ailleurs, l'étude du mutant *amp1* a permis également de confirmer le rôle du gène *SERK* dans le processus de l'embryogenèse somatique. En effet, l'expression de *AtSERK1* est augmentée de façon importante chez ce mutant lors de la formation des embryons somatiques.

Des études d'hybridation *in situ* pour étudier les profils d'expression de *STM* (*Shootmeristemless*) ont été réalisés chez les plantes sauvages d'*Arabidopsis* (Long *et al.*, 1996). Chez les embryons zygotiques, les transcrits apparaissent avant l'initiation du méristème apical et s'accumulent dans les cellules initiales du méristème à partir du stade globulaire dans une ou 2 cellules. Ces résultats révèlent également une relation entre les cellules qui accumulent le transcrit et les cellules prédestinées à donner naissance au méristème apical caulinaire (MAC). Au stade cœur, torpille et dans les embryons matures, l'expression est restreinte à la base des cotylédons.

Le gène *STM* a fait l'objet d'étude chez la chicorée pour son implication dans les mécanismes de dédifférenciation et redifférenciation au cours des phases précoces de l'embryogenèse somatique (Da Silva, 2004). Ces travaux ont abouti au clonage de l'ADNc complet de *STM* par des amorces dégénérées déduites à partir des régions conservées. La comparaison entre *AtSTM* et *CiSTM* montre une forte homologie de séquence avec 73% (Da Silva, 2004). Des analyses en temps réel à partir d'ARN extraits des explants foliaires du génotype K59 mis en condition d'embryogenèse somatique ont révélé une induction significative des transcrits *CiSTM* au 3<sup>ème</sup> jour de culture.

L'analyse des mutants s'est avérée un outil très enrichissant pour accroître la connaissance des gènes impliqués au cours de l'embryogenèse. Toutefois, l'analyse par mutagenèse ne permet pas d'élucider l'ensemble des mécanismes moléculaires et physiologiques de l'embryogenèse. Cette approche présente quelques limites puisque lorsqu'il existe des gènes homologues ou à fonction homologue, la mise en évidence de l'implication du gène reste difficile.

### 3.2. Analyse de la variation naturelle

La présence d'une variation naturelle pour la capacité à l'embryogenèse somatique chez différentes espèces a été exploitée pour identifier les gènes impliqués au cours de l'embryogenèse somatique.

#### 3.2.1. Modèles génétiques 'simples'

Plusieurs études sur le déterminisme de l'embryogenèse somatique chez la luzerne (*Medicago sativa*) ont suggéré un modèle à deux gènes à effets dominants et complémentaires. La ségrégation observée dans une F1 issue d'un croisement entre un génotype embryogène et un génotype non embryogène indique que le parent embryogène aurait un génotype AAaaBbbb, tandis que le parent non embryogène serait de génotype aaaabbbb. Les auteurs proposent que la présence d'une triple dose d'allèles favorables est nécessaire pour que les individus soient de phénotype embryogène (Hernandez-Fernandez and Christie, 1989 ; Kielly and Bowley 1992 ; Yu and Pauls 1993). Une autre étude sur la luzerne confirme les résultats précédents, à partir d'un croisement entre deux génotypes embryogènes. L'embryogenèse somatique ségrége selon un modèle à deux gènes dominants mais le phénotype embryogène est obtenu avec seulement un allèle favorable pour chaque gène (Crea *et al.*, 1995). Un modèle à deux gènes à effet complémentaire a été également décrit pour expliquer le déterminisme de l'embryogenèse somatique chez le trèfle (*Trifolium pratense*) (McLean and Nowak 1998) et le blé (De Buyser *et al.*, 1992).

Des études visant à identifier les gènes impliqués dans la régénération *in vitro* et l'organogenèse ont également été réalisées. Chez le pétunia (Dulieu, 1991), la tomate (Koorneef *et al.*, 1993), et la luzerne (Reich and Bingham, 1980), la capacité de régénération semble être contrôlée par deux allèles dominants présents sur deux locus différents. En effet, chez la tomate, la présence d'un allèle dominant *Rg-1* (*Regeneration1*) dans des cultures jeunes semble être suffisante pour régénérer les bourgeons alors que la présence d'un second allèle dominant *Rg-2* (*Regeneration 2*) semble être requise dans les cultures plus âgées. Alors que d'autres études chez le concombre (Nadolska-Oroczyk and Malepszy, 1989) et le seigle (*Secale cereale*)

(Rakoczy-Trojanowska and Malepszy, 1993) révèlent que ce caractère est contrôlé, respectivement, par des gènes complémentaires ou des allèles récessifs.

### 3.2.2. Analyses QTL

Les caractères liés à la régénération *in vitro* sont des caractères quantitatifs impliquant plusieurs locus ou QTL (Quantitative Trait Loci). L'analyse QTL permet de localiser ces locus sur une carte génétique.

Plusieurs analyses QTL ont été entreprises afin d'identifier les régions impliquées dans l'embryogenèse somatique chez les végétaux (Tableau 2). L'identification de plusieurs locus liés à l'embryogenèse somatique a été réalisée chez des espèces comme le maïs (Armstrong *et al.*, 1992), la luzerne (Yu and Pauls, 1993), l'orge (Mano *et al.*, 1996), le seigle (Bolibok *et al.*, 2007) et le blé (Haiyan *et al.*, 2007). Chez le tournesol, la mise en culture d'hypocotyles a permis d'identifier plusieurs QTL liés à l'embryogenèse somatique (Florres Berrios *et al.*, 2000b) expliquant chacun entre 7 et 20% de la variation observée (Tableau 2). Plus récemment, Huang et ses collaborateurs (2007) ont affiné les régions QTL détectées et ainsi identifié quatre marqueurs AFLP potentiellement liés à l'embryogenèse somatique.

Chez ces mêmes espèces mais aussi chez d'autres, les auteurs se sont également intéressés au caractère lié à l'organogenèse et à la callogenèse (Tableau 2). De nombreux auteurs ont ainsi identifié plusieurs locus liés à la régénération des bourgeons. Ces travaux ont été réalisés chez le maïs (Armstrong *et al.*, 1992), le riz (Tagushi-Shiobara *et al.*, 1997), l'orge (Mano *et al.*, 2002), le tournesol (Florres-Blerios *et al.*, 2002a), *Arabidopsis* (Schiantarelli *et al.*, 2004), le brocoli (Holme, 2004) et le blé (Haiyan *et al.*, 2007). Les différents QTL identifiés lors de la régénération *in vitro* ont un effet additif.

Récemment, une analyse QTL pour la capacité liée à la régénération *in vitro* chez le riz a permis d'identifier quatre QTL impliqués dans ce caractère (Nishimura *et al.*, 2005). Une analyse cartographique plus fine au niveau du QTL majeur, a permis d'isoler un gène codant pour une ferredoxine nitrite réductase. Des analyses par RT-PCR ont montré que ce gène est faiblement exprimé chez le parent Koshilikari ayant une capacité de régénération faible.

Ces analyses QTL ont été réalisées à partir de différentes populations (Tableau 2). Les populations les plus souvent utilisées pour la détection de QTL lié à la régénération *in vitro* sont les lignées recombinantes, principalement dérivées d'une population F2 ou d'une lignée haploïde doublée. Ce type de population présente un grand avantage pour l'analyse QTL puisqu'elle constitue des lignées homozygotes ou quasi homozygotes qui peuvent être propagées indéfiniment. Dans certains cas, une population F2 est également utilisée lors de l'analyse QTL (Armstrong *et al.*, 1992). Cependant, ce type de population ne permet pas de disposer d'une grande quantité de matériel végétal contrairement aux lignées recombinantes.

## Analyse Bibliographique

Espèce	Explant	Caractère étudié	Population		Marqueur		Analyse QTL			Références
			Type de descendance	Taille	Type	Nombre	Type d'analyse	Nombre de QTL détectés	% totale de la variation expliquée	
<b>Arabidopsis</b>	racine	Nombre de bourgeons différenciés	RILS	30	Marqueur moléculaire	163	IM	2	nd	Schiantarelli <i>et al.</i> , 2001
	feuille	Nombre de bourgeons différenciés						1	nd	
<b>Tournesol</b>	cotylédon	Nombre de bourgeons différenciés par explant	RILS	75	AFLP	254	CIM	6	52 (7-11)	Florres-Berrios <i>et al.</i> , 2000a
		Nombre de bourgeons différenciés par explant régénéré						7	67 (6-17)	
	tissu épidermique	Nombre d'explants embryonnaires						4	48 (7-15)	Florres-Berrios <i>et al.</i> , 2000b
	Nombre d'embryons formés						7	89 (7-20)		
<b>Riz</b>	graine mature	Nombre de bourgeons régénérés par cal	BC	98	RFLP	245	CIM	5	38,5 (6,2 -18,6)	Taguchi -Shiobara <i>et al.</i> , 1997
		% de cals régénérés						4	32,6 (6,4-14,1)	
<b>Brocoli</b>	protoplaste	Nombre de plantes régénérés pour 100 cals	F2	128	AFLP, SSR	54	IM	2	83	Holme <i>et al.</i> , 2004
<b>Blé</b>		% d'embryons formés par cal	RILS	136	RFLP	86	CIM	4	11-22,8	Haiyan <i>et al.</i> , 2007
		% de cals régénérés par plante						4	12,6-17,6	
		Nombre de plantes formées par cal régénéré						4	12,8-30,2	
<b>Orge</b>	embryon immature	Croissance des cals	RILS	58-99	AFLP	100	CIM	3	55 (11,6-31,3)	Mano <i>et al.</i> , 2002
		Nombre de bourgeons différenciés						3	57,8 (10-35,9)	
<b>Maïs</b>	embryon immature	% d'embryons formés	F2	86	RFLP	73	ANOVA	5	82 (10,4-26,7)	Armstrong <i>et al.</i> , 1992
		Nombre de plantes développées à partir d'embryon						4	70 (6,8-14,3)	
		Induction des cals	RILS	126	AFLP, SSR	111	CIM	8	39 (6-22)	
<b>Seigle</b>	embryon immature	% d'embryons immatures par cal	RILS	102	Marqueur moléculaire	99	CIM	2	(20,8-22,1)	Bolibok <i>et al.</i> , 2007
		% d'embryons immatures par cal embryogène						4	(21,5-41,6)	
	inflorescence immature	% d'inflorescence immature par cal						2	(11,4-20,6)	
		% d'inflorescence immature par cal embryogène						1	(10,8)	

**Tableau 2 :** Cartographie de QTL liés à la régénération *in vitro* et à l'embryogenèse somatique chez différentes espèces.

## Analyse phénotypique

III. La chicorée : une plante modèle de l'embryogenèse somatique chez les dicotylédones.

La chicorée est une plante dicotylédone ( $2n=18$ ), de la famille des Astéracées, traditionnellement cultivée dans la région du Nord-pas-de-Calais. C'est l'une des raisons pour lesquelles en partenariat avec différents sélectionneurs que notre laboratoire développe des projets de recherche sur cette plante.

### 1. La chicorée : classification botanique

Il existe trois espèces de chicorée présentes en Europe : *Cichorium spinosum* L., *Cichorium endivia* L. et *Cichorium intybus* L. *Cichorium endivia* L. et *Cichorium intybus* L. présentent un intérêt agronomique important contrairement à *Cichorium spinosum* L. qui est peu cultivée. *Cichorium endivia* L. qui a un cycle de vie annuel, est cultivée pour ses feuilles (scarole ou chicorée frisée). *Cichorium intybus* L. présente aussi bien un cycle de vie annuel, bisannuel ou vivace. On distingue deux variétés :

- *C. intybus Foliosum* ou Witloof cultivée pour la production de chicons : Semée en avril/mai, la première année, la plante va développer une rosette de feuilles et une racine pivotante tubérisée accumulant des réserves azotées et carbonées. A cette période, le cycle peut-être arrêté pour la production de chicons. Le cas échéant, après vernalisation, la plante érige une hampe florale portant un grand nombre de capitules composés de fleurs bleu-violet.

- *C. intybus Sativum* ou chicorée industrielle cultivée pour la récolte de ses racines utilisées pour la production d'inuline et pouvant être torréfiées.

### 2. Historique de la culture de chicorée

Les premières références à la chicorée remontent à l'Egypte ancienne où elle était utilisée et appréciée pour ses propriétés cicatrisantes, connue pour faciliter la digestion et son effet légèrement laxatif. La chicorée est introduite en France par les agriculteurs belges dans les années 1920. Elle est principalement cultivée dans le nord de la France pour la production d'endives (80% de la production française). La

plus grande partie de la production de chicorée industrielle est exploitée par le groupe Leroux situé à Orchies.

### 3. L'embryogenèse somatique chez la chicorée

Les premières études de l'embryogenèse somatique chez la chicorée datent de 1988. Des embryons somatiques étaient obtenus à partir d'organes reproducteurs tels que les styles (Dubois *et al.*, 1988) et les anthères (Guedira *et al.*, 1989) de l'hybride "474". Par la suite, les explants racinaires (Dubois *et al.*, 1990) et les feuilles (Dubois *et al.*, 1991) se sont révélés inductibles pour l'obtention des embryons somatiques. De plus, l'étude cytologique des embryons somatiques issus de l'hybride "474" a permis de caractériser les différentes phases précoces du processus d'embryogenèse (Blervacq *et al.*, 1995).

La multitude d'explants induisant la formation d'embryons somatiques et l'obtention de ces embryons par une voie directe, sans passage par une voie de callogenèse, ont fait de la chicorée une plante modèle pour l'étude de ce processus.

### 4. Objectifs de la thèse

Nous avons décidé dans notre travail d'utiliser une approche mixte pour identifier les gènes impliqués dans l'embryogenèse somatique. La présence d'une variation naturelle liée à l'embryogenèse somatique chez la chicorée au sein d'une descendance K28\*K59 nous a permis d'initier une analyse QTL. Notre approche a consisté à coupler une analyse QTL avec la localisation des gènes candidats. De nombreux gènes candidats ont été identifiés lors des différentes approches présentées dans les chapitres précédents. Leur co-localisation avec un QTL permet une hiérarchisation des gènes candidats pour des études plus approfondies de leur implication au cours de l'embryogenèse somatique.

Cette approche a nécessité plusieurs étapes :

- L'obtention d'une descendance K28\*K59 en ségrégation pour le caractère étudié

## Analyse phénotypique

- L'évaluation du potentiel embryogène des individus de la descendance
- La construction d'une carte génétique F1'a à partir des ADN des descendants de K59\*K28
- La détection de polymorphisme au sein des séquences des gènes candidats potentiels afin de les localiser sur la carte génétique F1'a

Après réalisation de ces différentes étapes, nous avons recherché la corrélation statistique entre les marqueurs localisés sur la carte génétique F1'a et la valeur phénotypique déterminée pour chacun des individus de la descendance.

Puis les marqueurs sous-jacents aux QTL ont été étudiés quant à leur implication dans l'embryogenèse somatique.

### 4.1. Obtention d'une descendance K28\*K59 en ségrégation pour l'embryogenèse somatique.

La première étape a été la recherche de parents à génotypes extrêmes pour l'embryogenèse somatique afin de générer une descendance en ségrégation. Les premières études de l'embryogenèse somatique ont été faites à partir de l'hybride « 474 » issu du croisement *Cichorium intybus* × *Cichorium endivia* (Dubois *et al.*, 1991 ; Blervacq *et al.*, 1995). Cet hybride a montré rapidement des limites puisque celui-ci, du fait de sa nature interspécifique, était très peu fertile. Il a été impossible d'utiliser l'hybride "474" comme parent afin de générer de nombreux descendants en ségrégation. C'est pourquoi, un crible de la population Koospol, d'où est issu le parent *C. intybus* du génotype "474", a été réalisé pour l'aptitude de cette population à produire des embryons somatiques *in vitro*.

Ainsi, deux génotypes, un très embryogène (K59) et un peu embryogène (K28) ont été sélectionnés et ont été utilisés pour obtenir une descendance en ségrégation. Le parent K28 a été utilisé comme parent femelle car il est mâle stérile. L'apport manuel de pollen de K59 sur les stigmates de K28 a permis d'obtenir de nombreux individus en ségrégation pour la capacité à former des embryons somatiques.

## Analyse phénotypique

### 4.2. L'évaluation du potentiel embryogène des individus de la descendance K28\*K59.

Le potentiel embryogène des parents K28 et K59, et de chacun des individus de la descendance, a été déterminé à partir des explants racinaires de plantules de chicorée cultivées *in vitro* (Chapitre 1). Après 30 jours en condition d'embryogenèse somatique, les structures formées c'est-à-dire les plantules issues de la formation des embryons somatiques et les structures chlorophylliennes unipolaires, ont été comptabilisées pour chacun des descendants. Ainsi, pour chaque génotype, nous avons attribué une valeur reflétant son potentiel embryogène.

### 4.3. La construction d'une carte génétique F1'a+b

La construction d'une carte génétique K28\*K59 a été réalisée dans notre laboratoire (Chapitre 2). La carte a été construite à partir d'une population F1' interspécifique de type "outbred" en utilisant la stratégie de "pseudo testcross" de Grattapaglia et Sederoff (1994). Depuis plusieurs années, notre laboratoire s'est attaché à générer une grande quantité de marqueurs (Encadré 2) afin de construire la carte génétique K28\*K59 de *Cichorium intybus* L. Des marqueurs SSR (Simple Sequences Repeats) produits à partir de 3 banques enrichies en motif SSR (Agrogènes, France ; Ecogenics, Suisse ; SADV, France) et des séquences STS (Sequences Tagged Sites) ont permis de couvrir les 9 groupes de liaison de la carte génétique K28\*K59. Ces travaux ont fait l'objet d'un article "A consensus genetic map for chicory (*Cichorium intybus* L.) " (Cadalen *et al.*, 2009) soumis à la revue *Molecular Breeding* (Annexe 1). Mon travail a consisté à inclure des marqueurs SSR dans la carte génétique K28\*K59. De plus, une nouvelle carte nommée F1'a construite à partir de marqueurs sélectionnés sur la carte génétique K28\*K59 et incluant des marqueurs AFLP (Amplification Fragment Length Polymorphism) a été réalisée.

**Les séquences ESTs** : elles correspondent à des gènes ou à des fragments de gènes de fonction connue ou non. Elles proviennent de gènes clonés ou d'ADNc transcrits en ARN messagers s'exprimant dans certaines conditions. Les ESTs sont utilisés pour définir des amorces spécifiques, permettant leur amplification par PCR. Ce sont généralement des marqueurs codominants.

**Les marqueurs AFLP** : cette technique, développée par Vos *et al.* (1995), permet d'amplifier de manière sélective des fragments d'ADN issus d'une digestion préalable de l'ADN génomique par des enzymes de restriction. Elle permet d'obtenir un grand nombre de marqueurs de type dominant. L'avantage de cette technique est qu'elle ne demande pas forcément la connaissance du génome de l'organisme étudié et de plus elle requiert une faible quantité d'ADN. Désavantage : dominants et marqueurs non directement transférables entre les descendance différentes

**Les marqueurs microsatellites** : ils sont constitués de répétitions en tandem de motifs mono-, di-, tri- ou tétranucléotiques. Les motifs sont variables comme (A)<sub>n</sub>, (TC)<sub>n</sub>, (TAT)<sub>n</sub>. L'avantage des microsatellites, est qu'ils sont distribués sur l'ensemble du génome donc très nombreux mais aussi qu'ils sont hautement polymorphes (polymorphisme du nombre de répétitions). Ce sont généralement des marqueurs codominants.

**Encadré 2** : Type de marqueurs utilisés pour la construction de la carte génétique K28\*K59.

## 4.4. La détection de polymorphisme au sein des gènes candidats

Certains des gènes identifiés au cours de l'embryogenèse somatique chez la chicorée (Tableau 2) par différentes approches décrites précédemment ont pu être localisés sur la carte génétique F1'a, après détection de polymorphisme au sein de leur séquence (Cadalen *et al.*, 2009). Ces différents gènes potentiellement impliqués au cours de la formation des embryons somatiques peuvent être considérés comme des gènes candidats.

Gènes cartographiés	Nature de la protéine déduite	Groupe de liaison
CLO22L1R1*	Somatic embryogenesis receptor kinase	1
GST5'-RACE4	Glutathione-S-transferase	1
ESTE002B3	GTP binding protein	3
ESTE002C9	Haloacid dehalogenase-like hydrolase family protein	3
ESTE002B7	Amino acid permease 6	4
MADS box	MADS box protein	4
SK95UTR9L19R6	Somatic embryogenesis receptor kinase	5
3262L1R1	Dehydrine	6
GTPB L3R1	GTP binding protein (GTP1)	6
3142L2R2*	Hydroxy-methylglutaryl-coenzyme A synthase	8
3361L1R1	Lipid Transfer Protein 1	8
3704L2R1	Lipid Transfer Protein 2	8
Tubulin L2R2	B-tubulin	8
3474 L1R1	Hémoglobine 1	9
GTPB L2R2	GTP binding protein (GTP1)	9

**Tableau 2 :** Gènes identifiés chez la chicorée comme potentiellement impliqués au cours de l'embryogenèse somatique et localisés sur la carte génétique K28\*K59 (Cadalen *et al.*, 2009).

\*Les marqueurs CLO22L1R1 et 3142L2R2 sont seulement cartographiés sur la carte génétique Rubis118.

Récemment, une analyse de l'embryogenèse somatique chez la chicorée par microarray a fourni un grand nombre de gènes candidats (Legrand *et al.*, 2007). En effet, la variabilité génétique pour la formation des embryons somatiques a été exploitée afin de générer une banque d'ADNc embryogène et une banque d'ADNc non embryogène. Les banques ont été construites à partir des génotypes K59 (embryogène) et C15 (non embryogène). C15 provient de l'autofécondation de K59, les deux génotypes présentent donc un fond génétique commun.

## Analyse phénotypique

Au 4<sup>ème</sup> jour de culture, dans les conditions d'induction de l'embryogénèse somatique, des analyses cytologiques ont mis en évidence les premières divisions cellulaires chez K59 (embryogène), non observables chez C15 (non embryogène). Les ARNm extraits des explants foliaires de K59 et C15 après 4 jours de culture (apparition des premières divisions cellulaires) ont été utilisés afin de générer par la technique SSH (suppression subtractive hybridisation) une banque embryogène avec 1944 EST et non embryogène avec 1404 EST. Ces 3348 EST ont été regroupés dans 2302 OC (Original Clusters) correspondant à 2077 gènes.

Une analyse par microarray a mis en évidence les gènes différentiellement exprimés au cours de l'embryogénèse somatique. Cette analyse a utilisé le génotype C15 comme témoin négatif de ce processus. Ainsi 105 gènes différentiellement exprimés entre le génotype K59, embryogène (37 gènes) et C15, non embryogène (68 gènes) ont pu être identifiés au 4<sup>ème</sup> jour de culture en condition d'induction de l'embryogénèse somatique (Legrand, 2006). Le choix du génotype non embryogène comme témoin a permis d'identifier les gènes présentant la même réponse chez les deux génotypes.

Par la suite, une soustraction *in silico* a été réalisée afin de vérifier l'efficacité de soustraction lors de la fabrication des banques soustractives.

Ainsi après la soustraction *in silico*, 33 gènes se sont révélés différentiellement exprimés ; 6 gènes sont préférentiellement exprimés chez le génotype embryogène K59 et 27 chez le génotype non embryogène C15. Le profil d'expression de ces gènes a été suivi par RT-PCR. Les résultats confirment l'expression différentielle de 14 gènes ; 2 chez K59 et 12 chez C15. Le profil d'expression a également été analysé pour 24 autres gènes, sélectionnés pour leur implication durant les phases précoces de l'embryogénèse somatique ou pour leur forte représentation dans les banques. Ainsi, 18 gènes différentiellement exprimés ont été identifiés ; 10 gènes chez K59 et 8 gènes chez C15.

Nous avons considéré les gènes différentiellement exprimés chez le génotype embryogène ainsi que chez le génotype non embryogène comme des gènes candidats. Afin de localiser ces gènes sur la carte génétique F1'a, nous avons fait une recherche de SNP (Single Nucleotide Polymorphisme) au sein des séquences mises en évidence par la technique SSCP (Single Strand Conformation Polymorphisme) (Orita *et al.*, 1989 ; Andersen *et al.*, 2003).

### 4. 5. L'analyse QTL

Le terme QTL signifie l'identification, sur une zone bien précise de la carte génétique, d'un facteur intervenant dans l'explication d'un caractère quantitatif. Il est connu depuis longtemps que les variations quantitatives d'un caractère dépendent de facteurs génétiques disséminés sur tout ou partie du génome. Sax en 1923 fut le premier à trouver une corrélation entre un caractère de pigmentation du tégument de la graine du haricot avec la taille du grain. Cette corrélation a été attribuée à la liaison génétique entre un ou plusieurs facteurs intervenant sur la taille du grain et les allèles du locus de pigmentation.

#### 4.5.1. Méthodes statistiques de détection des QTL

Les méthodes statistiques utilisées dans la recherche de QTL vont fortement influencer la précision et les effets des QTL.

##### a. Méthodes statistiques simples : analyse de variance

L'approche la plus classique (modèle à marqueur individuel) consiste à détecter la présence d'un QTL à proximité d'un marqueur génétique, en comparant les moyennes phénotypiques des classes génotypiques. La présence d'un QTL sera testée contre l'hypothèse nulle d'absence de différence entre les moyennes des classes génotypiques.

Cependant, ces méthodes ne permettent pas de localiser de manière précise les locus impliqués et ne donnent pas de réponse quant à l'effet du QTL. Pour ces raisons, afin de localiser les régions chromosomiques impliquées dans la variabilité des caractères, des méthodes utilisant l'information fournie par les cartes génétiques ont été développées.

##### b. Méthode de cartographie d'intervalles

L'interval mapping (IM) développé par Lander et Botstein (1989) permet de tester la présence d'un QTL non seulement au niveau des marqueurs (comme décrits précédemment) mais aussi dans l'intervalle entre ces marqueurs. Cette

## Analyse phénotypique

méthode est basée sur le calcul de vraisemblance maximale de détecter un QTL en chaque point du génome.

$$\text{LOD} = \log_{10} V(a_1, d_1) / V(a_0, d_0)$$

Où  $V(a_1, d_1)$  est la valeur de la fonction de vraisemblance sous l'hypothèse de présence d'un QTL et  $V(a_0, d_0)$  est la valeur de la fonction de vraisemblance d'absence de QTL c'est-à-dire lorsque  $a_0=0$  et  $d_0=0$  ( $a$ , effet additif ;  $d$ , degré de dominance) (Lander and Bolstein, 1989). Ce rapport est calculé à chaque point du génome (valeur prédéfinie), en progressant d'un marqueur flanquant vers l'autre pour un intervalle donné, cette opération est renouvelée pour tous les intervalles d'un groupe de liaison. Lorsque la valeur LOD dépasse le seuil prédéfini, un QTL est déclaré présent dans l'intervalle considéré. L'optimum de la courbe correspond alors à la position la plus probable du QTL.

Chaque QTL contribue à une certaine part de la variation phénotypique totale appelée  $R^2$ , qui est la somme des carrés des écarts liés au génotype du locus marqueur à la somme des carrés des écarts totaux.

Ensuite, plusieurs approches prenant en compte les effets de chaque QTL ont été développées et appelées Multiple QTL Mapping ou MQM.

Zeng et Jansen, en 1993, proposent la méthode de Composite Interval Mapping (CIM) incluant les marqueurs présentant un effet significatif comme co-facteurs. L'effet détecté au QTL n'est plus modifié par l'effet dû à d'autres QTL, ce qui permet d'affiner la précision de détection.

Cependant, cette méthode ne prend pas en compte les interactions entre les différents QTLs. La stratégie nommée Multiple Interval Mapping (MIM) proposée par Kao *et al.* (1999) et Zeng (2000) intègre les informations d'interaction entre les différents QTL. Prenant en compte les interactions, la méthode MIM consiste à poursuivre la recherche de QTL au cours des cycles de sélection. Au sein de chaque cycle, les positions et les interactions sont réévaluées.

Plusieurs logiciels de cartographie, permettent d'enchaîner les étapes décrites précédemment pour identifier les QTL (Encadré 3). Cependant, seul le logiciel MapQTL a offert la possibilité de traiter notre population de type F1'.

## Analyse phénotypique

Enfin, lorsque l'analyse QTL a été réalisée et que les locus impliqués dans l'embryogenèse somatique ont été identifiés, nous avons recherché les gènes candidats qui co-localisent avec ces QTL. Le rôle potentiel de ces gènes candidats a ensuite été discuté.

**Mapmaker/QTL** (Lincoln, Daly et Lander) utilise la méthode du maximum de vraisemblance par cartographie d'intervalles et traite les populations F2, HD, BC et les lignées recombinantes.

**QTL cartographer** (Basten, Weir et Zeng) recherche des QTL par régression marqueur par marqueur, cartographie d'intervalles (Lander et Botstein, 1989) et cartographie d'intervalles composite (Jansen, 1993 ; Zeng, 1993) et traite les populations F2 et BC.

**MapQTL** (Van Ooijen) travaille à la suite de JoinMap, et utilise des tests non paramétriques (Lehmann, 1975), la cartographie d'intervalles (IM, Stam, 1993 ; Jansen and Stam, 1994), et la cartographie de QTL multiples (MQM, Van Ooijen et *al.*, 2002). Il traite les populations de type F2, BC (Back-Cross), HD (Haploïdes Doublés), CP (Cross-Pollinated) et les lignées recombinantes.

**Encadré 3** : Les principaux logiciels de détection de QTL

# Chapitre I

Analyse des données phénotypiques des  
individus pour leur capacité liée à  
l'embryogenèse somatique et à  
l'organogenèse

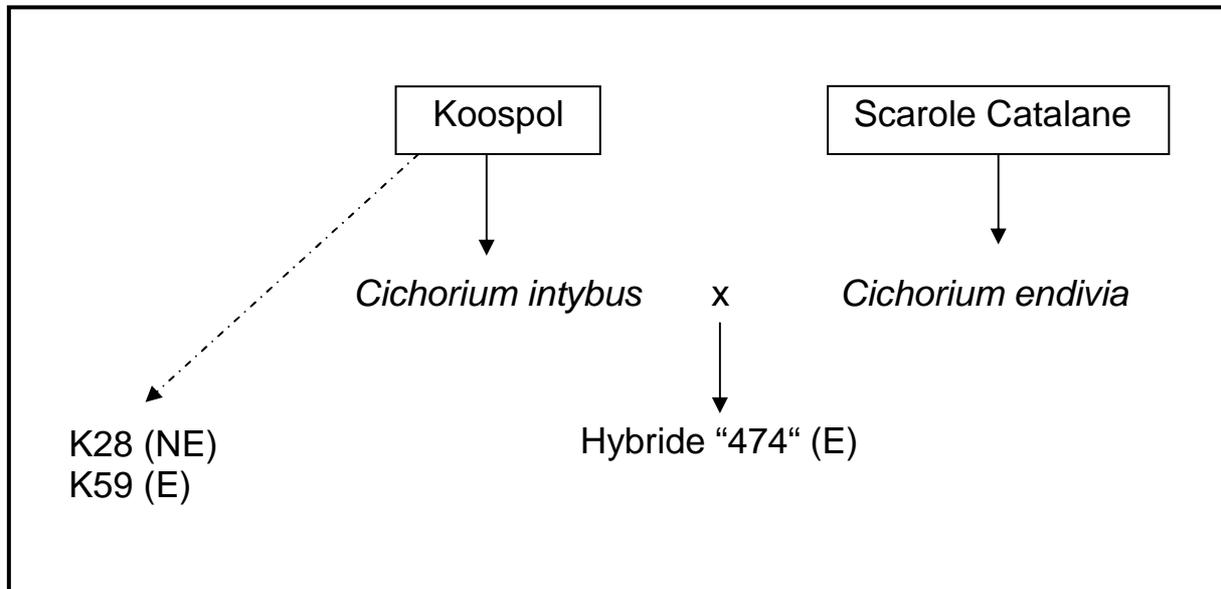
### I. Introduction

Chez les végétaux supérieurs, l'embryogenèse somatique est caractérisée par le développement d'une cellule somatique en un embryon, lequel présente une structure bipolaire avec un méristème caulinaire et racinaire, donnant naissance à un nouvel individu. Au contraire, l'organogenèse permet le développement de structures unipolaires aussi bien racinaire que foliaire. Ces deux processus sont contrôlés par des facteurs génétiques et environnementaux. L'exploitation de ces caractéristiques propres aux végétaux a ouvert la voie à de nombreuses applications en amélioration des plantes. Par exemple, elle a permis de multiplier par clonage de façon rapide des génotypes d'intérêts lorsque la reproduction sexuée est empêchée par la stérilité ou l'auto incompatibilité de l'espèce.

La chicorée est utilisée comme modèle d'étude au sein du laboratoire (SADV) depuis 1985 (Dubois *et al.*, 1988 ; Guedira *et al.*, 1989). Cette aptitude à l'embryogenèse somatique de la chicorée a fait l'objet de nombreuses études physiologiques et cytologiques.

Les premiers travaux sur l'embryogenèse somatique ont été réalisés sur l'hybride inter-spécifique "474" issu d'un croisement *Cichorium intybus* × *Cichorium endivia*, qui représentait la seule source d'embryogenèse somatique chez la chicorée (Figure 1). Le parent *C. intybus* provient de la variété Koospol (chicorée industrielle) et le parent *C. endivia* provient de la variété de scarole Catalane. Les contrôles non embryogènes utilisés pour les différentes études menées étaient des variétés commerciales de *C. intybus* génétiquement très éloignées de l'hybride embryogène. Parmi les nombreux travaux réalisés, l'utilisation de l'hybride "474" a permis de montrer l'importance de la température et l'influence de la lumière sur la formation d'embryons somatiques (Decout *et al.*, 1994). L'étude de l'hybride "474" a permis d'identifier les structures embryogènes et de caractériser les étapes de dédifférenciation, réactivation et de redifférenciation au niveau cellulaire par coloration des coupes cytologiques au bleu de toluidine de feuilles après induction en condition d'embryogenèse somatique (Decout *et al.*, 1994 ; Blervacq *et al.*, 1995).

## Analyse phénotypique



**Figure 1** : Origine des individus embryogènes (E), K59 et l'hybride "474", et de l'individu peu embryogène (NE), K28.

Plus récemment, les travaux réalisés au laboratoire (SADV) portent sur l'étude du déterminisme de l'embryogenèse somatique. Le pré-requis indispensable à cette étude génétique a été de disposer d'une variabilité génétique pour le caractère étudié, c'est-à-dire une variation phénotypique entre individus de la même espèce mais aussi d'obtenir de nombreux descendants en ségrégation issus de parents génétiquement variables. Dans le cadre d'études visant à mettre en évidence les éléments du déterminisme moléculaire du processus d'embryogenèse somatique, il est nécessaire de disposer de génotypes dont les fonds génétiques sont les plus proches et les plus homogènes possibles afin d'identifier avec une meilleure spécificité les gènes dont l'expression est associée à la réponse embryogène.

Des tentatives d'identification d'individus non embryogènes dans des descendances de l'hybride "474" se sont révélées impossibles du fait de sa forte nature interspécifique ; il est très peu fertile. Dès lors la capacité à former des embryons somatiques a été recherchée au sein d'individus de différentes variétés des 2 espèces dont est issu l'hybride "474". Aucun génotype n'a présenté une réponse à l'induction en condition d'embryogenèse somatique comparable à celle observée pour l'hybride "474" au sein de l'espèce *C. endivia*.

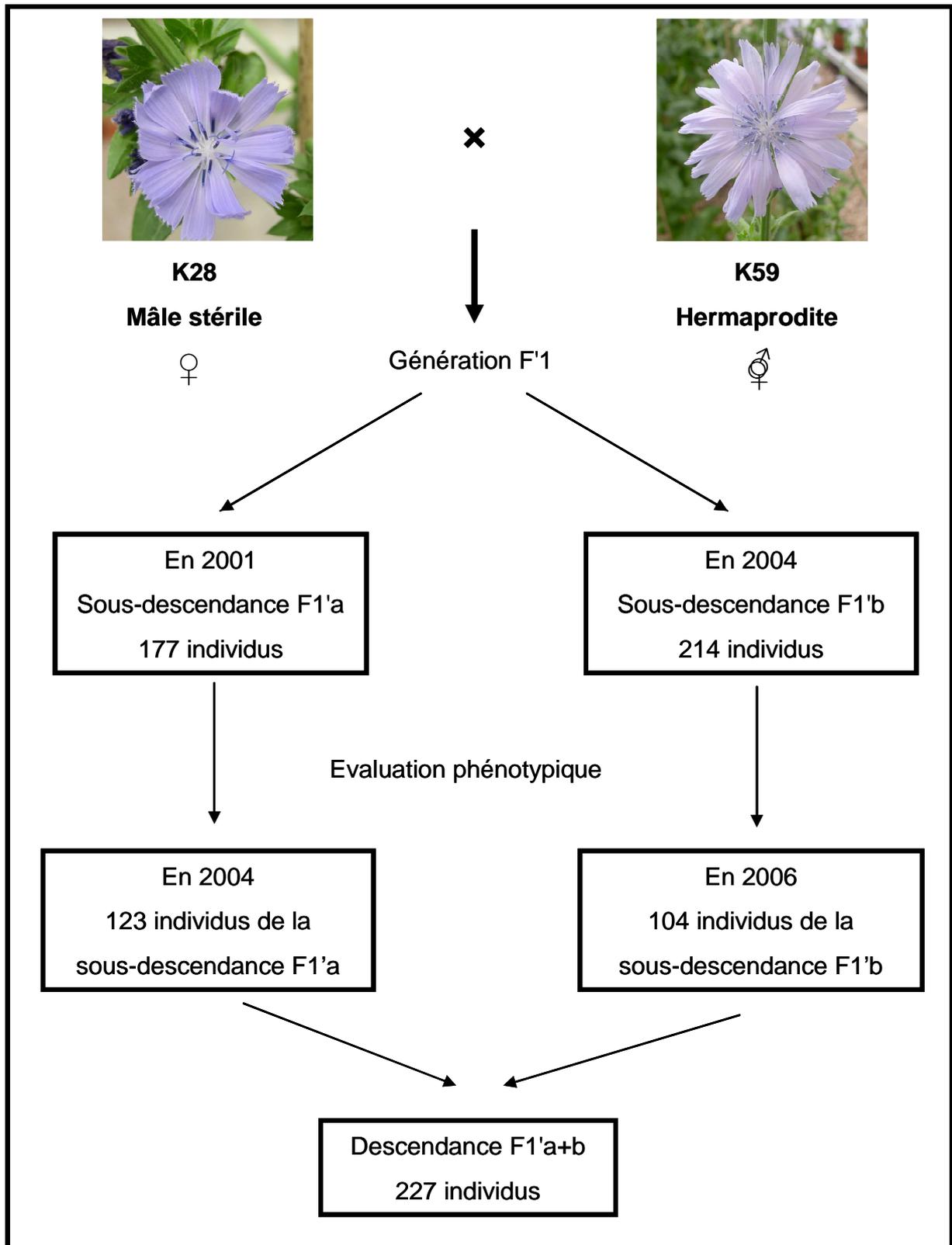
Un crible des individus de la population Koospol, à l'origine du parent *C. intybus* du génotype "474", pour leur aptitude à produire des embryons somatiques *in vitro* a été réalisé. Parmi les plantes testées, présentant des phénotypes contrastés

## Analyse phénotypique

en ce qui concerne l'intensité de l'embryogenèse somatique, des génotypes ont été identifiés (M.C Quillet et B. Delbreil, communication personnelle). Deux individus à phénotypes extrêmes ont été sélectionnés afin de générer une descendance F'1 (Figure 2). Il s'agit des génotypes K59, très embryogène et K28, peu embryogène. Le premier croisement a été réalisé en 2001 en apportant manuellement du pollen de K59 (hermaphrodite) sur les stigmates de K28 (mâle stérile). A partir de ce premier croisement, une sous-descendance F1'a de 177 individus a été obtenue et le potentiel embryogène de 123 de ces individus a été réalisé à partir des explants racinaires (MC Quillet, 2004). En 2004, de nouveaux croisements ont été réalisés entre les parents K28 et K59 afin de générer une nouvelle sous-descendance de 214 individus dont 104 individus ont été phénotypés (sous-descendance F1'b).

Ce chapitre s'articule sur l'étude des parents (K28 et K59) et des descendants (F1'a et F1'b) issus de croisement entre individus présentant une variabilité génétique pour l'embryogenèse somatique et l'organogenèse. Cette étude apportera des informations sur le degré d'héritabilité et sur la complexité du déterminisme génétique du caractère.

## Analyse phénotypique



**Figure 2** : Schéma du croisement K28\*K59 pour obtenir les sous-descendances F1'a et F1'b.

### II. Matériels et méthodes

#### 1. Matériel végétal

##### 1.1. Génération des plantules de la descendance F1'

Deux séries de croisement (en 2001 et 2004) entre les parents, K28 et K59 ont permis d'obtenir 2 lots de graines récoltées sur le parent K28 (Figure 2).

La germination du premier lot de graines obtenues en 2001 comprenant 177 individus a été réalisée *in vitro*. Pour le deuxième lot de graines, 133 graines ont été semées en serre en 2005 et 81 graines ont germé *in vitro* en 2006.

Les graines ont été stérilisées dans une solution de chlorure mercurique à 1% pendant 20 minutes puis rincées trois fois à l'eau stérilisée. Après aseptisation, les graines sont mises en germination dans des tubes contenant un milieu de Heller solide (1953) (Annexe 2) et cultivées *in vitro*. Les conditions de photopériode et de température étaient de 16h/8h et 24°C/16°C, respectivement.

Pour augmenter le nombre de plantes du 2<sup>ème</sup> lot de graines cultivées *in vitro*, des feuilles des 133 plantes cultivées en serre ont été prélevées et aseptisées. Des fragments sont découpés en incluant la nervure centrale de la feuille (zone la plus réactive) et sont déposés en boîte de pétri contenant du milieu de bourgeonnement (Bg 0,5) (Annexe 2). Seuls 23 génotypes ont pu être régénérés *in vitro* à cause de contaminations.

Le potentiel embryogène a été évalué à partir de 123 individus issus de la sous-descendance F1'a (1<sup>er</sup> lot de graines) et 104 individus de la sous-descendances F1'b (2<sup>ème</sup> lot de graine).

##### 1.2. Multiplication des individus par embryogenèse somatique et organogenèse

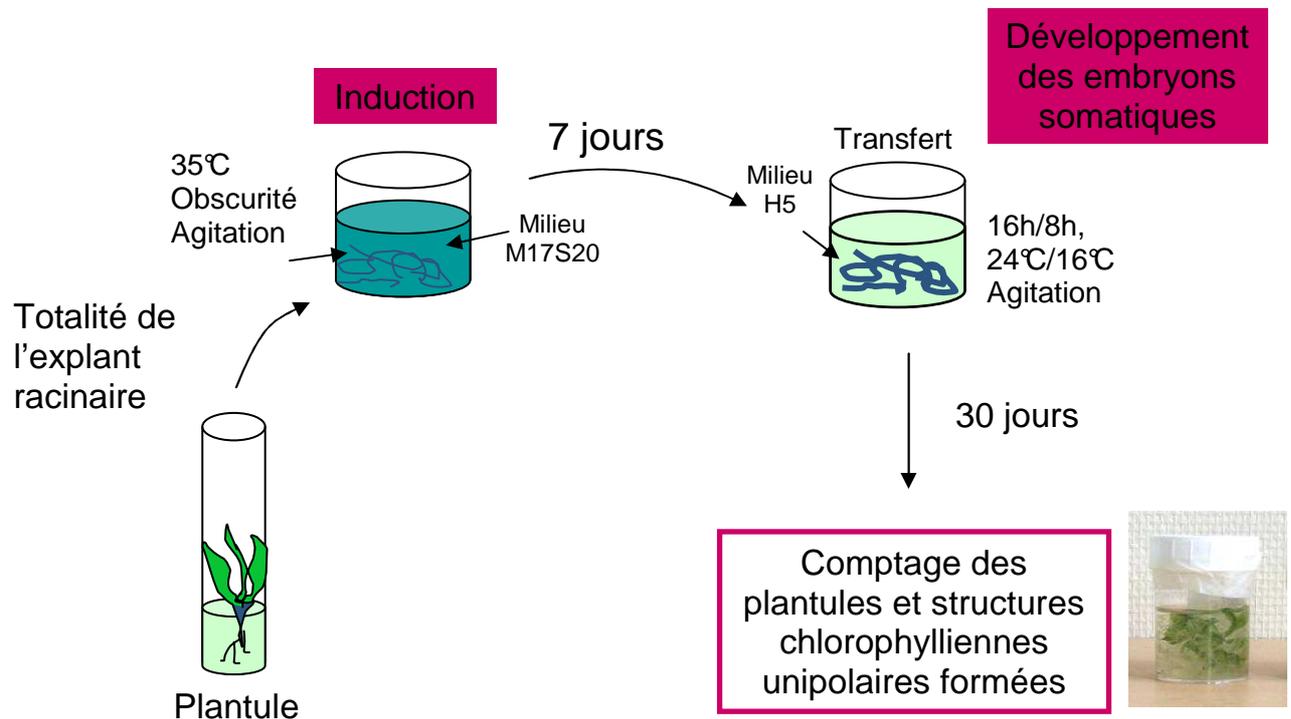
La multiplication des individus a été réalisée par embryogenèse somatique. Chaque individu a été multiplié afin d'obtenir 6 à 10 clones par génotype.

Les racines séparées de la plantule, ont été dégagées de la gélose qui l'entourait avant de réaliser l'induction de l'embryogenèse somatique (ES). Les

## Analyse phénotypique

vieilles feuilles ont été éliminées afin de ne garder que les jeunes feuilles émergentes au niveau du collet facilitant ainsi la reprise du développement de la plantule. Les collets ont été repiqués dans de nouveaux tubes de milieu H5 et remis en culture dans les mêmes conditions que ci-dessous.

La totalité des explants racinaires, d'une plantule âgée de 1 mois a été coupée puis placée dans un milieu standard de Murashige and Skoog (1962) modifié (milieu d'induction de l'ES) (Annexe 2) à 35°C, à l'obscurité sous agitation constante (80 rpm). Une semaine plus tard, les explants racinaires sont transférés dans un milieu liquide de Heller (1953), (Dubois *et al.*, 1991) (Figure 3) (Annexe 2). Les embryons somatiques formés ont été transférés en boîte de Pétri sur un milieu H5 solide en conservant la même photopériode que précédemment afin d'y poursuivre leur développement. Les embryons somatiques les plus développés ont été repiqués en tube contenant du milieu H5 solide.

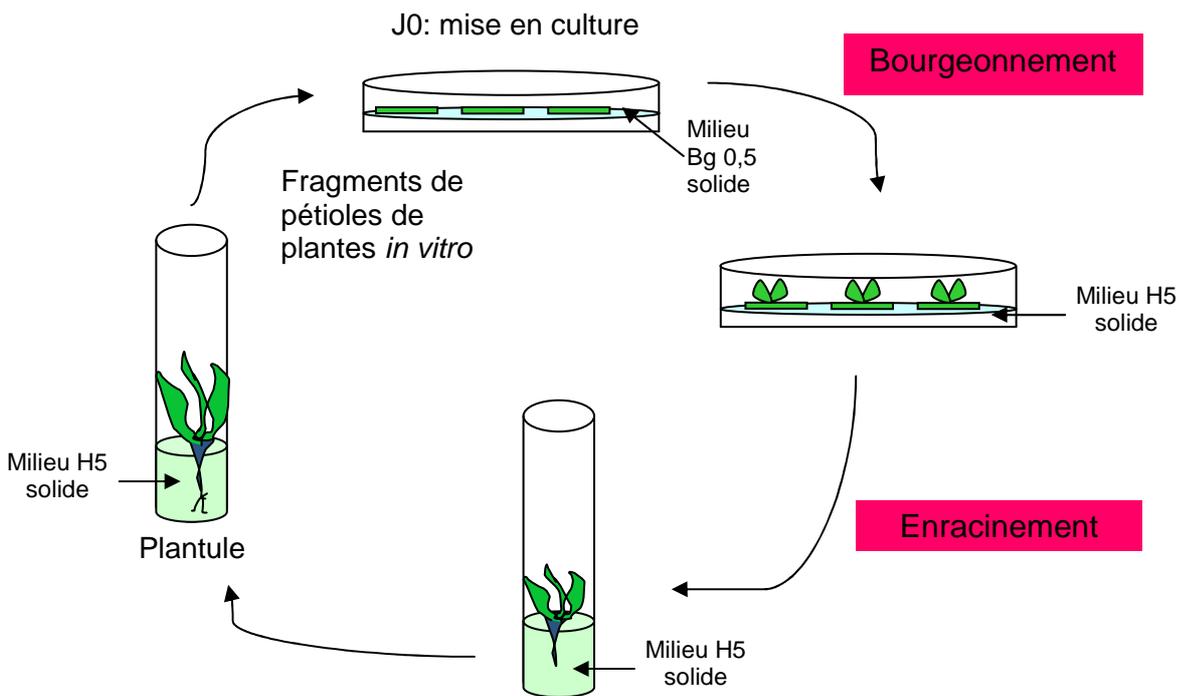


**Figure 3** : Culture en condition d'embryogenèse somatique

## Analyse phénotypique

Lorsque la capacité embryogène des génotypes était faible, les plantes ont été multipliées par organogénèse.

Pour la multiplication organogène (Figure 4), environ une dizaine de fragments de pétioles (1cm) prélevés sur les feuilles des plantules de la sous-descendance F1'b ont été mis en culture en boîte de pétri sur un milieu de bourgeonnement (bg 0,5) (Annexe 2) avec la même photopériode que les plantules. Après 10 jours de mise en culture, les calus formés ont été transférés sur un milieu H5 pour favoriser le développement et l'enracinement des bourgeons obtenus.



**Figure 4** : Culture en condition d'organogénèse

## Analyse phénotypique

### 1.3. Tests et notation du potentiel embryogène

Les tests d'embryogenèse somatique ont été effectués à partir des racines prélevées sur les plantules (Figure 3). Les plantules cultivées *in vitro* ont été prélevées lorsque les racines étaient développées, c'est à dire lorsque les racines commençaient à s'enrouler autour d'elle-même au fond du tube de culture. L'explant racinaire, prélevé sur des plantules cultivées *in vitro* présente l'avantage, par rapport à l'explant foliaire, de produire des embryons qui vont se détacher facilement de l'explant (Chapman *et al.*, 2000). Donc nous avons choisi ce type d'explant pour caractériser le potentiel embryogène des génotypes des sous-descendances F1'a et F1'b.

Après un test d'embryogenèse somatique (8 jours d'induction et 30 jours de développement), nous avons comptabilisé (voir résultats):

- Les plantules (PL), caractérisées par une structure bipolaire, avec une partie racinaire et foliaire.
- Les structures chlorophylliennes unipolaires (SH)

Pour chaque génotype des sous-descendances F1'a et F1'b, entre 10 et 26 tests d'embryogenèse somatique ont été réalisés.

### 1.4. Tests et notation du potentiel organogène

Un test d'organogenèse (Figure 4) a été réalisé à partir de fragments de pétioles sur la sous-descendance F1'b. Ce test n'a pas été réalisé sur la sous-descendance F1'a car les plantules cultivées *in vitro* n'existaient plus.

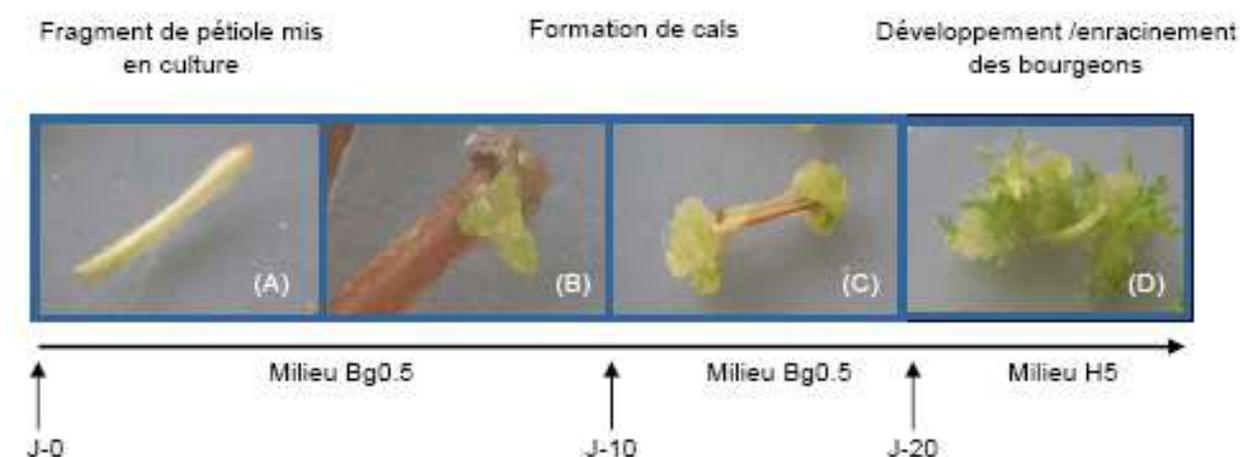
Environ une dizaine de fragments de pétioles (1cm) prélevés sur les feuilles des plantules cultivées *in vitro* ont été placés sur un milieu de bourgeonnement. Après 10 jours de mise en culture des fragments de pétioles (Figure 5A) sur un milieu de bourgeonnement, les cals formés (Figure 5B et 5C) ont été comptés pour chacun des génotypes. La formation des cals est plus prononcée au niveau de la blessure c'est-à-dire à chaque extrémité des fragments. Le dénombrement des cals a été réalisé précocement lors de leur initiation afin de faciliter le comptage.

Ensuite, les fragments de pétioles ont été transférés sur un milieu H5 de développement. Cette étape a favorisé le développement mais aussi l'enracinement des bourgeons (Figure 5D). Après 20 jours de mise en culture, les bourgeons

## Analyse phénotypique

développés ont pu être comptés. Nous avons fait le choix de compter précocement le nombre de bourgeons formés car après un délai de culture supérieur à 20 jours, il était difficile de distinguer les bourgeons développés entre eux.

Le nombre de fragments de pétioles mis en culture a pu être variable (entre 7 et 13) en fonction des plantules prélevées pour réaliser le test d'organogenèse. Pour chaque génotype, le nombre de cals et de bourgeons comptabilisés a alors été rapporté à 10 fragments de pétioles mis en culture. Les tests du potentiel organogène ont été réalisés entre 4 et 5 fois par génotype.



**Figure 5** : Développement et enracinement des bourgeons à partir des fragments de pétioles en condition d'organogenèse.

### 2. Analyse des données

Les données pour chacun des caractères étudiés, la formation de plantules (PL) à partir des embryons somatiques, les structures chlorophylliennes unipolaires (SH) et les événements totaux (ET) (voir résultats), de chaque génotype des sous-descendances F1'a et F1'b ont été inspectées (F1' a en 2004 par M.C. Quillet et F1'b en 2006 dans le cadre de ma thèse). Nous appellerons événements totaux, l'ensemble des données correspondant aux plantules développées à partir des embryons somatiques et aux structures chlorophylliennes unipolaires pour un génotype donné.

## Analyse phénotypique

Dans un premier temps, 90% des valeurs les plus élevées pour chaque génotype ont été moyennées. Les 10% éliminés correspondent le plus souvent à des valeurs nulles. Ces valeurs nulles peuvent être dues à un effet expérimental et/ou environnemental (stade de développement des racines, milieu d'induction et/ou de développement, génotype). Nous avons décidé de les écarter dans le calcul de la moyenne. Les valeurs moyennes ( $PL_{90}$ ,  $SH_{90}$  et  $ET_{90}$ ) traduiront, dans ce cas, l'intensité moyenne de réponse obtenue pour un individu.

Dans un deuxième temps, seulement 50% des valeurs les plus élevées de chaque génotype ont été moyennées ( $PL_{50}$ ,  $SH_{50}$  et  $ET_{50}$ ), traduisant ainsi le potentiel de réponse pour un individu ; cette valeur est proche de la valeur maximale obtenue pour chacun des génotypes mais, comme statistiquement, il n'est pas correcte de choisir la valeur maximale seule, nous avons choisi de moyennier 50% des valeurs les plus élevées, ce qui correspond à  $\pm 5$  tests d'embryogenèse somatique par génotype.

Pour chaque caractère, la distribution des valeurs moyennes a été représentée. L'observation des données brutes a révélé une distribution log normale. Pour tenter de normaliser les résidus, nous avons choisi de transformer les données brutes en  $\log_{10}$  (Dagnelie, 1986). Chaque valeur obtenue après comptage correspondant aux nombres de plantules et/ou aux structures chlorophylliennes unipolaires formées a été transformée en  $\log_{10}$  puis la moyenne des valeurs en  $\log_{10}$  a été calculée.

Le calcul de l'héritabilité et du coefficient de corrélation entre chaque caractère a été réalisé par un modèle d'analyse de variance en utilisant le logiciel SAS (Procédure GLM, SAS INSTITUTE 1990). L'héritabilité du caractère au sens large ( $h^2_L$ ) a été calculée selon la formule suivante :  $h^2 = \sigma^2_g / (\sigma^2_g + \sigma^2_e)$  où  $\sigma^2_g$  représente la variance génotypique et  $\sigma^2_e$  la variance expérimentale. Cette dernière prend en compte le choix de l'explant racinaire, le milieu de culture préparé, les conditions de la chambre de culture et l'expérimentateur.

### III. Résultats

Nous avons choisi d'évaluer le potentiel embryogène des individus de la descendance F1' à partir des explants racinaires. L'induction des explants racinaires mais aussi foliaires permet le développement des embryons somatiques. Cependant les embryons somatiques se détachent plus facilement à partir des explants racinaires ce qui facilite leur comptage (Chapman *et al.*, 2000).

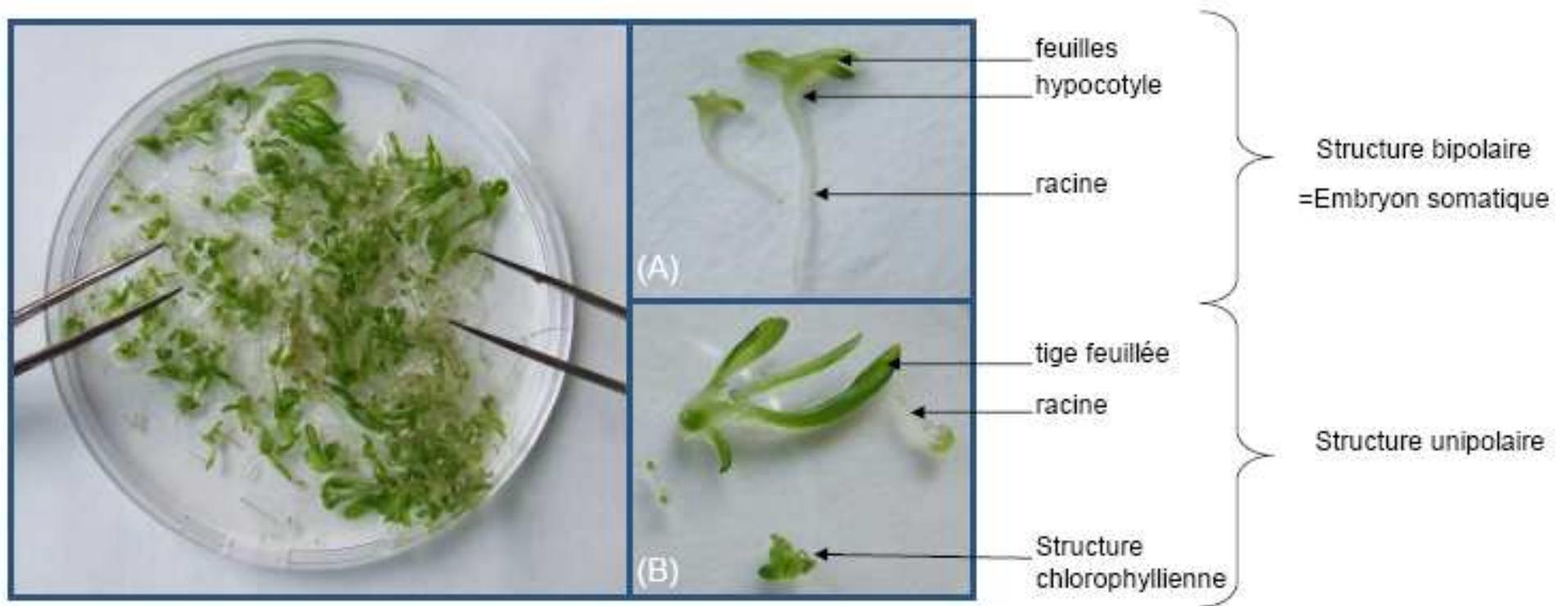
Dans notre laboratoire, la coloration au bleu de toluidine des coupes cytologiques d'explants foliaires, placés préalablement en condition d'embryogenèse somatique, nous a permis de distinguer les structures embryogènes et ainsi de les comptabiliser (Decout *et al.*, 1994 ; Blervacq *et al.*, 1995). De même, l'éclaircissement des racines permet de visualiser la formation des embryons somatiques après un montage entre lame et lamelle. Cependant, il était difficilement envisageable d'utiliser ces méthodes dans notre étude sur un grand nombre de descendants dans le temps qui nous était imparti. En effet, le nombre d'explants racinaires était trop nombreux pour réaliser pour chaque génotype et chaque répétition un comptage sous microscope. Nous avons donc défini une méthode indirecte moins laborieuse et plus rapide afin de faciliter l'identification des structures embryogènes pour chacun des génotypes. Ainsi, lors de notre étude, le dénombrement des embryons somatiques a été réalisé à partir des plantules issues du développement des embryons.

#### 1. Evaluation du potentiel embryogène des génotypes parentaux

Les premières observations des structures formées en condition ES ont été réalisées après 8 jours d'induction et 30 jours de culture des explants racinaires issus des génotypes parentaux K59 et K28. Parmi l'ensemble des événements observés, on distingue (Figure 6):

- une structure bipolaire, avec une partie racinaire et foliaire développée dans l'axe de la plantule.
- une structure unipolaire avec soit, une partie chlorophyllienne (tige feuillée ou structure chlorophyllienne), ou une partie racinaire développée.

## Analyse phénotypique



**Figure 6** : Développement de plantules (A), de structures unipolaires chlorophylliennes (B) et racinaires (B) après 8 jours d'induction en condition d'embryogenèse somatique et 30 jours de développement.

## Analyse phénotypique

Dans cette étude, seules les plantules et les structures chlorophylliennes ont été considérées et nous avons défini que l'ensemble des structures chlorophylliennes unipolaires et les plantules d'un génotype donné, constitue les événements totaux. Les structures racinaires n'ont pas été prises en compte de part la difficulté de distinguer les racines néoformées des racines issues des plantules utilisées pour l'induction. Les résultats des comptages montrent en moyenne  $363 \pm 149$  événements totaux chez le génotype embryogène K59 et  $19 \pm 15$  chez le parent peu embryogène K28. Les valeurs moyennes calculées sont regroupées dans le tableau 1.

Notre objectif principal étant d'évaluer le potentiel embryogène de chacun des génotypes, nous nous sommes alors efforcés de distinguer les structures observées. Afin de différencier les structures chlorophylliennes observées, nous avons supposé qu'une structure bipolaire, avec une partie racinaire et foliaire développée dans l'axe de la plantule, s'était développée à partir des embryons somatiques (Figure 6A). Les structures chlorophylliennes unipolaires (Figure 6B) n'étaient pas toujours faciles à distinguer entre elles. En effet, après 30 jours en condition d'embryogenèse somatique, il était possible d'observer des tiges feuillées et/ou des structures chlorophylliennes de tailles et de formes différentes. Ces observations macroscopiques pourraient indiquer que les plantules et les structures chlorophylliennes unipolaires ont une origine de développement différente. Nous les avons donc regroupées dans une classe différente de celle des plantules. Dans un deuxième temps, nous avons considéré que les structures chlorophylliennes unipolaires avaient peut-être suivi le même processus de développement que les plantules. Nous avons alors regroupé les plantules et les structures chlorophylliennes unipolaires dans une seule classe ; les événements totaux.

Ces critères ont été appliqués pour caractériser le potentiel embryogène des individus des sous-descendances F1'a et F1'b.

Ainsi pour chaque génotype et lors de chaque répétition, nous avons comptabilisé :

- les plantules issues du développement des embryons somatiques (PL)
- les structures chlorophylliennes unipolaires (SH, Shoot)

Ainsi en moyenne,  $259 \pm 148$  plantules et  $134 \pm 81$  structures chlorophylliennes unipolaires ont été comptées chez le parent embryogène, K59, contre seulement

## Analyse phénotypique

11±9 plantules et 7±6 structures chlorophylliennes unipolaires chez le parent peu embryogène, K28. La sélection de parents à phénotypes contrastés pour la capacité à former des embryons somatiques nous a permis d'obtenir une descendance F1' (227 individus) en ségrégation pour les caractères étudiés.

Pour chaque série d'expérience et chaque répétition, nous avons inclus le plus souvent possible les parents comme référence. Le test d'embryogenèse appliqué aux explants racinaires de K59 et K28 en 2006 a révélé un nombre d'évènements totaux bien inférieur à celui obtenu en 2001 pour le génotype K59. En moyenne, 57±41 évènements totaux (contre 363±149 en 2001) ont été comptabilisés pour le parent K59 et 41±30 évènements totaux (contre 19±15 en 2001) pour le parent K28. De plus, la capacité embryogène du parent K59 est fortement diminuée avec en moyenne seulement 21±18 (contre 259±148 en 2001) plantules comptabilisées. Lors de la deuxième série, les explants racinaires provenant des génotypes K59 et K28 ont été prélevés sur des plantules maintenues en culture *in vitro* depuis 6 ans. Le fait de repiquer de nombreuses fois les plantules a apparemment entraîné une diminution de la capacité embryogène du génotype parental K59. Cet effet a été observé chez l'hybride '474' après induction des explants foliaires en condition d'embryogenèse somatique (C. Rambaud, données personnelles).

Ainsi, dans la suite de l'étude, nous avons choisi de conserver les valeurs parentales obtenues en 2001 puisqu'elles reflètent mieux le potentiel embryogène de K59 et K28.

### 2. Evaluation du potentiel embryogène des individus des sous-descendances F1'a et F1'b

Les critères de discrimination des plantules et structures chlorophylliennes unipolaires définis chez les parents ont été appliqués aux individus des sous-descendances F1'a et F1'b.

Le croisement K28\*K59 en 2001 a généré une sous-descendance de 177 individus, mais l'évaluation du caractère embryogène n'a été réalisée que sur 123 individus. Pour 54 génotypes de la sous-descendance F1'a, le maintien des plantules en culture *in vitro* n'a pas été possible pour cause de contaminations.

Dans la deuxième descendance F1'b, le potentiel embryogène a été évalué pour 104 individus. Pour cette descendance, la présence de contaminations ne nous

## Analyse phénotypique

a pas permis de régénérer *in vitro* la totalité des plantes cultivées en serre. Au total, nous avons déterminé le potentiel embryogène de 227 individus (descendance F1'a+b) issus du croisement K28\*K59 (Figure 7A). Les résultats révèlent un nombre moyen d'évènements totaux compris entre 1 et 442, avec une moyenne de  $110 \pm 95$ . Parmi les individus de la descendance F1'a+b, le nombre moyen de plantules formées s'étend de 0 à 370, avec une moyenne de  $46 \pm 64$  et un nombre moyen de structures chlorophylliennes unipolaires entre 0 et 286, avec une moyenne de  $67 \pm 60$ .

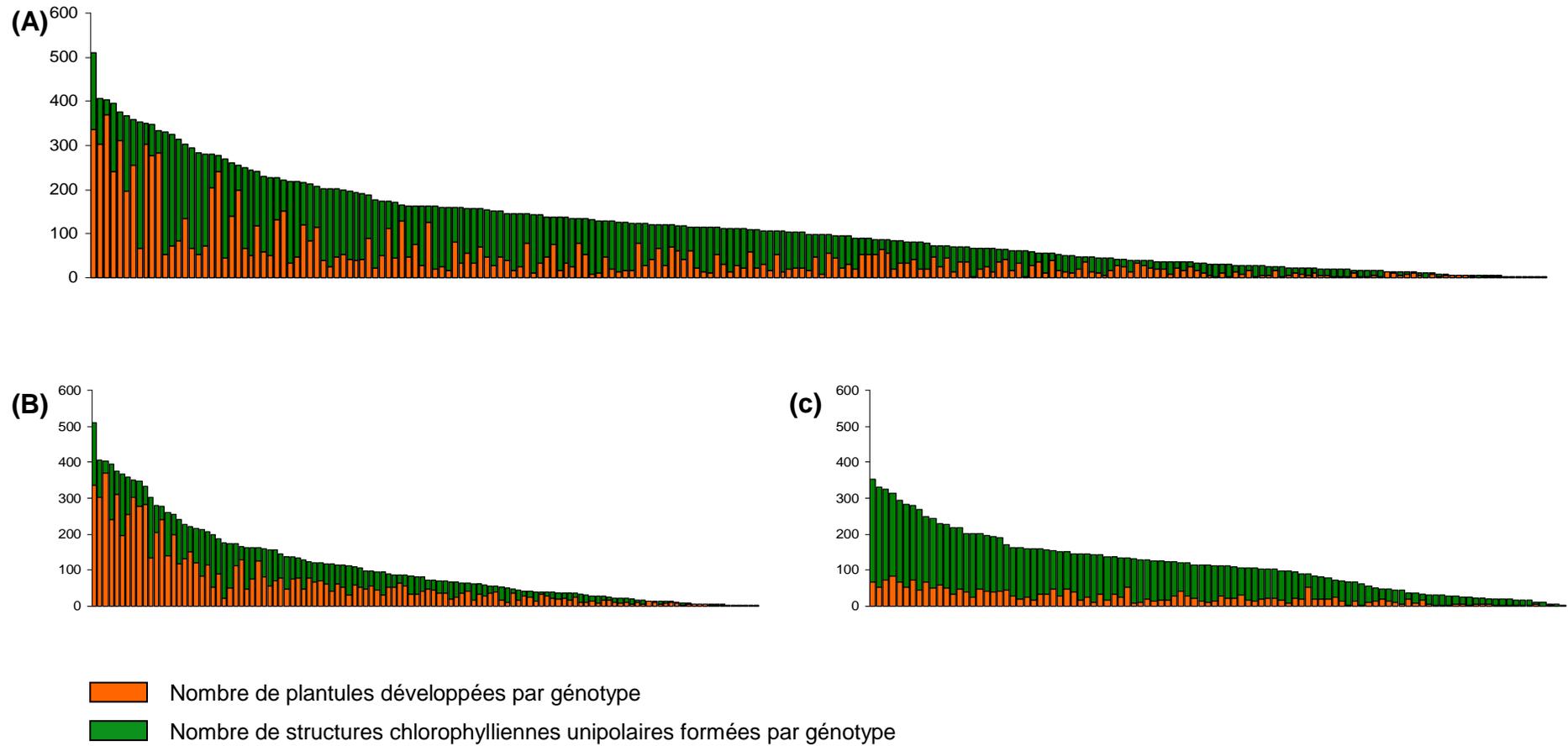
### 2.1. Comparaison des sous-descendances F1'a et F1'b

Dans un premier temps, les résultats des évaluations du niveau d'embryogenèse somatique obtenus pour les 227 individus ont été étudiés.

L'évaluation du nombre d'évènements totaux dans les deux sous-descendances révèle en moyenne  $104 \pm 105$  et  $117 \pm 82$  structures formées dans les sous- descendances F1'a (Figure 7B) et F1'b (Figure 7C), respectivement.

Cependant, l'inspection des données des comptages obtenues pour chaque génotype pour chacun des caractères (plantules et structures chlorophylliennes unipolaires) entre les deux sous-descendances a révélé plusieurs différences. D'une manière assez étonnante, lorsqu'on observe les deux histogrammes (F1'a et F1'b), la proportion de plantules développées (indiquée en orange) par rapport aux structures chlorophylliennes unipolaires (indiquée en vert) pour chaque génotype est élevée dans la sous-descendance F1'a alors que ce rapport semble être inversé dans la descendance F1'b. Entre 0 et 370 plantules formées ont été comptabilisées avec une moyenne de  $64 \pm 81$  dans la sous-descendance F1'a contre 0 et 83 plantules avec une moyenne de  $24 \pm 19$  dans la sous-descendance F1'b. A l'inverse, le nombre moyen de structures chlorophylliennes unipolaires est plus élevé dans la sous-descendance F1'b où il est compris entre 3 et 286 avec une moyenne de  $94 \pm 66$  alors que dans la sous-descendance F1'a, ce nombre s'étend de 0 à 174 avec une moyenne de  $44 \pm 43$ .

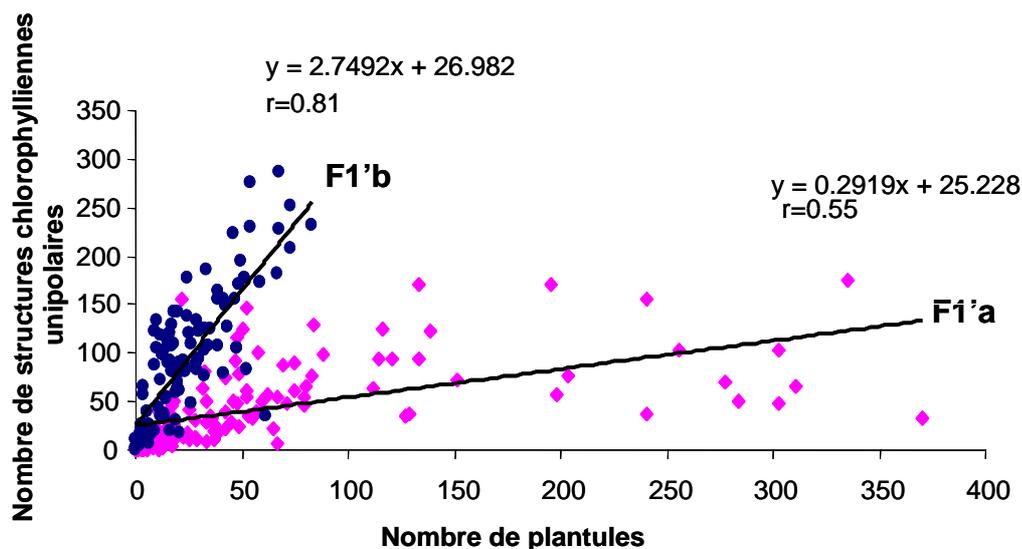
## Analyse phénotypique



**Figure 7 :** Histogrammes représentant les valeurs moyennes des plantules et structures chlorophylliennes unipolaires comptabilisées par génotype dans la descendance F1'a+b (A) et les sous-descendances F1'a (B) et F1'b (C)

## Analyse phénotypique

L'analyse des corrélations entre le nombre de plantules et le nombre de structures chlorophylliennes unipolaires formées pour chaque génotype suggère la présence d'une forte corrélation ( $r$ ) (Figure 8) entre ces deux caractères dans la sous-descendance F1'b avec un  $r$  proche de 0,81. Avec un  $r$  égale à 0,55, la corrélation est suggérée plus faible dans la sous-descendance F1'a.



**Figure 8** : Corrélation ( $r$ ) entre le nombre de plantules développées et de structures chlorophylliennes unipolaires formées dans les sous-descendances F1'a et F1'b.

L'analyse des résultats révèle plusieurs différences entre les deux sous-descendances F1'a et F1'b : d'une part un nombre de plantules formées bien plus important dans la sous-descendance F1'a et au contraire un nombre de structures chlorophylliennes unipolaires plus élevées dans la sous-descendance F1'b. De plus, lorsqu'on observe la corrélation entre ces deux caractères, on peut suggérer une plus forte corrélation ( $r=0,81$ ) dans la sous-descendance F1'b (F1'a,  $r=0,55$ ). Nous avons donc envisagé, dans la suite de l'étude, de traiter séparément les deux sous-descendances F1'a et F1'b.

## Analyse phénotypique

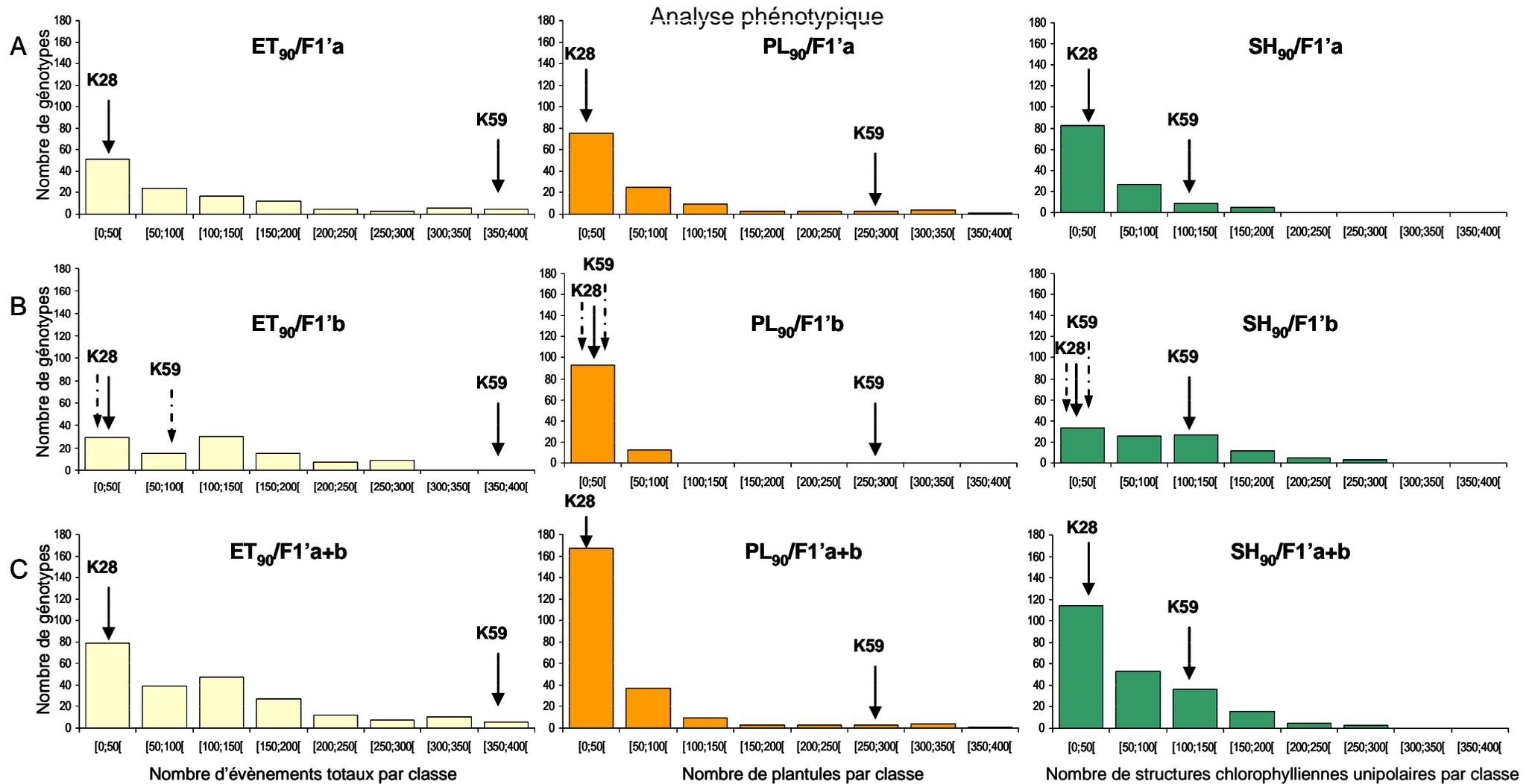
### 3. Analyse des distributions

#### 3.1. Transformation des valeurs des caractères

La normalité des distributions des variables est une condition de l'application et de l'interprétation des résultats d'un modèle d'analyse de variance (Dagnelie, 1986). Les données brutes des distributions obtenues pour les variables PL, SH et ET sont fortement déviées par rapport à la normale. Nous avons donc réalisé des transformations logarithmiques décimales des valeurs des trois variables, ce qui a permis d'homogénéiser les variances entre les génotypes non embryogènes et les génotypes très embryogènes et de diminuer l'écart entre les moyennes d'autre part. La figure 9 représente les distributions des valeurs obtenues pour les caractères PL, SH et ET avant la transformation  $\log_{10}$  (A, B, C) et après la transformation  $\log_{10}$  (D, E, F).

Les histogrammes (Figure 9) obtenus, à partir des moyennes ( $\log_{10}$ ) des comptages des plantules et des structures chlorophylliennes unipolaires dans les sous-descendances F1'a (Figure 9A) et F1'b (Figure 9B), révèlent une distribution quasi normale et continue, ce qui suggère que les caractères sont sous la dépendance de plusieurs gènes. Pour les données 90% et 50%, la moyenne des comptages des plantules et des structures chlorophylliennes unipolaires des sous-descendances F1'a et F1'b se situent entre les valeurs parentales. De plus, les histogrammes suggèrent la présence de transgressions pour les caractères liés à la formation des plantules et des structures chlorophylliennes unipolaires. Environ 30 et 26 individus ont une valeur moyenne plus faible que K28 et aucun individu suggère un valeur moyenne significative supérieur à K59 pour le nombre plantules et de structures chlorophylliennes unipolaires, respectivement, dans les sous-descendances F1'a et F1'b. Ceci suggère qu'il existe chez les deux parents des allèles intervenant pour une augmentation ou une diminution des caractères. La recombinaison se traduit par l'apparition de génotypes plus extrêmes, dûs à l'accumulation d'allèles favorables ou défavorables au sein du même génotype.

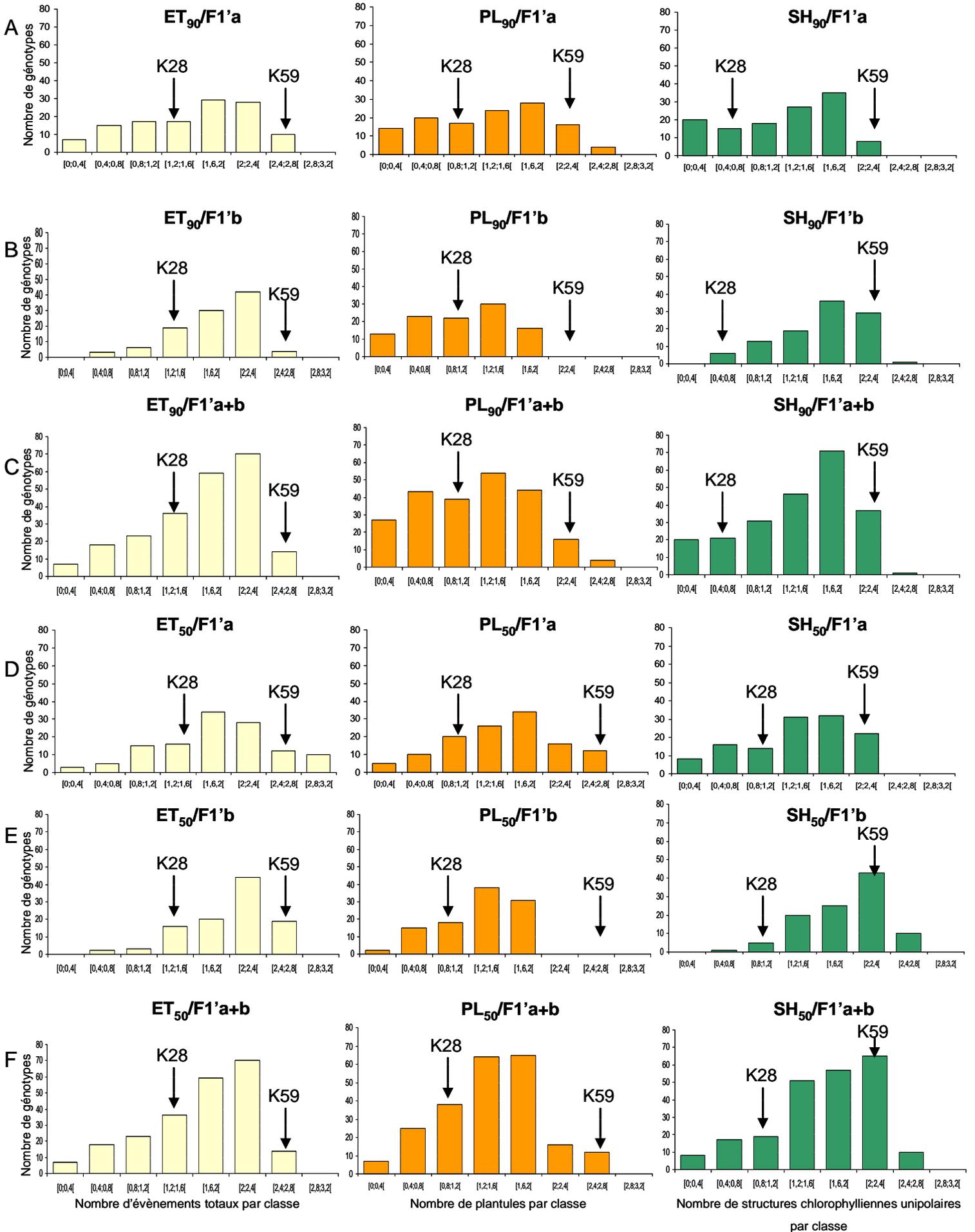
Par conséquent, les sous-descendances F1'a et F1'b sont caractérisées par une grande variabilité pour les deux caractères mesurés.



—————> Valeur moyenne obtenue pour les parents K28 et K59 au cours de la 1<sup>ère</sup> série d'induction en 2004  
 - - - - -> Valeur moyenne obtenue pour les parents K28 et K59 au cours de la 2<sup>ème</sup> série d'induction en 2006

**Figure 9 :** Distributions des valeurs moyennes obtenues pour les caractères (ET<sub>90</sub>, PL<sub>90</sub> et SH<sub>90</sub>) avant la transformation log<sub>10</sub> pour les sous-descendances F1'a (A), F1'b (B) et, la descendance F1' a+b (C). La valeur moyenne calculée pour les parents K28 et K59 pour chaque caractère (ET<sub>90</sub>, PL<sub>90</sub> et SH<sub>90</sub>) a été indiquée par une flèche.

# Analyse phénotypique



## Analyse phénotypique

### 3.2. Optimisation de la valeur moyenne des caractères étudiés

La détection de QTL nécessite d'analyser au mieux les données phénotypiques de chacun des génotypes de la descendance. Pour augmenter ce pouvoir de détection, nous avons défini une autre variable qui correspond à la moyenne des 50% des valeurs les plus élevées pour chaque génotype. Le calcul de cette variable avait pour but de discriminer le mieux possible les génotypes entre eux. Dans ce cas, la variable ( $PL_{50}$ ,  $SH_{50}$  et  $ET_{50}$ ) traduit le potentiel d'intensité de réponse de l'embryogenèse somatique d'un individu.

Les distributions obtenues pour les caractères  $PL_{50}$ ,  $SH_{50}$  et  $ET_{50}$  sont normales et continues dans les sous-descendances F1'a et F1'b et la descendance F1'a+b après transformation en  $\log_{10}$  (Figure 10, A-F). Les moyennes des trois caractères calculées pour F1'a, F1'b et F1'a+b ont été regroupées dans le tableau 1.

## 4. Variabilité génotypique et héritabilité

### 4.1. Variabilité génotypique

Les résultats obtenus par l'analyse de variance sont présentés dans le tableau 1. L'effet génotype a été calculé pour chaque variable et dans chaque descendance F1'a et F1'b. Pour chaque variable, comme dans chaque descendance, les niveaux de signification de l'effet du génotype sont très élevés ( $p (F_g) < 0.0001$ ) sur les explants racinaires. L'intensité de l'embryogenèse somatique mesurée est, donc fortement dépendante du génotype pour les deux sous-descendances.

**Figure 10 :** (A gauche) Distributions des valeurs moyennes obtenues pour les caractères (ET, PL et SH) et les deux paramètres d'analyses étudiés après transformation en  $\log_{10}$  dans les sous-descendances F1'a (A,D), F1'b (B,E) et la descendance F1'a+b (C,F). La valeur moyenne calculée pour les parents K28 et K59 pour chaque caractère et paramètre a été indiquée par une flèche.

## Analyse phénotypique

### 4.2. Héritabilité au sens large ( $h_L^2$ )

Les estimations de l'héritabilité pour les variables ( $ET_{90}$ ,  $PL_{90}$  et  $SH_{90}$ ) et ( $PL_{50}$ ,  $SH_{50}$  et  $ET_{50}$ ) sont présentées dans le tableau 1.

Les estimations de l'héritabilité pour les caractères ( $ET_{90}$ ,  $PL_{90}$  et  $SH_{90}$ ) sont plus faibles par rapport aux caractères ( $PL_{50}$ ,  $SH_{50}$  et  $ET_{50}$ ) et ceci est observable dans les deux sous-descendances. Pour le calcul des données 50%, nous avons moyenné les 5 meilleures valeurs pour chacun des génotypes. Ainsi, nous avons diminué la variation intra-génotype et par conséquent augmenté la variation inter-génotype. Dans ce cas, on assiste à une réduction de la variance ce qui explique pourquoi on a une meilleure estimation de  $h_L^2$  pour les données 50%.

Cependant, quelle que soit la méthode de calcul des données, l'héritabilité est élevée pour les 2 sous-descendances, elle est supérieure à 70% et 63% pour les descendances F1'a et F1'b, respectivement. Toutefois, nous pouvons remarquer qu'elle est légèrement plus faible dans la sous-descendance F1'b.

D'après ces résultats d'analyse de l'effet du génotype et des estimations de l'héritabilité, on peut conclure que l'intensité de l'embryogenèse somatique sur les explants racinaires présente une forte héritabilité (entre 63% et 83%) quel que soit le caractère ou la sous-descendance étudiée.

## Analyse phénotypique

		Données brutes				Données log10				héritabilité			
		K59	K28	F1'	intervalle	K59	K28	F1'	intervalle	p (F <sub>g</sub> )	σ <sub>e</sub> <sup>2</sup>	σ <sub>g</sub> <sup>2</sup>	h <sub>L</sub> <sup>2</sup> (%)
<b>F1'a</b>	PL <sub>90</sub>	259±148	11±9	64±81	0-370	2.34±0.25	0.98±0.30	1.30±0.65	0.06-2.53	<0.0001	0.150	0.403	73
	SH <sub>90</sub>	134±81	7±6	44±43	0-174	2.05±0.29	0.79±0.29	1.20±0.62	0.02-2.22	<0.0001	0.158	0.387	70
	ET <sub>90</sub>	363±149	19±15	104±105	0-442	2.51±0.24	1.20±0.29	1.54±0.66	0.08-2.58	<0.0001	0.162	0.420	72
	PL <sub>50</sub>	356±130	15±10	85±109	0-498	2.51±0.16	1.11±0.27	1.54±0.61	0.11-2.70	<0.0001	0.066	0.344	83
	SH <sub>50</sub>	165±71	10±6	57	0-256	2.18±0.19	0.99±0.22	1.42±0.59	0.03-2.39	<0.0001	0.082	0.323	79
	Et <sub>50</sub>	477±77	27±15	141	0-646	2.67±0.07	1.40±0.21	1.78±0.62	0.13-2.81	<0.0001	0.092	0.967	79
<b>F1'b</b>	PL <sub>90</sub>	21±18	5±3	24±19	0-83	1.12±0.39	0.70±0.27	1.04±0.49	0-1.87	<0.0001	0.133	0.212	63
	SH <sub>90</sub>	43±26	38±28	94±66	3-286	1.58±0.23	1.49±0.28	1.66±0.49	0.42-2.41	<0.0001	0.120	0.218	65
	ET <sub>90</sub>	57±41	41±30	117±82	3-354	1.67±0.27	1.53±0.31	1.85±0.43	0.44-2.50	<0.0001	0.084	0.169	67
	PL <sub>50</sub>	22±18	6±3	34±27	0-115	1.20±0.37	0.84±0.15	1.29±0.46	0-1.99	<0.0001	0.074	0.196	74
	SH <sub>50</sub>	56±26	41±28	129±91	6-413	1.71±0.18	1.54±0.28	1.91±0.42	0.74-2.59	<0.0001	0.053	0.157	75
	Et <sub>50</sub>	77±42	41±30	160±113	6-514	1.83±0.22	1.61±0.26	2.00±0.44	0.58-2.69	<0.0001	0.054	0.169	76
<b>F1'a + b</b>	PL <sub>90</sub>	259±148	11±9	46±64	0-370	2.34±0.25	0.94±0.26	1.18±0.60	0-2.53	<0.0001	0.144	0.332	71
	SH <sub>90</sub>	134±81	7±6	67±60	0-286	2.05±0.29	1.42±0.30	1.41±0.61	0.02-2.41	<0.0001	0.144	0.320	70
	ET <sub>90</sub>	363±149	19±15	110±95	0-442	2.51±0.24	1.42±0.29	1.68±0.59	0.08-2.58	<0.0001	0.133	0.325	72
	PL <sub>50</sub>	356±130	15±10	61±86	0-498	2.51±0.16	1.01±0.25	1.43±0.55	0-2.70	<0.0001	0.069	0.280	80
	SH <sub>50</sub>	165±71	10±6	90±83	0-413	2.18±0.19	1.24±0.35	1.64±0.57	0.03-2.59	<0.0001	0.071	0.258	78
	Et <sub>50</sub>	477±77	27±15	150±135	0-646	2.67±0.07	1.45±0.27	1.88±0.56	0.13-2.81	<0.0001	0.064	0.293	82

**Tableau 1** : Données phénotypiques (avant et après transformation en log10), effet génotypique (σ<sub>e</sub><sup>2</sup>) et héritabilité (h<sub>L</sub><sup>2</sup>%) calculés pour chaque caractère (PL, SH et ET) et paramètre d'analyse pour les parents (K28 et K59) et pour les sous-descendances F1'a, F1'b et la descendance F1'a+b.

## Analyse phénotypique

### 5. Evaluation du potentiel organogène à partir des fragments de pétioles

Les génotypes non embryogènes ou très peu embryogènes ont été multipliés par organogénèse. Pendant la multiplication, une variation quantitative a été observée lors de la formation des cals et des bourgeons pour ces génotypes. Nous nous sommes alors intéressés à évaluer le potentiel organogène de ces génotypes.

De plus, la collecte de ces données avait pour objectif, après analyse QTL, de déterminer la présence ou non de QTL communs pour la formation des plantules et des structures chlorophylliennes unipolaires issues de l'embryogénèse somatique et, la formation des cals et des bourgeons issus de l'organogénèse.

#### 5.1. Evaluation du potentiel organogène des génotypes parentaux

La mise en culture des fragments de pétioles prélevés à partir des plantules de K59 et K28 a révélé un forte organogénèse pour K59 contrairement à K28 (Tableau 2). Ainsi en moyenne  $10 \pm 5$  cals et  $16 \pm 7$  bourgeons ont été comptabilisés pour K59 (Tableau 2). Pour K28, aucune formation de cal donc aucun développement de bourgeon n'a été observé sur les fragments de pétioles mis en culture.

#### 5.2. Evaluation du potentiel organogène des génotypes de la sous-descendance F1'b

Le même principe de calcul, que celui utilisé pour analyser les résultats des tests d'embryogénèse somatique a été appliqué aux comptages des cals et bourgeons : La moyenne des valeurs obtenues pour chaque génotype a été calculée avant (Figure 11A) et après transformation (Figure 11B) des valeurs en  $\log_{10}$ .

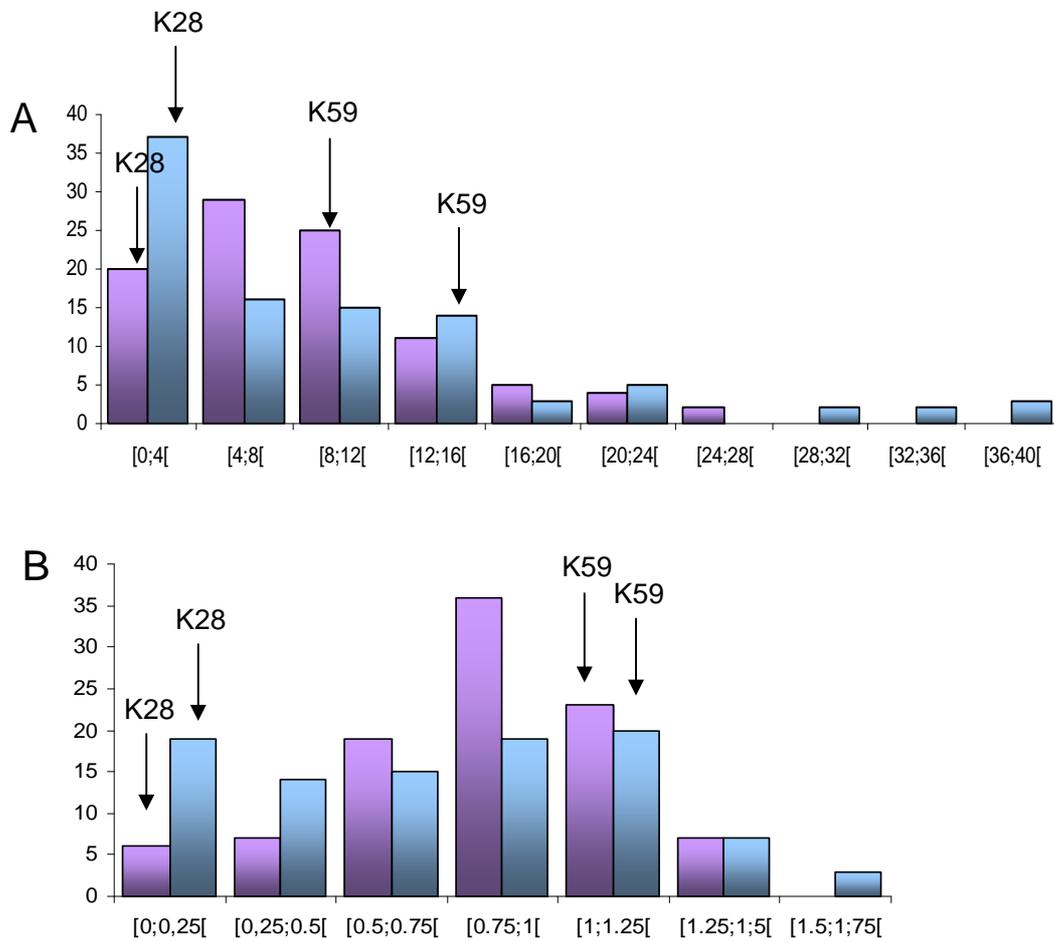
L'histogramme obtenu à partir des moyennes en  $\log_{10}$  (Figure 11B) des comptages des cals et bourgeons au niveau des fragments de pétioles dans la sous-descendance F1'b révèle une distribution normale et continue. Comme pour l'embryogénèse somatique, l'organogénèse semble être sous la dépendance de plusieurs gènes.

## Analyse phénotypique

	Données brutes				Données log10				héritabilité			
	K59	K28	F1'	intervalle	K59	K28	F1'	intervalle	p (F <sub>g</sub> )	$\sigma_e^2$	$\sigma_g^2$	h <sub>L</sub> <sup>2</sup> (%)
Cal	10±5	0	9±6	0-24	1.02±0.19	0	0.83±0.30	0-1.39	<0.0001	0.574	0.30	35
Bg	16±7	0	9±10	0-38	1.20±0.15	0	0.72±0.43	0-1.59	<0.0001	0.480	0.82	65

**Tableau 2 :** Données phénotypiques (avant et après transformation en log10), effet génotypique ( $\sigma_e^2$ ) et héritabilité ( $h_L^2$ ) calculés pour chaque caractère (Cal et Bg) et paramètre d'analyse chez les parents (K28 et K59) et la sous-descendance F1'b.

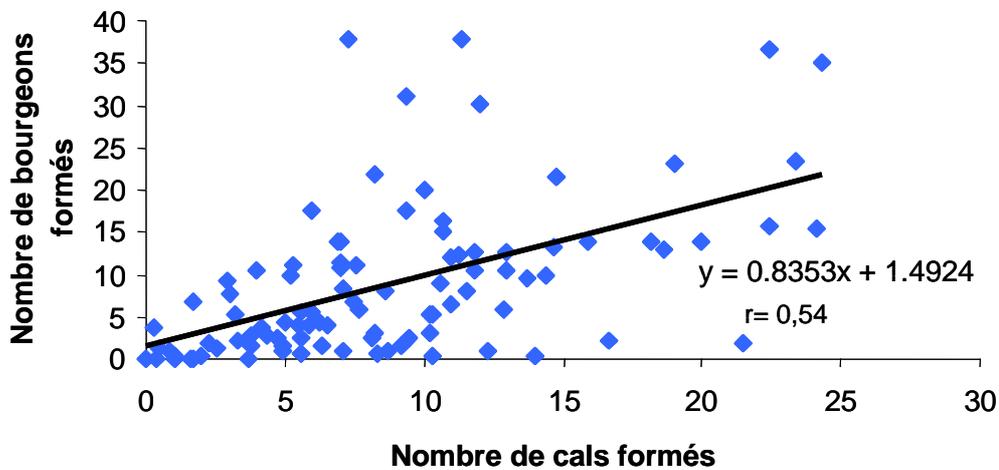
## Analyse phénotypique



**Figure 11** : Distributions des valeurs moyennes obtenues pour la formation des cals et le développement des bourgeons dans la sous-descendance F1'b avant (A) et après (B) transformation en log10.

## Analyse phénotypique

Pour les deux caractères étudiés, 33 et 14 géotypes suggère une valeur moyenne plus élevée que K59 et, 1 et 6 géotypes ont une valeur moyenne égale à 0 comme le géotype parental K28 pour la formation des cals et le développement des bourgeons, respectivement. La corrélation entre les deux caractères étudiés, déterminée à partir du coefficient de détermination est relativement faible avec un  $r=0.54$  (Figure 12).



**Figure 12** : Corrélation ( $r$ ) entre le nombre de cals et de bourgeons formés dans la sous-descendance F1'b.

### 5.3. Calcul de l'héritabilité et effet génotypique

De la même manière que pour les explants racinaires, l'effet génotype a été calculé pour la formation des cals et des bourgeons au niveau des fragments de pétioles. Les résultats obtenus par l'analyse de variance sont regroupés dans le tableau 2.

Pour chaque caractère étudié, les niveaux de signification de l'effet du génotype sont très élevés ( $p(F_g) < 0.0001$ ) pour la formation des cals et le développement des bourgeons sur les fragments de pétioles. Ainsi, comme pour le processus d'embryogenèse somatique, l'organogenèse est fortement dépendante du génotype dans la sous-descendance F1'b. De même, l'héritabilité estimée à 65%

## Analyse phénotypique

pour le développement des bourgeons est proche de celle obtenue pour le processus ES. Au contraire, avec 35%, l'héritabilité calculée pour la formation des cals est beaucoup plus faible et très différente de celle obtenue pour la formation des bourgeons. Toutefois, même si les valeurs expérimentales ( $\sigma^2_e$ ) pour les cals (0,574) et les bourgeons (0,480) sont comparables, les valeurs génotypiques ( $\sigma^2_g$ ) sont différentes. En effet, la variance génotypique des cals (0,300) est très faible par rapport à celle des bourgeons (0,820). Le comptage des cals s'est avéré difficile dans certains cas notamment pour les génotypes très callogène. Pour pallier à cette difficulté, un comptage précoce des cals a été réalisé. Cette démarche ne nous a peut-être pas toujours permis de bien distinguer les cals formés entre eux et ainsi d'évaluer au mieux le potentiel callogène pour chacun des génotypes d'où une héritabilité plus faible.

### IV. Discussion

Les parents choisis pour cette étude varient fortement dans leur potentiel embryogène et organogène, et les deux sous-descendances F1'a et F1'b développées à partir de ces deux parents ségrégent pour les caractères étudiés : l'embryogenèse somatique et l'organogenèse.

Une première sous-descendance F1'a a été obtenue en 2001 et la capacité embryogène de 123 individus de cette sous-descendance a été déterminée en 2004 (M.C. Quillet, données personnelles). Cependant, la recherche de QTL nécessite un nombre élevé de descendants pour accroître la puissance de détection et la précision de la localisation des QTL, alors nous avons choisis de réaliser d'autres croisements à partir des génotypes parentaux K28 et K59 générant ainsi 1333 nouveaux individus correspondant à F1'. Le potentiel embryogène de 104 de ces individus a été évalué en 2006. Les données phénotypiques après transformation en  $\log_{10}$  ont révélé une distribution normale et continue. Le nombre d'événements totaux entre les deux sous-descendances F1'a et F1'b est comparable alors qu'au contraire le nombre de plantules et structures chlorophylliennes s'est avéré différents entre F1'a et F1'b. En effet, les descendants de F1'b semblent moins embryogènes que les descendants de F1'a. De plus, la corrélation entre le nombre de plantules et les structures chlorophylliennes unipolaires formées s'est avérée plus grande avec  $r=0,81$  dans F1'b comparée à F1'a avec  $r$  égale à  $0,55$ . Afin de comprendre les différences de potentiel embryogène entre F1'a et F1'b, nous avons réfléchi à plusieurs paramètres environnementaux et expérimentaux qui pourraient expliquer la présence d'une telle variabilité. Ces paramètres seront présentés et discutés en parallèle avec les données génotypiques (Chapitre II) dans le dernier chapitre "Discussion générale et Perspectives".

Toutefois, nous avons décidés dans la suite de notre travail de prendre en compte ces différences c'est-à-dire d'une part, de traiter séparément les deux sous-descendances lors de la construction de la carte génétique mais aussi lors de la recherche de QTL et d'autre part, de traiter ces deux sous-descendances comme une seule descendance F1'a+b afin de voir si les différences phénotypiques observées ont un effet sur la détection des QTL.

## Analyse phénotypique

Une forte héritabilité pour chaque variable liée à l'embryogenèse (PL) et l'organogenèse (Bg) a été calculée et, celle-ci est en concordance avec les estimations décrites dans la littérature. La présence d'une forte héritabilité traduit la présence de locus majeurs contrôlant la variabilité du nombre de plantules et de bourgeons formés.

Cependant, pour les caractères liés à l'embryogenèse somatique peu d'études sur l'héritabilité du caractère ont été réalisées. Chez le tournesol, Flores Berrios et ses collaborateurs (2000) ont estimé l'héritabilité au sens large pour la formation des embryons somatiques à 0,64 (nombre d'embryons formés pour 40 explants cultivés) et 0,77 (nombre d'explants embryogène pour 40 explants cultivés) à partir de lignées recombinantes.

Au contraire, de nombreux articles ont estimé l'héritabilité liée à l'organogenèse et à la callogenèse. Toujours chez le tournesol, les estimations de l'héritabilité pour la capacité de régénération *in vitro* à partir d'explants cotylédonaire s'élèvent à 0,72 et 0,70 (Flores Berrios et al., 2000). De même, chez le riz (Tagushi et al., 1997) l'héritabilité liée à la régénération *in vitro* a été estimée à partir de graines matures dans une descendance back cross. Pour le nombre de bourgeons régénérés et le pourcentage de cals régénérés par rapport aux nombres total de cals formés, l'héritabilité s'élèvent à 0.919 et 0.736, respectivement. Chez l'orge (Mano et al., 2002), une valeur  $h_L^2$  de 0.804 a été estimée pour le nombre de bourgeons différenciés par cal à partir d'une lignée recombinante. Dans notre étude, l'estimation de l'héritabilité pour la formation des bourgeons à partir de fragment de pétioles dans une descendance F1' est légèrement plus faible avec 0,65.

Par contre, l'héritabilité de 35% que nous avons estimé pour la formation des cals par rapport aux estimations décrites dans la littérature est faible. Par exemple, à partir d'une lignée recombinante chez le maïs, les auteurs ont calculé une valeur  $h_L^2$  de 0.52 pour l'initiation des cals à partir d'embryons immatures (Krakowski et al., 2006).

Cependant, pour un caractère donné, les estimations de l'héritabilité peuvent varier selon le dispositif expérimental, le modèle végétal et la descendance étudiée, ce qui rend difficile la comparaison. Chez le blé, la réponse à l'induction de cals à partir d'anthers (embryogenèse haploïde) est un caractère fortement influencé par les conditions de cultures des plantes sur lesquelles son prélevés les plantes. Pour une même descendance F2, on observe une héritabilité au sens large qui varie de 0

## Analyse phénotypique

à 0.34, selon le milieu de culture des plantes. Chez le maïs, Murigneux et ses collaborateurs (1994) ont étudié deux descendance, l'une de type haploïde doublée, l'autre de type back cross, toutes les deux issues des mêmes lignées parentales. Ils ont estimé pour les mêmes critères d'androgenèse *in vitro*, mesurés dans les mêmes conditions, une valeur de  $h_L^2=0.90$  pour la descendance haploïdes doublées et une valeur de  $h_L^2=0.78$  pour la descendance back cross. Cette différence d'estimation de l'héritabilité entre les deux descendance est expliquée par l'état homozygote des lignées haploïdes doublées qui permet d'avoir une meilleur estimation des facteurs génétiques qui contribuent à l'expression du génotype.

## Chapitre II

Construction d'une carte génétique F1'a+b  
pour la détection de QTL

### I. Introduction

La disponibilité d'une carte génétique dense est un outil essentiel pour des études génétiques détaillées. En outre, son exploitation est indispensable à l'identification de QTL (Quantitative Trait Loci) afin de fournir des gènes candidats pour la sélection assistée par marqueurs ou encore faciliter et mieux contrôler l'introgession de gènes d'intérêt agronomique.

Le développement de marqueurs moléculaires dominants, essentiellement RAPD et AFLP, a permis de construire deux cartes génétique chez la chicorée. La première a été construite à partir d'une population outbred interspécifique F1' entre chicorée et une lignée de *Cichorium endivia*. La carte est couverte par 351 marqueurs (AFLP, RAPD et SAMPL) dispersés sur 13 groupes de liaisons (1204 cM) (de Simone *et al.*, 1997). La deuxième est basée sur une population F2 intervariétale dérivée d'un croisement entre deux chicorées « witloof ». La carte est couverte par 129 marqueurs RAPD sur 9 groupes de liaison (609 cM) (Van Stallen *et al.*, 2003). Par la suite, la carte a été complétée par des marqueurs AFLP afin de détecter des QTL d'intérêts (Van Stallen *et al.*, 2005a ; Van Stallen *et al.*, 2005b).

Dans le cadre du projet Gis-Cartochic, impliquant l'Université des Sciences et Technologies de Lille1 (USTL) et des partenaires privés (Florimond Desprez Veuve & Fils SAS, Leroux SAS et SARL Hoquet Endives), une carte génétique consensus (881,7 cM) de la chicorée ( $2n=2x=18$ ) a été construite à partir des données des marqueurs moléculaires de 3 descendances (Cadalen *et al.*, 2009). La première, appelée K28\*K59, est une population outbred de 177 individus F1' obtenue après croisement entre deux génotypes hétérozygotes K28 et K59 (USTL, laboratoire SADV, France). La deuxième population, appelée Rubis 118, est une F2 de 96 descendants obtenue après croisement entre un clone mâle stérile MS8 et un cultivar de chicorée industrielle "Cassel" (Florimond-Desprez, Capelle-en-Pevele, France). La troisième, appelée BR, est une population F2 de 145 descendants obtenue à partir d'un croisement entre 2 lignées parentales de chicorée witloof (Hoquet Endives, Raillencourt, France). Un hybride F1 a été autofécondé pour générer la population.

Trois banques enrichies en motifs SSR (Simple Sequences Repeats) [Agrogène (AGR), France ; Ecogenics (SW), Suisse ; SADV (EU), France] ont été produites à partir d'ADN génomique résultant de 420 marqueurs AGR, 719

marqueurs SW et 1251 marqueurs EU. De plus, 81 marqueurs STS (Sequences Tagged Sites), incluant des séquences EST (Expressed Sequences Tags) et les gènes isolés au laboratoire, ont été utilisés. Les marqueurs SSR et STS sont des marqueurs codominants et ont été utilisés comme marqueurs ancrés pour l'intégration des trois cartes.

Mon travail a consisté à améliorer la densité de la carte K28\*K59 en augmentant le nombre de marqueurs cartographiés. En effet, il s'agit de rechercher du polymorphisme au sein de marqueurs SSR issus de la banque SW afin de les cartographier. Des marqueurs AFLP générés en adaptant le protocole de Vos et ses collaborateurs (Vos *et al.*, 1995) pour la chicorée (C. Blassiau, M.C Quillet, L. Gonthier, données non publiées) ont été inclus à la carte génétique, couvrant les 9 groupes de liaison. Enfin, la recherche de polymorphisme au sein des séquences EST mis en évidence au 4<sup>ème</sup> jour d'induction de l'embryogenèse somatique par analyse microarray (Legrand *et al.*, 2007) a été entreprise. Les séquences polymorphes ont été cartographiées en vue d'une analyse QTL liée à l'embryogenèse somatique (Chapitre 3). Ces gènes impliqués dans les phases précoces de l'embryogenèse somatique peuvent s'avérer de bons gènes candidats dans la détection de QTL liée à l'embryogenèse somatique.

Pour réaliser l'analyse QTL, des marqueurs SSR, STS et AFLP présents sur la carte génétique F1'a ont été sélectionnés selon plusieurs critères et génotypés à partir des ADN de la descendance F1'b afin de construire une carte génétique F1'a+b.

## II. Matériels et méthodes

### 1. Extraction d'ADN à partir de plantes cultivées en serre

L'extraction de l'ADN a été réalisée au moyen du kit GenElute Plant genomic DNA kit (Sigma) selon les instructions du fabricant. Environ 100 mg de tissus foliaires frais, prélevés sur des plantes cultivées en serre, ont été broyés dans de l'azote liquide, puis incubés 10 min à 65°C dans 400 µl d'un mélange de tampon de lyse. Après précipitation des débris cellulaires, des protéines et des polysaccharides, par ajout de 130 µl d'un tampon de précipitation et centrifugation 5 min à 15000 g, le surnageant a été filtré. L'ADN a été purifié par fixation sur une colonne

préalablement équilibrée puis élué, après lavage de la colonne, par deux fois 100 µl d'une solution d'éluion. La concentration de la solution d'ADN ainsi obtenue a été déterminée par spectrophotométrie à 260 nm.

### 2. Détection de polymorphisme au sein des séquences

L'objectif de ce travail a été de détecter des variations de type insertion-délétion (InDel) ou des variations de séquences qui, lorsqu'elles concernent une ou quelques bases sont appelés SNP (*Single Nucleotide Polymorphism*).

#### 2.1. Les microsatellites ou SSR

Afin de générer des marqueurs co-dominants, 3 banques (AGR, SW, EU) enrichies en motifs SSR (motif di-nucléotidique de type CA et GA) ont été construites à partir des séquences d'ADN génomique résultant de 420 séquences AGR, 719 séquences SW et 1251 séquences EU (Cadalen *et al.*, 2009 ; Annexe 1). Le niveau d'enrichissement a été élevé (75-85%), et la proportion de séquences SSR non redondantes a été similaire pour les 3 banques (42-57%), avec un total de 730 séquences SSR disponibles pour la recherche de polymorphisme. Des amorces ont été définies autour des motifs SSR détectés (Cadalen *et al.*, 2009).

##### 2.1.1. Définition des amorces

Pour chaque paire d'amorces définies, une queue (ou tail) M13F (5'-CACGACGTTGTAAAACGAC) ou M13R (5'-GGATAACAATTTTCACACAGG) a été rajoutée à l'une des deux amorces, suivant la procédure dite " Tailling " de Oetting *et al.* (1995).

##### 2.1.2. Réaction PCR

Les réactions PCR ont été réalisées dans un volume de 15 µl contenant:

- Tampon PCR (1X)
- MgCl<sub>2</sub> (2 mM)

## Carte Génétique

- dNTP (100 µM de chaque)
- BSA (0,2 mg/ml)
- Amorces étendues par une queue M13 (33 nM)
- Amorce non étendue (333 nM)
- Amorce étendue par une queue M13 liée à un fluorochrome (6'Fam, Vic, Ned et Pet)
- Taq polymérase (Applied Biosystems) (0,3 U)
- ADN (2 ng)

Les conditions PCR ont été les suivantes :

Cinq minutes de dénaturation à 94°C suivies de,

10 cycles :

- Dénaturation : 30 secondes à 94°C
- Hybridation des amorces : 35 secondes à T<sub>m</sub>
- Diminution de -1°C/cycle jusqu'au 10<sup>ième</sup> cycle
- Elongation : 45 secondes à 72°C

35 cycles :

- Dénaturation : 30 secondes à 94°C
- Hybridation des amorces : 20 secondes à 56°C
- Elongation : 45 secondes à 72°C
- Extension terminale : 7 minutes à 72°C

Pour les séquences présentant un faible signal après migration par électrophorèse capillaire (ce qui rend l'analyse du profil difficile), les conditions précédentes ont été modifiées. L'amorce étendue avec une queue M13 a été remplacée par une amorce liée à un fluorochrome (6-Fam ou Hex). Dans ce cas, les conditions PCR sont identiques à celles utilisées pour les séquences ESTs.

### 2.2. Les marqueurs AFLP

La génération des marqueurs AFLP (Vos *et al.*, 1995) a été réalisée en collaboration avec Christelle Blassiau, Marie-Christine Quillet et Lucie Gonthier. Mon

travail a consisté à analyser les profils AFLP obtenus à partir des différentes combinaisons d'amorces nucléotidiques utilisées.

### 2.2.1. Digestion des ADN

Cinq cents ng d'ADN ont été digérés dans un volume de 30 µl contenant 10,5 µl de mélange de restriction [3 µL de Tampon Ligase 10 x (New England Biolabs), 1,5 µl de BSA 100 X, 3 µl de NaCl 0,5 µl d'enzyme *EcoRI* 20 U/µl (New England Biolabs) et 2,5 µl d'enzyme *MseI* 4 U/µl (New England Biolabs)]. La digestion a été réalisée pendant 3 heures à 37°C.

### 2.2.2. Ligature des adaptateurs

Cinq µl de mélange de ligature [1 µl d'adaptateur *EcoRI* 5 pmol/µl, 1 µl d'adaptateur *MseI* 50 pmol/µl, 0,5 µl de tampon ligase 10 X (New England Biolabs), 0,5 µl d'ATP 10 mM et 0,16 µl (1 U Weiss) de T4 DNA ligase (New England Biolabs)] ont été ajoutés aux 30 µl de restriction. Ce mélange réactionnel a été incubé 16h à 37°C.

Les séquences des adaptateurs sont :

- Adaptateur *MseI* : 5'-GACGATGAGTCCTGAG  
TACTCAGGACTCAT-3'
- Adaptateur *EcoRI* : 5'-CTCGTAGACTGCGTACC  
CATCTGACGCATGGTTAA-3'

### 2.2.3. Pré-amplification

La pré-amplification a été réalisée à partir d'une dilution au 1/ 30 des réactions de restriction/ ligature. Les amorces utilisées possèdent une base de sélection en 3'. Les séquences des amorces sont :

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

Les PCR ont été réalisées dans un volume de 30 µl contenant :

## Carte Génétique

- $MgCl_2$  (1,5 mM final)
- dNTP (0,2 mM final)
- Taq Polymerase (Applied Biosystems) 15 U/ $\mu$ l, (1,5 U par réaction)
- Amorces *EcoRI*-A (8 pmol par réaction)
- Amorces *Msel*-C (8 pmol par réaction)

Réaction de restriction/ligature diluée au 1/ 30 (3  $\mu$ l par réaction)

Les conditions de PCR sont les suivantes :

30 cycles :

- Dénaturation 30 secondes à 94°C
- Hybridation des amorces 30 secondes à 56°C
- Elongation 1 minute à 72°C
- Extension terminale : 5 minutes à 72°C

### 2.2.4. Amplification sélective

Les amplifications sélectives ont été réalisées à partir de dilutions au 1/10 des produits de pré-amplification. Les amorces utilisées étaient les mêmes que celles utilisées pour la pré-amplification avec une extension en 3' de 2 bases sélectives supplémentaires. Les amorces *EcoRI* ont été liées à un fluorochrome (6-Fam ou Hex).

Les séquences de base des amorces :

- Amorces *EcoRI*-ANN : 5'-GACTGCGTACCAATTCANN-3' (avec N=A, C, G ou T)
- Amorces *Msel*-CNN : 5'-GATGAGTCCTGAGTAACNN-3' (avec N=A, C, G ou T)

Les PCR ont été réalisées dans un volume de 10  $\mu$ l contenant :

- Tampon Taq Polymérase standard (5 U/ $\mu$ l)

## Carte Génétique

- $MgCl_2$  (1,5 mM final)
- dNTP (0,2 mM final)
- Amorce *EcoRI*-ANN marquée (2 pmol par réaction)
- Amorce *Msel*-CNN (2,5 pmol par réaction)

Produit d'amplification diluée au 1/10 (3  $\mu$ l par réaction)

Les conditions PCR ont été les suivantes :

13 cycles :

- Dénaturation : 30 secondes à 94°C
- Hybridation des amorces : 30 secondes à 65°C
- Diminution de 0,7°C/cycle soit une température de 56°C au 13<sup>ème</sup> cycle
- Elongation : 1 minute à 72°C

35 cycles

- Dénaturation : 30 secondes à 94°C
- Hybridation des amorces : 30 secondes à 56°C
- Elongation : 1 minute à 72°C
- Extension terminale : 5 minutes à 72°C

### 2. 3. Les séquences EST

Les séquences ESTs ont été obtenues à partir des ARNm extraits des explants foliaires de K59 et C15 (autofécondation de K59) après 4 jours de culture en condition d'embryogénèse somatique (Legrand *et al.*, 2007). Ces ARNm ont ensuite été utilisés afin de générer une banque embryogène et non embryogène par la technique SSH (Suppression Subtractive Hybridisation) (Diatchenko *et al.*, 1996), et au total 3,348 ESTs ont été générés pour les 2 banques.

Les EST ont été regroupés en OC (Original Cluster, regroupant les EST identiques) suite à l'application de 2 critères successifs : Les ESTs présentant entre eux une probabilité d'alignement (E-value) inférieure ou égale à  $1.10^{-30}$  ont été groupés dans un même OC primaire, puis les OC contenant plus de 1 séquence EST ont été alignés afin de vérifier qu'elles présentaient au moins 95% d'homologie entre

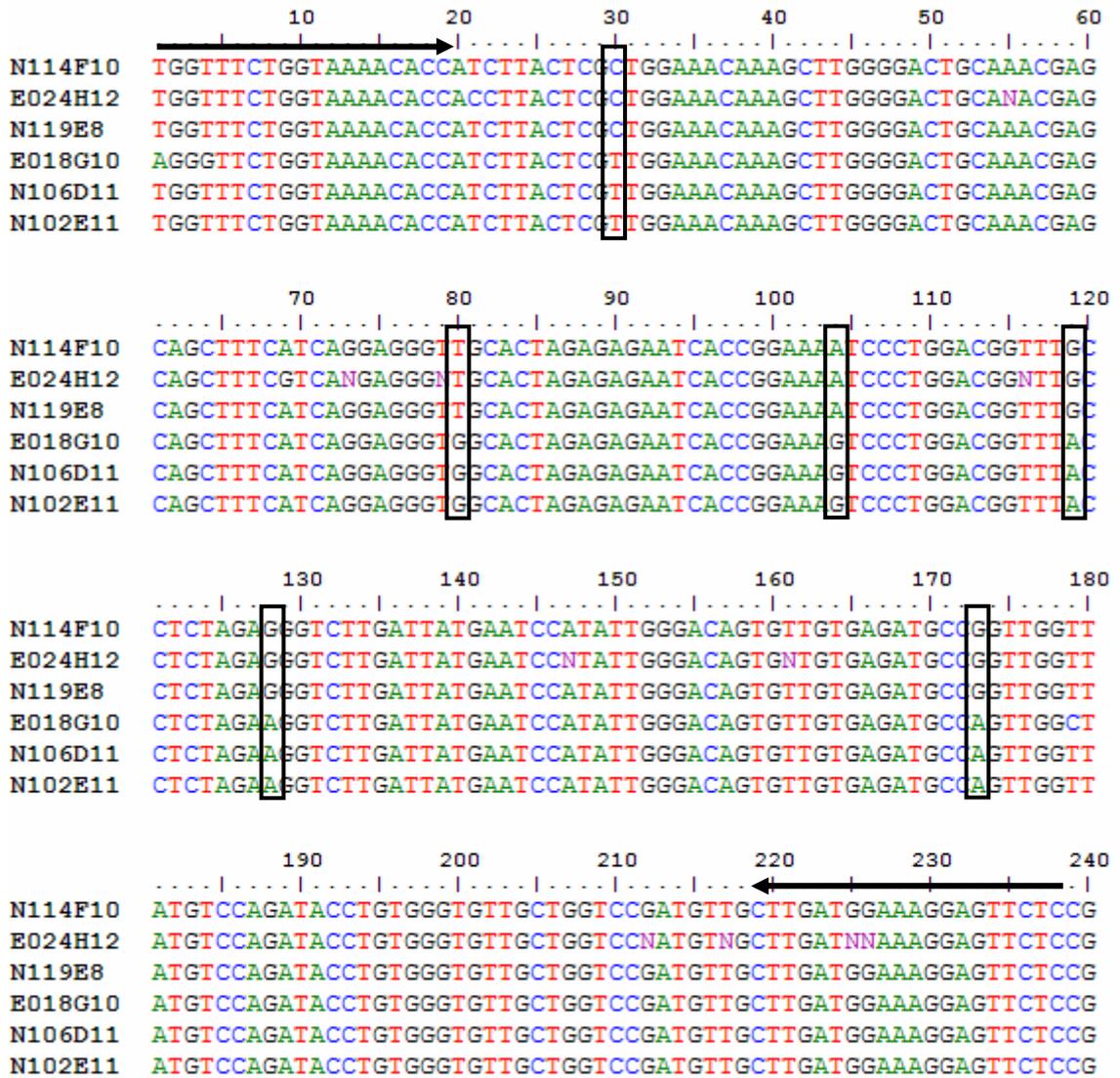
elles sur 150 pb. Les OC primaires contenant 2 ou plusieurs groupes d'EST, ont été divisés en sous-groupes.

Les 3,348 EST regroupés dans 2,302 OC ont été utilisés pour des travaux d'alignement avec les EST de laitue, de tournesol (CGPD, Compositae Genome Project Database) et de (PGDB : Plant Genome Database) afin d'identifier la fonction putative des gènes correspondants. Les alignements ont établi que ces EST représentent 2,077 gènes. Une analyse par microarray a mis en évidence 105 gènes différentiellement exprimés chez le génotype K59 (37 gènes) et non embryogène C15 (68 gènes) au 4<sup>ème</sup> jour de culture en condition d'induction de l'embryogenèse somatique (Legrand, 2006). Une soustraction *in silico* a été réalisée afin de vérifier l'efficacité de soustraction lors de la fabrication des banques soustractives et 33 gènes se sont révélés différentiellement exprimés ; 6 gènes sont préférentiellement exprimés chez le génotype embryogène K59 et 27 chez le génotype non embryogène C15. Le profil d'expression de ces gènes a été suivi par RT-PCR. Les résultats confirment l'expression différentielle de 14 gènes (2 chez K59 et 12 chez C15). Des analyses par RT-PCR ont également été réalisées pour 24 autres gènes, sélectionnés pour leur implication durant les phases précoces de l'embryogenèse somatique ou pour leur forte représentation dans les banques : Dix huit gènes différentiellement exprimés ont été identifiés (10 gènes chez K59 et 8 gènes chez C15).

- Recherche de SNP

Dans un premier temps, pour les OC contenant au moins 4 séquences ESTs des alignements de séquences ont été réalisés à l'aide du logiciel BioEdit v. 7.0.5.3. afin de distinguer les formes alléliques potentielles. Lorsque des changements de base (SNPs) ont été détectés *in silico*, alors des amorces ont été définies autour de SNP (Figure 1).

## Carte Génétique



**Figure 1 :** Alignement de séquences EST identiques regroupées en OC (*Original Cluster*) pour identifier la présence de SNP (*Single Nucleotide Polymorphism*). Le changement de base entre les séquences est indiqué par un cadre noir. Les amorces ont été définies (flèche noire) autour des SNP afin de différencier les formes alléliques. Les EST correspondants à un gène d'*AGP* (OC0687R2L2) ont été choisis pour illustrer cet exemple.

## Carte Génétique

- Alignement des séquences EST de chicorée avec les gènes d'*Arabidopsis thaliana* pour définir les régions introniques.

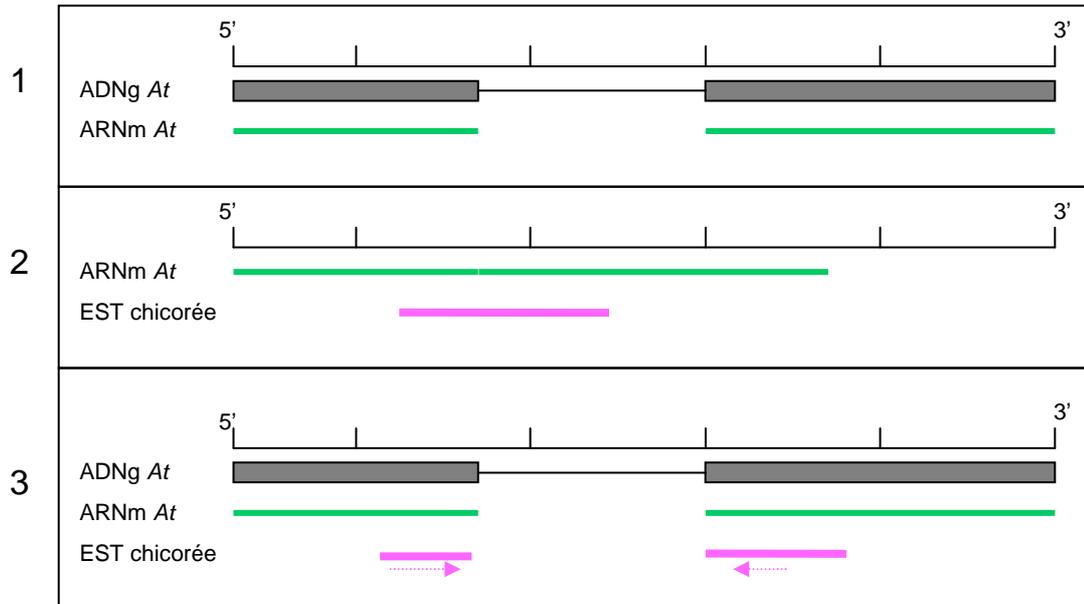
Nous nous sommes intéressés à rechercher du polymorphisme au sein des introns car le taux de mutations y est élevé.

L'expérience acquise au laboratoire a révélé que dans la grande majorité des cas les jonctions intron-exon sont bien conservées entre la chicorée et *Arabidopsis* autorisant l'alignement et l'identification des régions introniques dans les séquences de chicorée. Ainsi, pour les OC ne présentant pas de SNP *in silico* ou ne permettant pas la recherche de SNP (certains OC présentaient moins de trois séquences EST), les bases de données d'*Arabidopsis* ont été prospectées dans NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) afin d'identifier les régions homologues à partir d'un BLASTX.

Si les résultats du Blast ont permis d'associer un gène d'*Arabidopsis* à un EST de chicorée, alors la séquence EST de chicorée et la séquence génomique d'*Arabidopsis* ont été alignées par l'intermédiaire du logiciel BioEdit v. 7.0.5.3. L'alignement a permis de définir la position de la séquence EST de chicorée par rapport aux régions introniques et exoniques de la séquence d'*Arabidopsis*. L'alignement a été réalisé en plusieurs étapes (Figure 2) :

1. Alignement de la séquence génomique d'*Arabidopsis* et l'ARNm correspondant par le logiciel BioEdit v. 7.0.5.3 pour positionner les régions exoniques et introniques.
2. Alignement de la séquence ARNm d'*Arabidopsis* avec la séquence EST de chicorée par le logiciel BioEdit v. 7.0.5.3.
3. Positionnement manuel de la séquence EST de chicorée avec la séquence ARNm et génomique d'*Arabidopsis* préalablement alignées.

L'alignement a permis, dans certains cas, d'identifier la présence de régions introniques au sein des séquences chicorée (comme dans l'exemple ci-contre) et de définir des amorces autour du ou des introns.



**Figure 2 :** Alignement de séquences d'*Arabidopsis* (*At*) et de chicorée pour identifier les jonctions "intron-exon" à l'aide du logiciel BioEdit v. 7.0.5.3. L'alignement entre la séquence génomique (ADNg *At*) et l'ARNm (ARNm *At*) d'*Arabidopsis* correspondant (1) a été réalisé à l'aide des informations de séquences obtenues dans NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) pour positionner les régions exoniques et introniques. Un deuxième alignement entre la séquence ARNm d'*Arabidopsis* et la séquence chicorée correspondante a été réalisé (2). Ces alignements ont permis d'identifier au sein des séquences chicorée la présence d'introns (3). Lorsque des régions introniques ont été identifiées, des amorces (F, sens et R, antisens) ont été définies à partir de la séquence chicorée (3).

- Construction d'amorces à partir de séquences chimériques : *Lactuca sp.* /chicorée

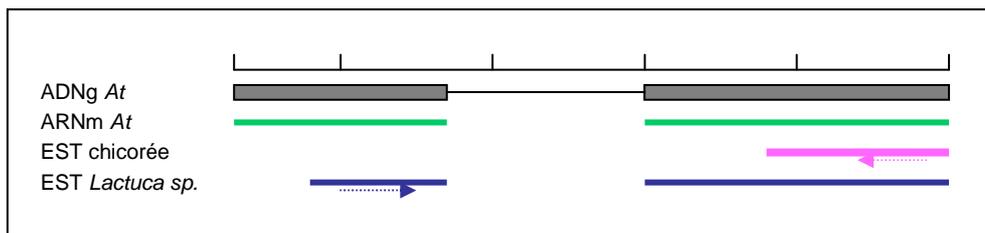
Si les séquences EST de chicorée était située au niveau d'un exon de la séquence génomique d'*Arabidopsis* alors celles-ci ont été confrontées aux séquences de *Lactuca sp.* de la base de données CPDG (Compositae Genome Project Database, <http://cgpdb.ucdavis.edu/>). L'intérêt de cette démarche a été de définir des amorces à partir des séquences *Lactuca sp.* et d'autres amorces à partir des séquences chicorée (Figure 3) encadrant des régions introniques pour construire des couples d'amorces : *Lactuca sp.* /chicorée

- Utilisation du logiciel de prédiction des jonctions "exon-intron"

Enfin, les séquences EST, ne présentant ni SNP ni homologie avec les séquences d'*Arabidopsis* ont été confrontées au logiciel de prédiction des jonctions "intron-exon" NetGene 2 v 2.4.

### 2. 3.1 Définition des amorces

Les amorces sens et anti-sens spécifiques de chaque séquence étudiée ont été définies à l'aide du logiciel Primer 3 v0.3.0. Pour chaque paire d'amorces définie, une amorce a été liée à un fluorochrome (6-Fam ou Hex). Les paramètres fixés pour la définition des amorces ont été : une température d'hybridation (55°C à 60°C), une longueur des amorces (de 19 à 23 pb) et une taille de l'amplicon de 150-1200 pb. Les amorces définies pour les séquences EST et utilisées au cours de cette étude sont présentées dans les annexes 3, 4, 5, 6 et 7.



**Figure 3 :** Construction d'un couple d'amorces *Lactuca sp./chicorée*. Lorsque la séquence de chicorée associée à une séquence d'*Arabidopsis* ne couvre pas la région intronique, alors un blastX entre une séquence de laitue *sp.* et chicorée a été réalisé. Si une association a pu être établie et la séquence de laitue *sp.* couvre au moins une région intronique, alors des amorces ont été définies autour de l'intron. Une amorce a été définie à partir de la séquence de chicorée et une autre amorce a été définie à partir de la séquence laitue *sp.* La position des amorces définies sens (en bleue) et antisens (en rose) est indiquée par une flèche.

### 2.3.2. Réactions PCR

Les réactions PCR ont été réalisées dans un volume de 15 µl contenant :

- Tampon PCR (1X)
- MgCl<sub>2</sub> (2 mM)
- dNTP (100 µM de chaque)
- BSA (0,2 mg/ml)
- Amorces liées à un fluorochrome 6-Fam ou Hex (400 nM)
- Amorces non marquées (400 nM)
- Taq polymérase (Applied Biosystems) (0,3 U)
- ADN (2 ng)

Les conditions PCR ont été les suivantes :

Cinq minutes de dénaturation à 94°C suivies de,

5 cycles :

- Dénaturation : 30 secondes à 94°C
- Hybridation des amorces : 30 secondes à T<sub>m</sub> (=température d'hybridation)
- Diminution de -1°C/cycle jusqu'au 5<sup>ième</sup> cycle
- Elongation : 30 secondes à 72°C

35 cycles :

- Dénaturation : 30 secondes à 94°C
- Hybridation des amorces : 30 secondes à T<sub>m</sub> – 5°C
- Elongation : 30 secondes à 72°C
- Extension terminale : 7 minutes à 72°C

### 3. Analyse des produits d'amplification

#### 3.1. Electrophorèse sur gel d'agarose

Dans un premier temps, l'amplification des produits PCR, à partir des amorces définies pour les séquences EST, a été observée sur gel d'agarose 1%. De manière

générale, 7 µl de produits PCR ont été mélangés avec 0,7 µl de tampon de charge et a été déposé sur gel contenant du bromure d'éthidium (10 mg/ml). La migration a été effectuée à 130 volts durant 40 minutes dans du tampon TBE 0,5 X. Enfin, le gel a été visualisé sous lampe UV. La taille de fragments a été évaluée par une échelle de poids moléculaire (Smart, Ladder-Eurogentech).

### 3.2. Séparation des fragments d'amplification par électrophorèse capillaire

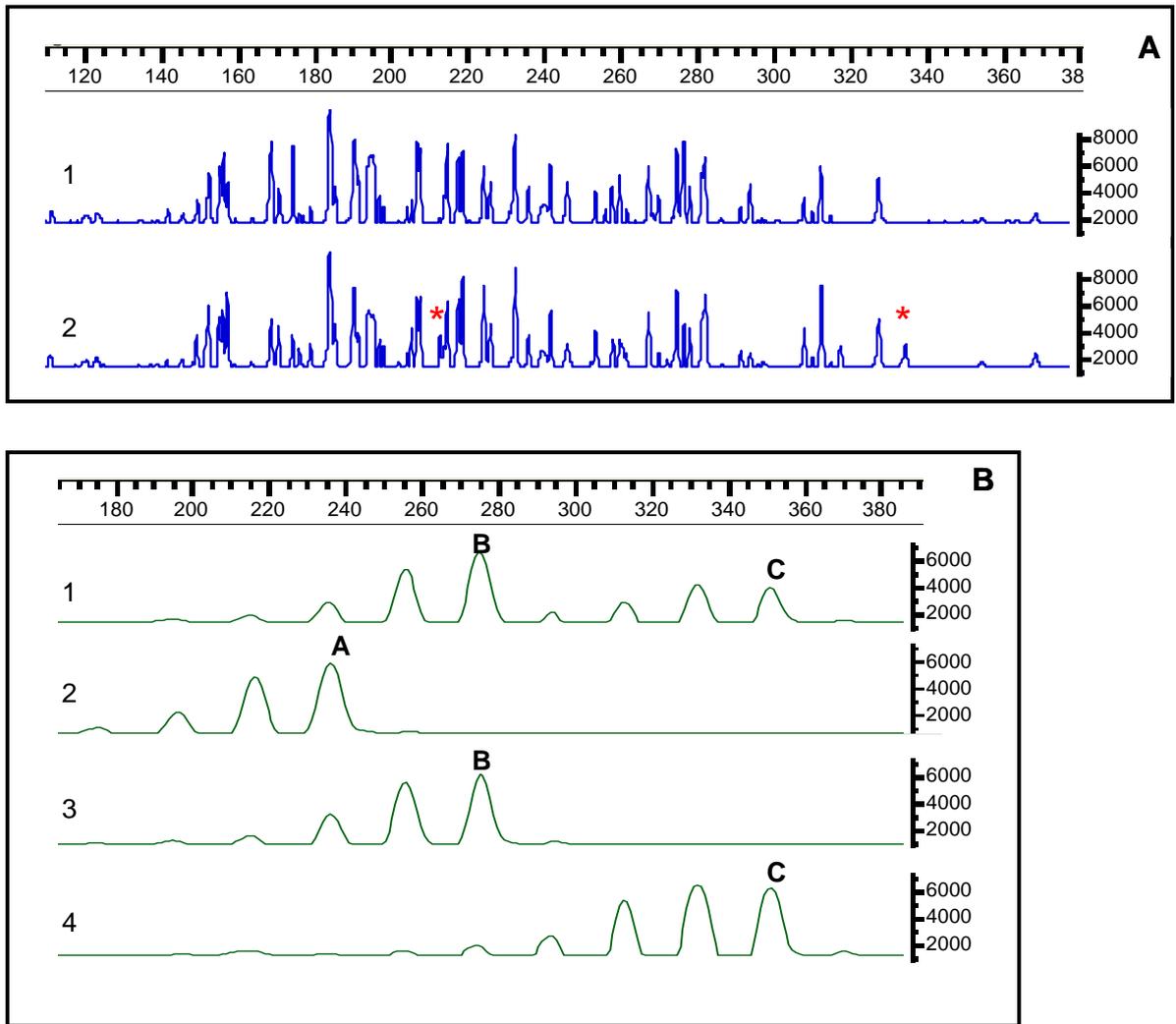
#### 3.2.1. Les marqueurs SSR et AFLP

Pour révéler les variations de type InDel (insertion-délétion) dans les séquences SSR et le polymorphisme de longueur pour les AFLP, les échantillons ont été soumis à migration en condition dénaturante pour détecter un polymorphisme de taille (Figure 4).

En pratique, 1 µl de produits PCR contenant des fragments de tailles différentes ont été mélangés, et combinés avec 9 µl de formamide et un marqueur de taille interne (Lys ou Rox). Après dénaturation (4 min à 95°C), les échantillons ont été fractionnés sur le séquenceur capillaire ABI PRISM® 3100-*Avant* (Applied Biosystems, Foster City, CA, USA). La migration a été réalisée en condition dénaturante (POP4).

Les échantillons fractionnés ont été analysés avec les logiciels GeneScan® analysis 3.7.1 et Genotyper 3.7.

## Carte Génétique



**Figure 4 :** Profils AFLP (A) et SSR (B) obtenus après migration des échantillons sur électrophorèse capillaire en condition dénaturante.

- (A) La présence des pics pour l'individu 2 mais absent chez l'individu 1 est indiqué par \*
- (B) L'individu 1 est hétérozygote (BC) et les individus 2, 3 et 4 sont homozygotes (A, B ou C).

### 3.2.2. Les marqueurs STS (comprenant les séquences EST)

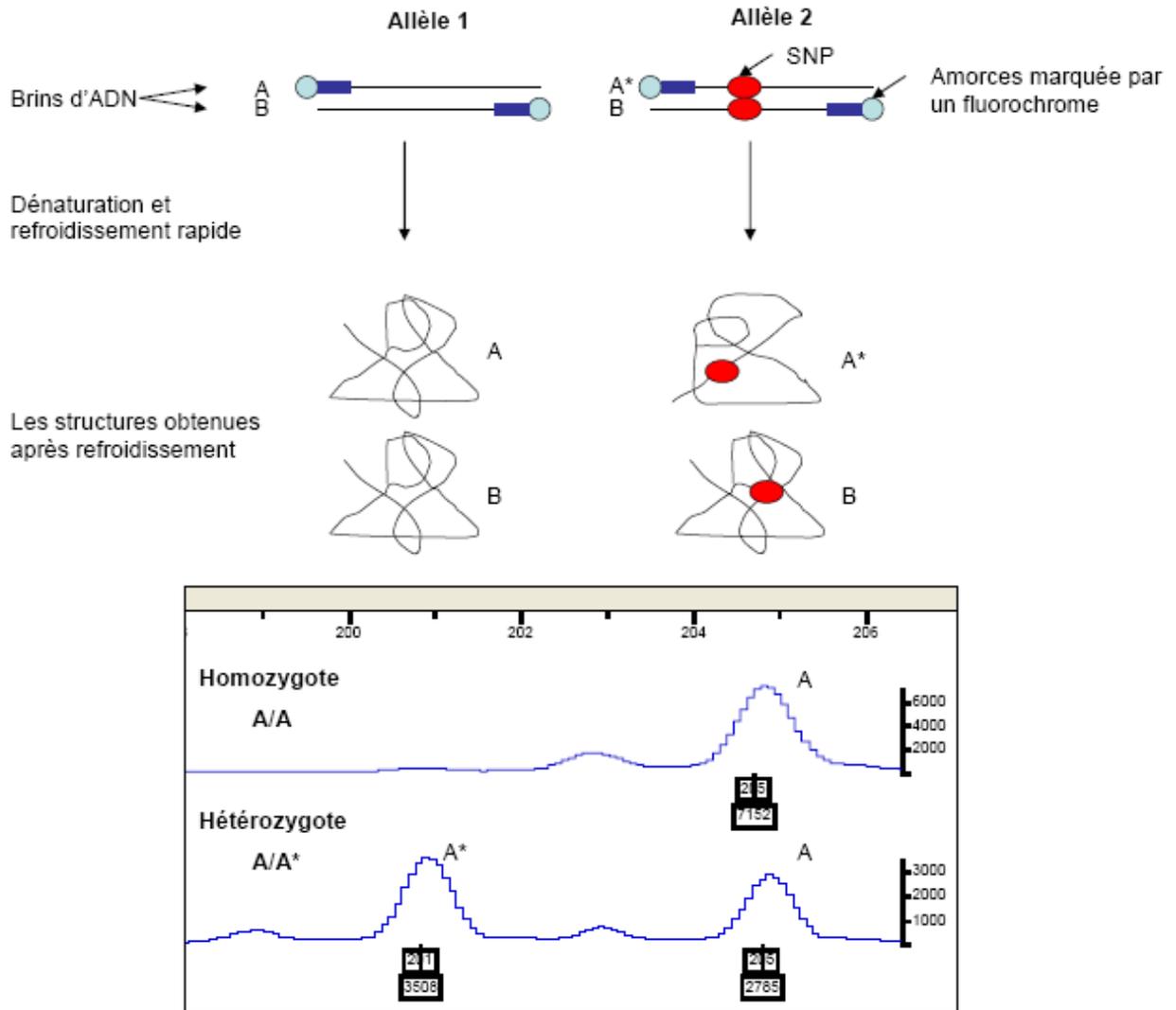
La technique SSCP (*Single Strand Conformation Polymorphism*) (Figure 5), nous a permis de révéler le polymorphisme des séquences amplifiées pour chacun des marqueurs STS analysés (Orita *et al.*, 1989 ; Andersen *et al.*, 2003).

En pratique, 1  $\mu$ l de produits PCR contenant des fragments de tailles différentes ont été mélangés, et combinés avec 9  $\mu$ l de formamide, 0,5  $\mu$ l de soude et un marqueur de taille interne (Lys ou Rox). Après dénaturation (4min à 95°C), les échantillons ont été fractionnés sur le séquenceur capillaire ABI PRISM® 3100-*Avant* (Applied biosystems, Foster City, CA, USA). La migration a été réalisée en condition non dénaturante (CAP).

Par défaut, la recherche de polymorphisme au sein des séquences a été réalisée à 25°C. La température ayant un rôle important dans la mobilité de l'ADN simple brin, elle a parfois été modifiée (Figure 6). Selon les produits amplifiés, la température optimale pour la migration varie. Ainsi, lorsque les profils étaient complexes à analyser la température a été augmentée à 30°C ou diminuée à 18°C en fonction de l'échantillon.

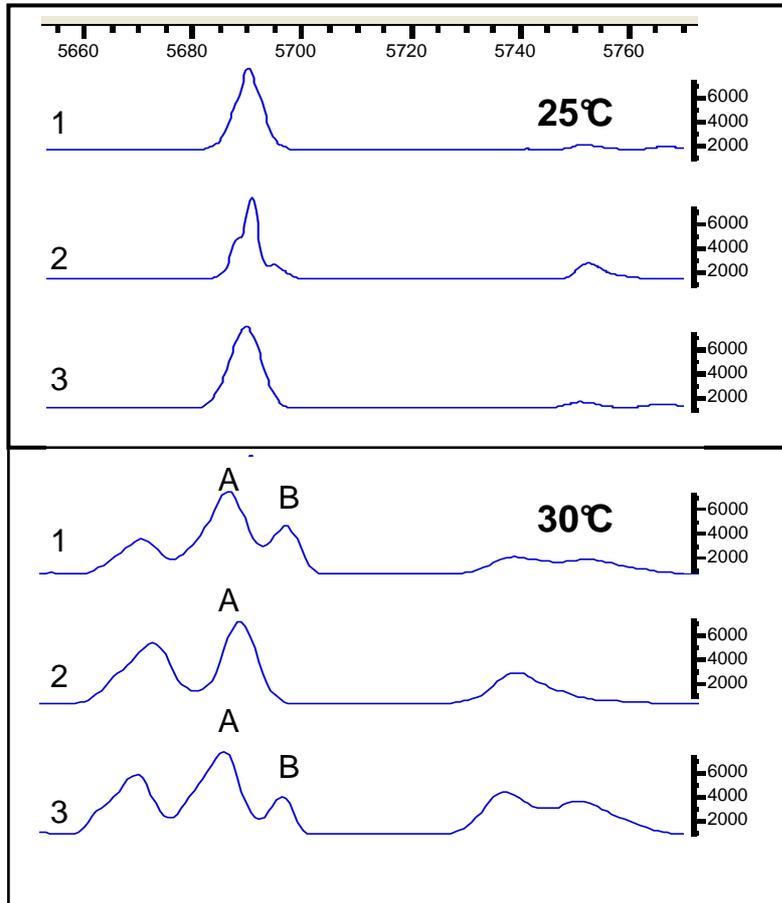
Une fois les échantillons fractionnés, ils ont été analysés avec les logiciels GeneScan® analysis 3.7.1 et Genotyper 3.7.

## Carte Génétique



**Figure 5 :** La technique SSCP (Single Strand Conformation Polymorphism) (Orita et al., 1989, Andersen et al., 2003) consiste à faire migrer des fragments d'ADN dénaturés, préalablement amplifiés avec des amorces spécifiques marquées par un fluorochrome en 5', dans un gel de polyacrylamide non dénaturant. L'ADN double brin est dénaturé par chauffage, puis rapidement refroidi dans la glace, ce qui empêche les molécules simple brin de se réassocier entre elles, et forment une structure secondaire stable du fait de la présence de SNP. Ces molécules peuvent prendre des conformations différentes décelables par une différence de migration lors de l'électrophorèse capillaire en fonction des conditions de migration (voltage, température, composition du gel).

## Carte Génétique



**Figure 6:** Variation de la température (25°C et 30°C) lors de la migration des échantillons par électrophorèse capillaire en condition non dénaturante. La migration des échantillons à 30°C, a permis de distinguer les allèles A et B, présents chez les individus hétérozygotes 1 et 3. Pour l'individu homozygote 2, seul l'allèle A est présent.

### 3.3. Electrophorèse sur gel d'acrylamide

La détection de polymorphisme de taille a aussi été réalisée par migration des produits PCR par électrophorèse sur gel d'acrylamide.

Pour mettre en évidence le polymorphisme, nous avons réalisé des migrations sur gel d'acrylamide à 8% (Long Ranger) en condition dénaturante en utilisant le séquenceur automatique d'ADN (modèle 4200<sup>TM</sup>, Li-Cor). Brièvement, le gel d'acrylamide a été coulé entre deux plaques de verre puis laissé polymériser pendant 2 heures. Quinze µl de produits PCR ont été mélangés avec 7,5 µl de tampon de charge, dénaturés 4 min à 95°C puis 1 µl de mélange a été déposé sur gel.

Contrairement aux données analysés automatiquement sur le séquenceur capillaire par les logiciels GeneScan® analysis 3.7.1 et Genotyper 3.7, les données visualisées sur le séquenceur Li-Cor ont été relevés manuellement.

### 4. Construction des cartes génétiques de liaison

La carte génétique a été construite à l'aide du logiciel JoinMap v 3.0 (van Ooijen & Voorrips 2001) puisqu'il traite les populations de type F1' issues de parents hétérozygotes (cross pollinated population).

Les paramètres d'analyses de JoinMap ont été fixés: fréquence de recombinaison = 0.45, LOD = 3 et les distances génétiques ont été calculées suivant la fonction Kosambi (Kosambi, 1944).

La construction de la carte génétique, à partir de la descendance F1', issue du croisement K28\*K59 a été réalisée. Les marqueurs analysés ont été codés selon les instructions du logiciel JoinMap (Tableau 1).

La stratégie de cartographie des marqueurs a consisté à établir :

- Une carte parentale de type E+H (K59) construite avec les marqueurs hétérozygotes chez le parent K59 et les marqueurs de type H (codominant H1, H3 et H5 ; dominant H4).
- Une carte parentale de type N+H (K28) construite avec les marqueurs hétérozygotes chez le parent K28 et les marqueurs de type H (codominant H1, H3 et H5 ; dominant H4).

## Carte Génétique

- La présence de marqueurs codominants H1, H3 et H5 dans les deux cartes parentales a permis ensuite de construire une carte parentale intégrée E+H+N (K28\*K59) avec les marqueurs de type E et N, les marqueurs codominants H1, H3 et H5 et les marqueurs dominants H4.

**Tableau 1 :** Code utilisé par le logiciel JoinMap pour chaque type de ségrégation pour une population CP (cross-pollinated). Les données manquantes pour un génotype sont indiquées par (--).

Code	Type de ségrégation	Description	Génotype possibles
<b>E</b>	<lmxll>	Locus hétérozygote chez le parent K59	ll, lm, --
<b>N</b>	<nnxnp>	Locus hétérozygote chez le parent K28	nn, np, --
<b>H1</b>	<hkxhk>	Locus hétérozygote chez les deux parents, 2 allèles, marqueur codominant	hh, kk, hk, --
<b>H4</b>	<hkxhk> (h- ; kk)	Locus hétérozygote chez les deux parents, 2 allèles, marqueur dominant	h-, kk, --
<b>H3</b>	<efxeg>	Locus hétérozygote chez les deux parents, 3 allèles, marqueur codominant	ef, eg, ee, fg, --
<b>H5</b>	<abxcd>	Locus hétérozygote chez les deux parents, 4 allèles, marqueur codominant	ac, ad, bc, bd, --

### III. Résultats

#### 1. Analyse des marqueurs cartographiés dans la descendance F1'a

Toutes les séquences révélées polymorphes ont été génotypées sur les 123 individus phénotypés de la sous-descendance F1'a

##### 1.1 Les marqueurs microsatellites ou SSR

Sur les 39 paires d'amorces testées, 21 ont généré au total 23 marqueurs microsatellites polymorphes. L'amplification du marqueur sw2f10 a généré plusieurs profils correspondant à 3 locus, nous avons ainsi sw2f10d.H4 sur le groupe de liaison LG7, et sw2f10.104E et sw2f10cH4 sur le groupe de liaison LG8. Sur les 23 marqueurs microsatellites polymorphes détectés, 5 sont hétérozygotes chez le parent femelle K28, 5 chez le parent mâle K59, et 11 chez les deux parents, présentant une ségrégation de type H3 (9), H5 (1) ou H1 (1), et 2 sont des marqueurs dominants (H4). Sur les 18 couples d'amorces qui n'ont pas été retenus, 12 séquences étaient monomorphes et 6 séquences présentaient un profil d'amplification très complexe. Ces dernières ont été abandonnées dans la suite de l'étude.

Sur l'ensemble des marqueurs microsatellites polymorphes, 6 présentaient une distorsion de ségrégation. L'intégration de ces marqueurs en distorsion au niveau de la carte a entraîné une perturbation dans l'ordre des marqueurs sur les groupes de liaison. Ces marqueurs ont alors été éliminés. Finalement, 17 marqueurs SSR ont été cartographiés à partir de la descendance F1'a. Ces mêmes marqueurs ont également été intégrés à la carte génétique K28\*K59 (Cadalen *et al.*, 2009).

### 1.2. Les marqueurs AFLP

Au total, 269 marqueurs AFLP générés par 16 couples d'amorces ont été intégrés à la carte génétique F1'a (Figure 7). Parmi ces marqueurs, 75,8% (204 marqueurs) et 22,8% (61 marqueurs) présentaient un ratio de ségrégation de type 1 :1 (type E et N) et de type 3 :1 (H4), respectivement. Le nombre de marqueurs d'origine paternel (K59 : 124 marqueurs) est plus élevé que le nombre de marqueurs d'origine maternel (K28 : 80 marqueurs). Dix huit marqueurs présentaient une distorsion de ségrégation, soit 5,8% des marqueurs AFLP [ $p < 0.05$  (9 marqueurs),  $p < 0.01$  (8 marqueurs),  $p < 0.005$  (1 marqueurs)].

Les marqueurs AFLP ont permis de couvrir rapidement les 9 groupes de liaison. Ces marqueurs ont augmenté de manière significative la densité des marqueurs (densité moyenne de 6,9 à 2,6) pour le groupe de liaison 9, groupe pour lequel peu de marqueurs SSR et STS ont été cartographiés. De plus, sur les groupes de liaison 2 (de 34 à 50 cM), 3 (de 21 à 39 cM) et 4 (de 6 à 19 cM et de 23 à 30 cM), les marqueurs AFLP ont couvert des régions pour lesquelles aucun marqueur SSR et ou STS n'étaient assignés.

### 1.3. Les séquences EST

La première démarche a consisté à identifier la présence de SNP *in silico* (Annexe 3) afin de distinguer les formes alléliques. Lorsqu'un ou plusieurs SNP ont été détectés, les séquences EST ont été confrontées aux séquences contenues dans les bases de données d'*Arabidopsis* afin d'identifier la présence ou non de régions introniques. Ceci avait pour but d'éviter de se situer dans une zone de jonction intron-exon lors de la définition des amorces, autour du ou des SNP.

Quatorze OC ont révélé la présence de SNP et 20 couples d'amorces ont été définis autour de ceux-ci. Lorsqu'un intron était présent à proximité d'un SNP, par exemple pour OC0378, OC0466, OC0748, les amorces ont été définies afin d'encadrer l'intron et le SNP, ceci pour augmenter les chances de détecter du polymorphisme. Les amorces ont permis d'amplifier l'ensemble des séquences et 8 couples d'amorces ont révélé la présence de polymorphisme au sein des OC : OC0082\_a, OC0206, OC0378\_a, OC0603\_a, OC0687, OC0748, OC1635\_a, OC1976. De nouvelles amorces ont été définies autour des introns, lorsqu'ils étaient présents, pour les OC pour lesquels l'amplification de la séquence présentant un SNP *in silico* n'a pas révélé de polymorphisme (OC0170, OC0316, OC0183, OC0564). Seul l'OC0183 a révélé un polymorphisme.

Les séquences EST, ne présentant pas de SNP *in silico* ou pour lesquelles ceux-ci n'étaient pas détectés (certains OC présentaient moins de 3 séquences EST), ont été directement confrontées aux séquences contenues dans les bases de données d'*Arabidopsis* afin d'identifier la présence ou non de région intronique. Quarante-cinq OC (Annexe 4) présentaient au moins une région intronique autour desquelles nous avons défini 111 couples d'amorces. À partir des amorces définies, 13 couples d'amorces n'ont pas permis d'amplification et 31 ont révélé du polymorphisme. Dans la majorité des cas, les régions introniques définies à partir des séquences génomiques d'*Arabidopsis* sont plus grandes chez la chicorée. L'amplification de la séquence correspondante à l'OC0347 a révélé plusieurs profils. Pour cette séquence, nous avons pu lire deux locus : OC0347aE et OC0347bN, localisés sur les groupes de liaisons 3 et 8, respectivement.

Les séquences EST associées à un gène d'*Arabidopsis* seulement au niveau des régions exoniques ont été confrontées aux séquences de *Lactuca sp.* afin de déterminer des homologies de séquences (Annexe 5). Le blast des séquences EST de chicorée avec les séquences *Lactuca sp.* a révélé des homologies de séquences de 85 à 95%. Les alignements de séquences entre la laitue, le gène d'*Arabidopsis* associé et la séquence chicorée ont été réalisés pour les OC : OC0168, OC0745, OC1359\_a, OC1568 et OC1867\_b. Ces alignements ont permis de construire 5 couples d'amorces (*Lactuca sp./chicorée*), d'amplifier les séquences et de révéler une seule séquence polymorphe (OC1359\_a).

Si aucune démarche ne permettait d'identifier des régions introniques, les séquences EST étaient confrontées au logiciel de prédiction des jonctions « exons/introns ». Pour 4 EST (OC0067, OC1056, OC1824 et OC1932), la prédiction exons/introns a été positive (Annexe 6). Cependant, dans notre cas, cette approche ne s'est pas révélée très efficace. Sur les 4 couples d'amorces définies autour des introns, 3 couples d'amorces ont permis une amplification PCR. Seules les amorces définies pour l'OC1056 ont permis l'amplification d'une région intronique. Cependant, la migration du produit PCR par électrophorèse capillaire n'a pas révélé de polymorphisme. Les amorces que nous avons définies pour l'OC1932 n'ont pas permis d'amplifier la séquence intronique. Par contre, le produit PCR a révélé la présence d'un polymorphisme après migration par électrophorèse capillaire. Ces amorces ont certainement été définies autour d'une variation de type InDel que nous avons détecté par électrophorèse capillaire.

Nous avons également défini des amorces à chaque extrémité des séquences EST (Annexe 7) pour lesquelles aucune des démarches préalablement décrites n'avait permis de détecter de polymorphisme. Ainsi, 22 couples d'amorces ont été définis et 20 amplifications PCR réalisées ; 5 couples ont révélé du polymorphisme (OC0081, OC1145, OC2147, OC0344, OC1796).

Au total, sur les 162 paires d'amorces testées, 146 ont généré un produit d'amplification et 47 ont révélé la présence de polymorphisme (Annexe 8).

Les 47 marqueurs révélant du polymorphisme (Annexe 8) ont été cartographiés sur la carte génétique F1'a. Dix huit marqueurs sont de type E (K59) et

13 de type N (K28). Treize marqueurs codominants présentent une ségrégation de type H3 (4 marqueurs), H5 (1 marqueur) ou H1 (8 marqueurs) et seuls 3 marqueurs sont dominants (H4). Le marqueur OC0185R3L3 (LG 1) présente une distorsion de ségrégation ( $p < 0.005$ ).

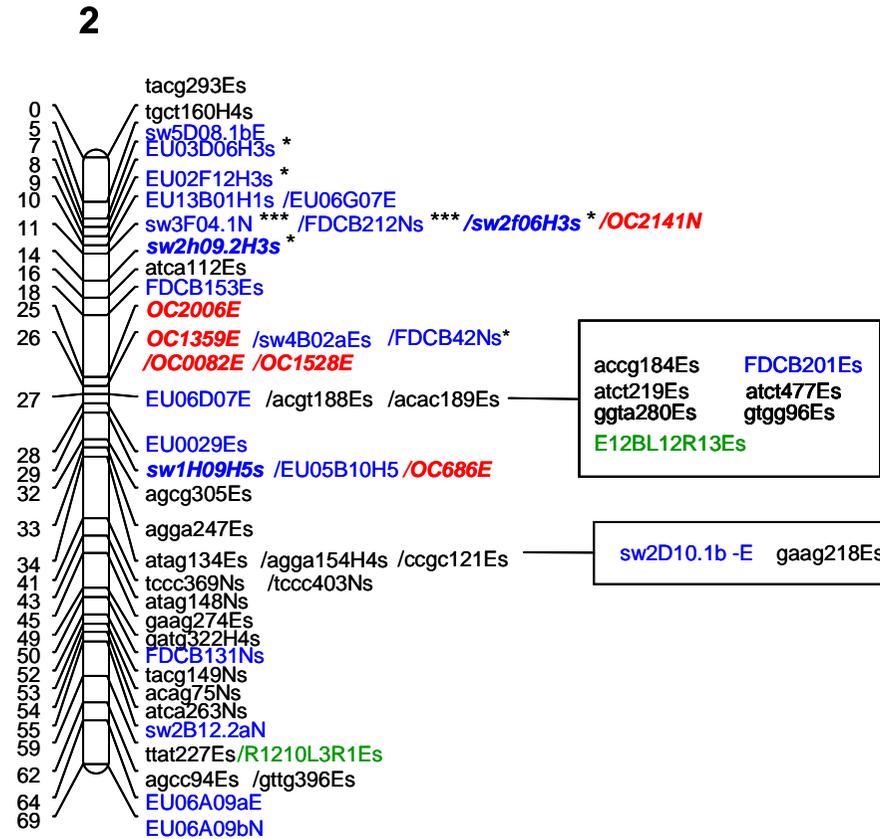
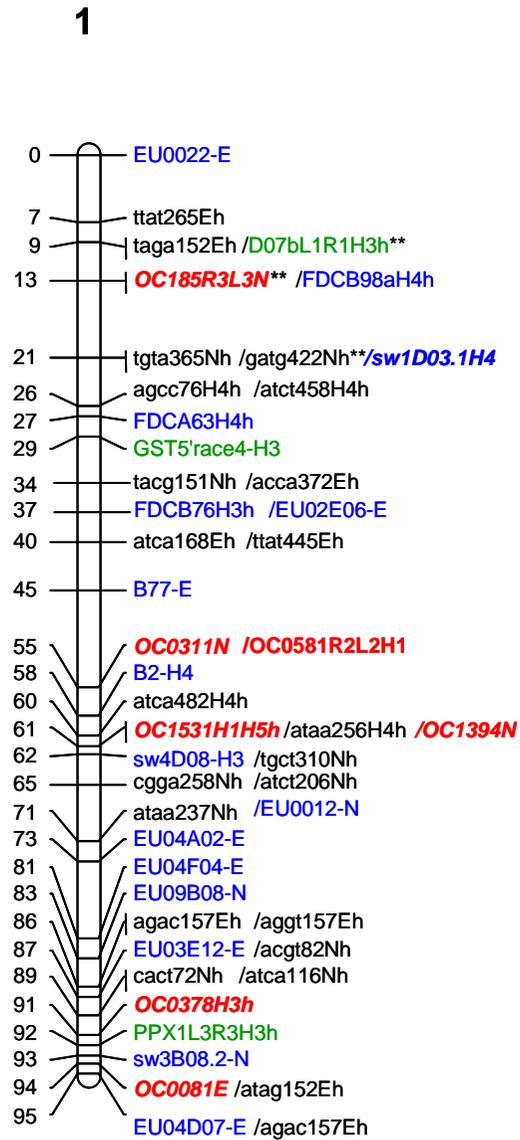
Nous avons également recherché du polymorphisme dans les séquences correspondantes au gène *STM* (*Shoot Meristemless*) afin de le cartographier. *STM* est un gène lié au processus de l'embryogenèse somatique et son rôle dans la formation des embryons somatiques a été étudié au cours d'une thèse (S. Da Silva, 2004). Plusieurs clones du gène *STM* ont été obtenus chez K59 et K28 (B. Huss, communication personnelle) et l'alignement des séquences nous a permis d'identifier des variations de type InDel.

Mon travail a consisté à définir des amorces autour des InDel afin de détecter du polymorphisme après migration des produits PCR par électrophorèse capillaire. Trois couples d'amorces (Annexe 9) ont été définis et les produits PCR obtenus pour 2 couples d'amorces ont révélé du polymorphisme. Après analyse du polymorphisme dans la descendance F1'a, *STM* a été cartographié sur le groupe de liaison 6 (Figure 7).

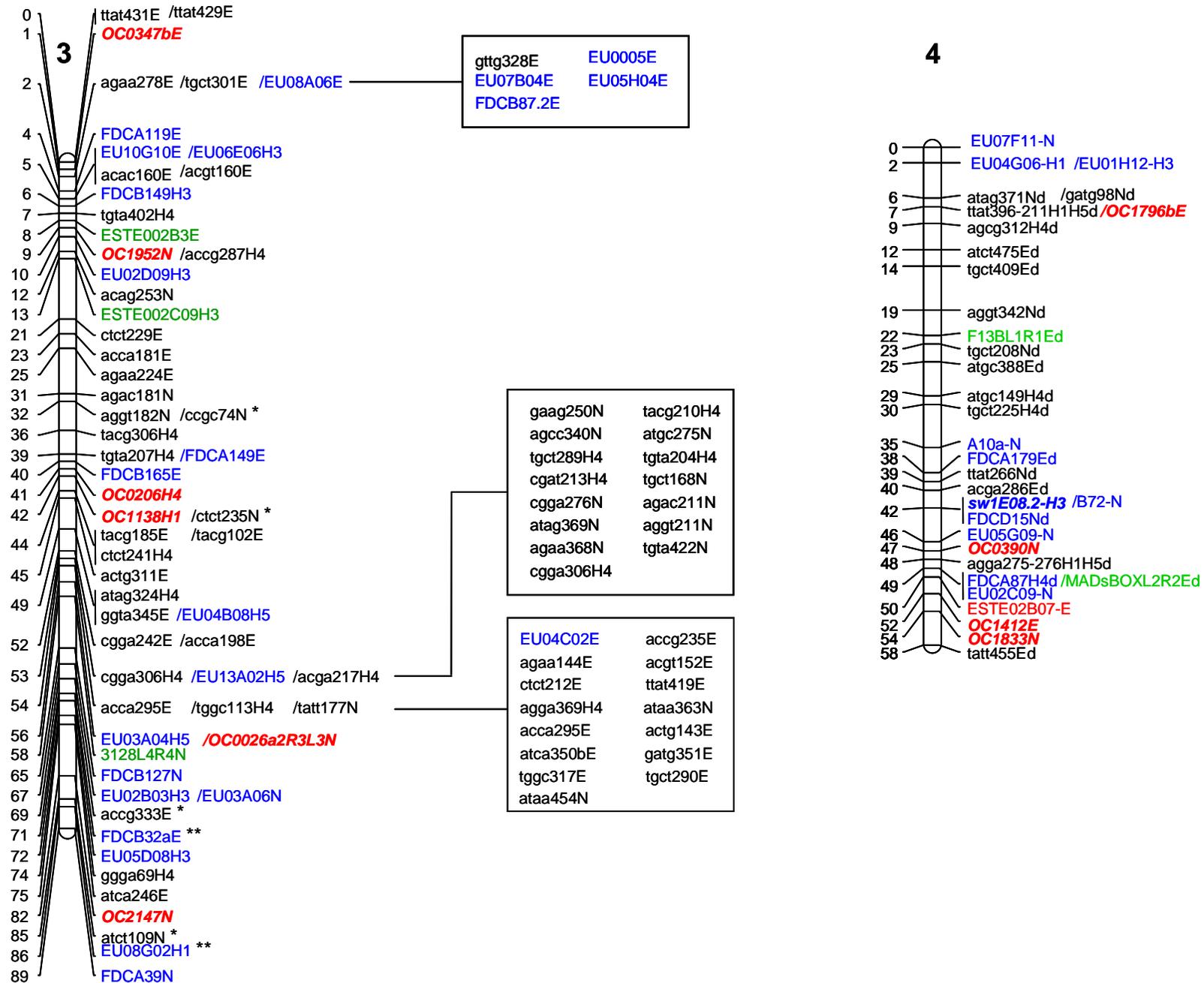
**Figure 7** : Carte génétique F1'a de la population de chicorée (*Cichorium intybus* L.) N=123.

Les marqueurs SSR, STS et AFLP ont été indiqués en bleu, vert et noir, respectivement. Les marqueurs EST ont été distingués des marqueurs STS par la couleur rouge. Les marqueurs distordus ont été suivis par un symbole (\*) représentant la probabilité associée au test  $X^2$  (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.005$ ).

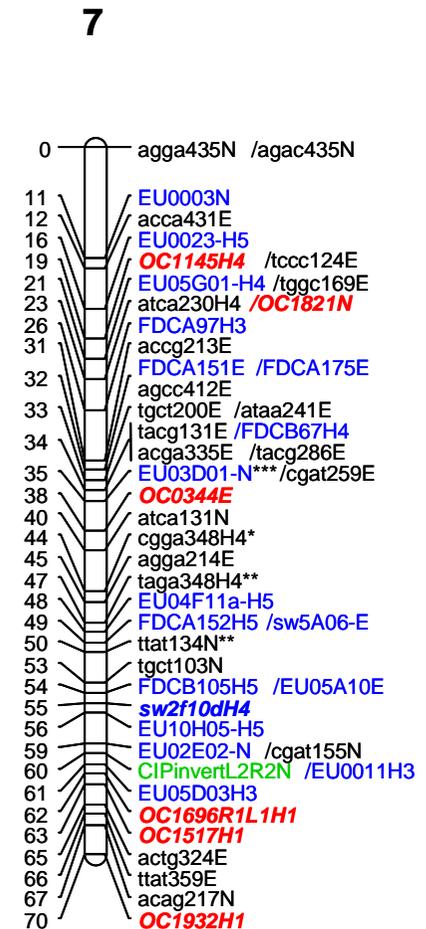
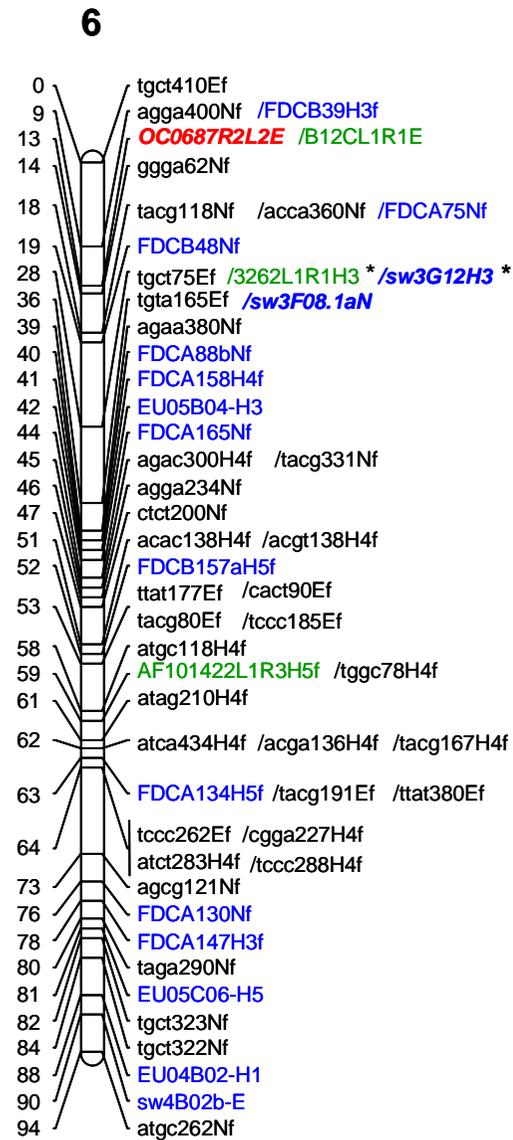
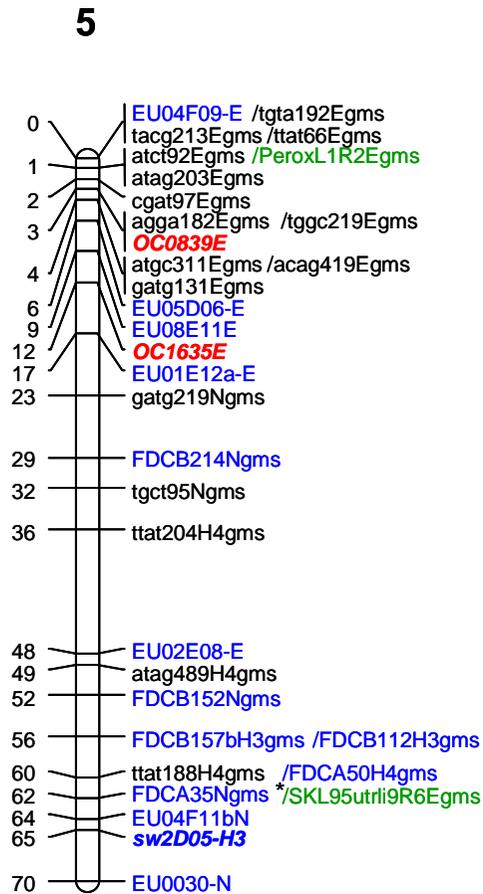
# Carte Génétique



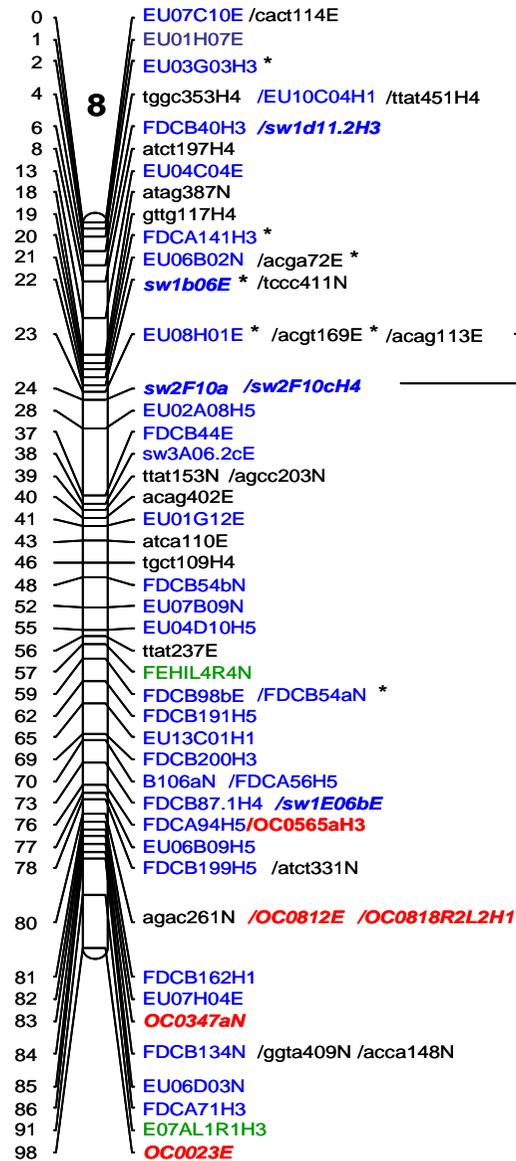
# Carte Génétique



## Carte Génétique



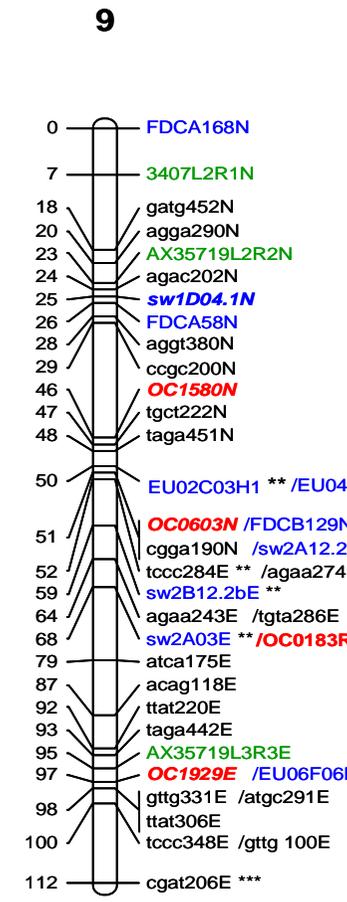
# Carte Génétique



atgc336E *	acac169E *
ttat336H4	acgt303H4

acag403H4	acgt343H4
	acac343H4

OC0748E	actg237E
EU07E01H5	cAMP2bE
actg237E	EU07A01H1
OC2060H3	cgat193H4
actg160E	acgt63-121H3
FDCB162H1	EU09B03H4
FDCB168H1	agaa213N
gatg265N	OC1976H1
tgta97H4	gaag302H4
gttg141H4	3704L2R1H1
OC0031H3	OC1822H1
ttat311E	atct229H4
3361L2R2H3	TUBL2R2H3
gatg232H4	EU04F07N
atca198-199-219H5	



3474L1R1H1 **
EU08F09H3 **
sw1E06aH1 **

2. Construction d'une carte génétique F1'a+b pour l'analyse QTL liée à l'embryogenèse somatique et à l'organogenèse à partir des génotypes des sous-descendances F1'a et F1'b.

Sachant qu'il n'était pas envisageable sur la durée de la thèse d'augmenter le nombre d'individus génotypés pour l'ensemble des marqueurs déjà présents sur la carte F1'a (Figure 7), nous avons entrepris la démarche suivante :

- une sélection des différents marqueurs (SSR, STS et AFLP) de la carte F1'a (Figure 7) selon différents critères : le type de ségrégation du marqueur, la localisation du marqueur sur le groupe de liaison et la présence ou non de distorsion de ségrégation au sein du marqueur.

- un génotypage de la descendance F1'b (n=104) pour les marqueurs sélectionnés.

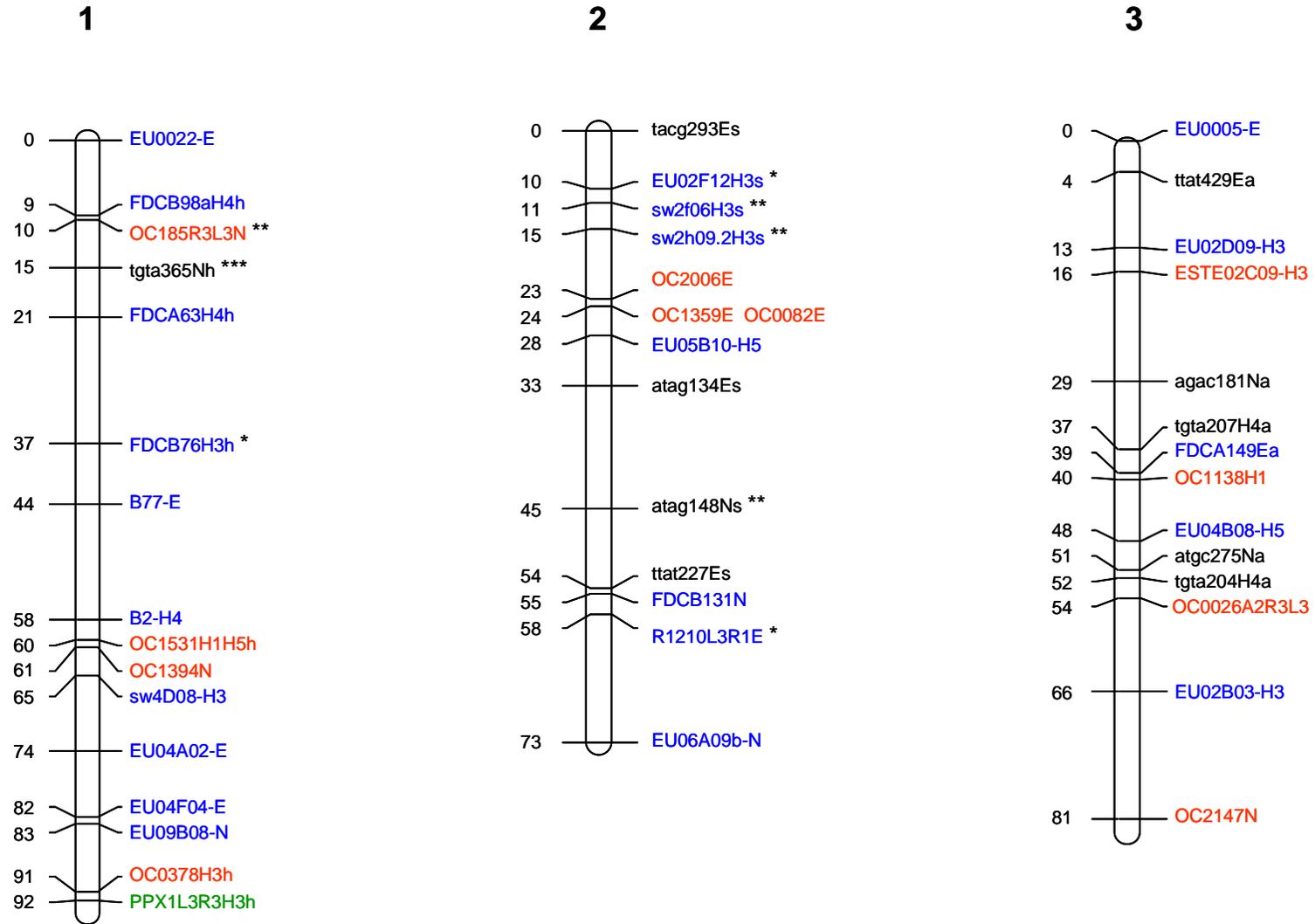
Ainsi pour l'analyse QTL, 124 marqueurs (52 SSR, 33 STS et 39 AFLP) ont été sélectionnés et génotypés à partir des ADN de la descendance F1'b (Figure 9). Ces marqueurs couvrent l'ensemble des 9 groupes de liaison de la carte génétique F1'a (Figure 8). Parmi ces marqueurs, 43 sont de type E, 33 de type N et 37 sont codominants : 21 marqueurs de type H3, 11 marqueurs de type H5 et 5 marqueurs de type H1. Seuls 11 marqueurs sont dominants (type H4). Plusieurs marqueurs, notamment sur le groupe de liaison 7, présentent une faible distorsion de ségrégation avec une probabilité P inférieure à 0.05. Le nombre de marqueurs par groupe de liaison s'élève de 13 à 18.

Une carte génétique F1'a+b (Figure 7) a été construite à partir des marqueurs sélectionnés et à partir des génotypes des sous-descendances F1'a (N=123) et F1'b (N=104) pour lesquels une évaluation phénotypique a été réalisée.

**Figure 8** : Carte génétique F1'a+b de la population (N= 227) de chicorée (*Cichorium intybus* L.) élaborée pour la détection de QTL.

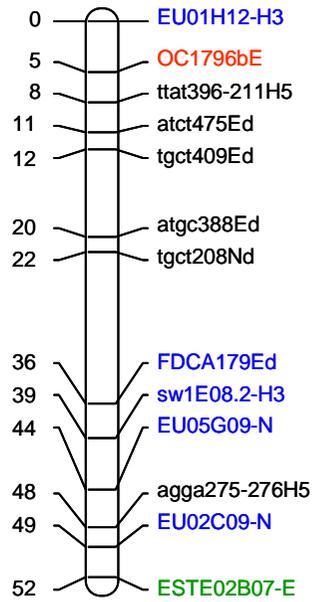
Les marqueurs SSR, STS et AFLP ont été indiqués en bleu, vert et noir, respectivement. Les marqueurs ESTs ont été distingués des marqueurs STS par la couleur rouge. Les marqueurs distordus ont été suivis par un symbole (\*) représentant la probabilité associée au test  $X^2$  (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.005$ ).

## Carte Génétique

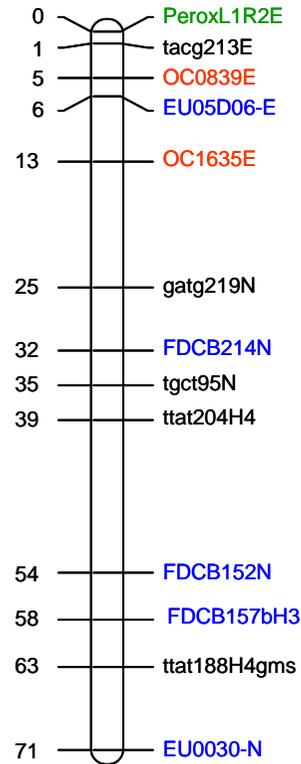


## Carte Génétique

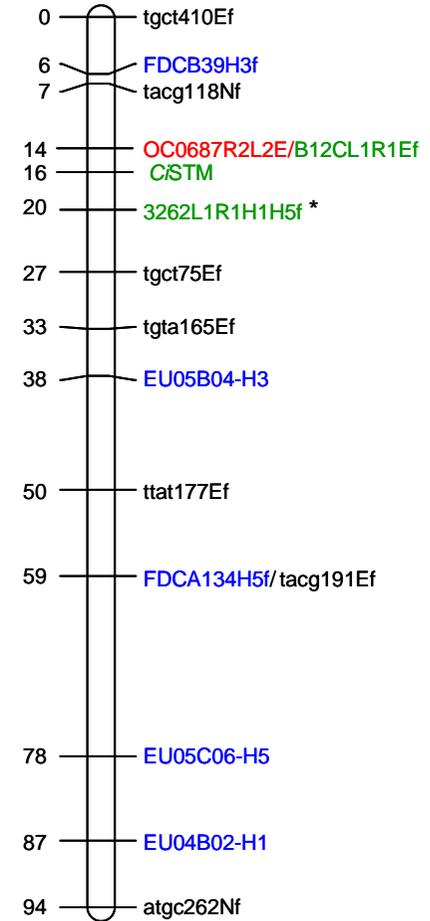
**4**



**5**

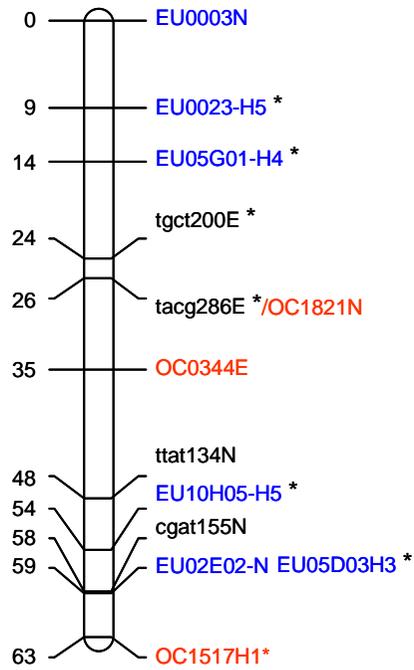


**6**

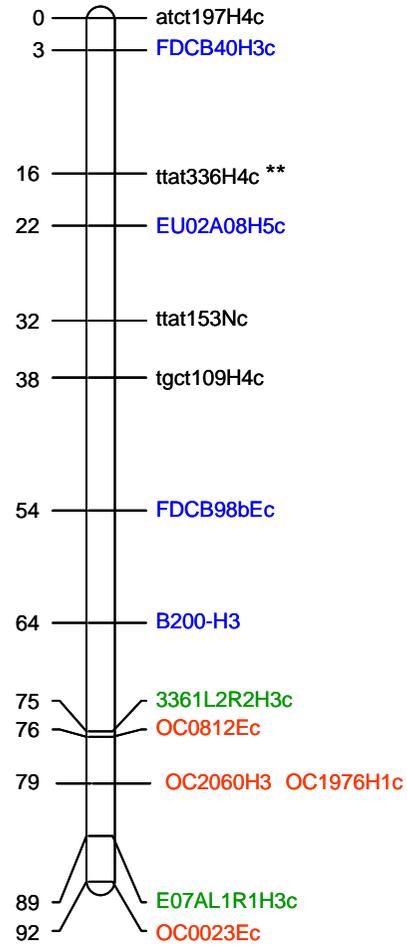


# Carte Génétique

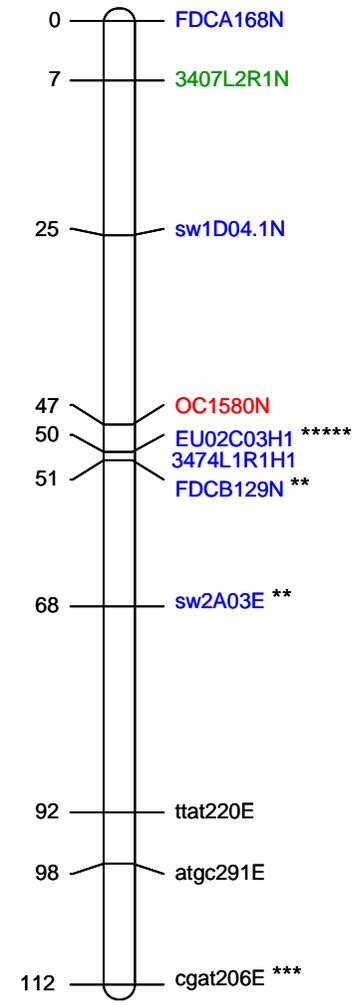
**7**



**8**



**9**



### 3. Distorsion de ségrégation des marqueurs de la sous-descendance F1'b

Lors du génotypage des marqueurs sélectionnés à partir des ADN de la sous-descendance F1'b, nous avons observé une forte distorsion de certains marqueurs qui n'avait pas été observée lors de la construction de la carte F1'a. Les résultats de l'analyse des données phénotypiques (Chapitre I) ayant également révélés une différence entre les deux sous-descendances (F1'a et F1'b), deux cartes génétiques ont été construites, une à partir des données génotypiques de la population F1'a (n=123) et l'autre à partir des données génotypiques de F1'b (n=104). Les deux cartes génétiques ne présentent pas de différences au niveau de l'ordre des marqueurs sur les groupes de liaison. La différence marquante entre les deux cartes génétiques se situe sur le groupe de liaison 7. De manière significative, les marqueurs du groupe 7 présentent une forte distorsion de ségrégation dans la descendance F1'b comparée à la sous-descendance F1'a (Tableau 2).

**Tableau 2 :** Distorsion de ségrégation des marqueurs du groupe de liaison 7 présents pour les sous-descendances F1'a et F1'b. Les marqueurs en distorsion sont suivis par un symbole (\*) représentant la probabilité associée au test  $\chi^2$  (\* p<0.05, \*\* p<0.01, \*\*\* p<0.005, \*\*\*\* p<0.001, \*\*\*\*\* p<0.0005, \*\*\*\*\*, p<0.0001, \*\*\*\*\*, p<0.00005)

Marqueurs	Position (cM)	Type de ségrégation <K59xK28>	Probabilité associée au test $\chi^2$	
			descendance F1'a	descendance F1'b
EU0003N	0	<nnxnp>		*
EU0023H5	9	<abxcd>		**
EU05G01H4	14	<hkxhk>(h- ; kk)		****
tgct200E	24	<llxlm>		*****
OC1821N	26	<nnxnp>		*
tacg286E	26	<llxlm>		*****
OC0344E	35	<llxlm>		*****
ttat134N	48	<nnxnp>	**	
EU10H05H5	54	<abxcd>		*****
cgat155N	58	<nnxnp>		
EU02E02N	59	<nnxnp>		
EU05D03H3	59	<efxeg>		*****
OC1517H1	63	<hkxhk>		*****

## Carte Génétique

Les marqueurs hétérozygotes chez les deux parents (type H1, H3, H4 et H5) et les marqueurs de type E (K59) présentent une très forte distorsion avec une probabilité associée au test de  $X^2$  de 0,00005, contrairement aux marqueurs hétérozygotes de type N (K28) (les marqueurs OC1821N et EU0003N présentent seulement une très faible distorsion avec  $p < 0.05$ ). Pour chaque marqueur en distorsion, l'observation des effectifs obtenus dans les différentes classes génotypiques (Tableau 3) montre qu'une ou plusieurs classes présentent un effectif faible par rapport aux effectifs attendus. Si l'on s'intéresse aux différentes classes pour les marqueurs fortement en distorsion ( $p < 0.00005$ ) ; le marqueur EU10H05 (ségrégation de type  $\langle ab \times cd \rangle$ ) présente un effectif faible pour les classes  $ac$  et  $ad$  comparées aux classes  $bc$  et  $bd$ . L'allèle  $a$ , apporté par K59 semble impliquer dans cette diminution d'effectif puisque les allèles  $c$  et  $d$ , en association avec l'allèle  $b$  ( $bc$  et  $bd$ ) présente un effectif relativement élevé. De même, pour le marqueur EU5D03H3 (ségrégation de type  $\langle ef \times eg \rangle$ ) où l'allèle apporté par K59, (l'allèle  $f$ ) semble également impliqué dans la diminution des effectifs pour les classes  $ef$  et  $eg$ . Les allèles  $e$  et  $g$  (apportés par K28) associés à l'allèle  $e$  ( $ee$  et  $eg$ ) (apporté par K59), présente un effectif relativement élevé.

L'effet allélique apporté par K59 et ou K28 ne peut être directement déterminé pour les marqueurs de type E et H1. Pour ce type de marqueurs, ce sont les classes  $ll$  (marqueur de type E,  $tacg200E$ ,  $tacg286E$ ) et  $hh$  (marqueur de type H1, OC1517H1) qui présentent un effectif observé faible par rapport aux autres classes du marqueur. Cependant, en se basant sur les marqueurs de type H3 et H5 (analysés ci-dessus) présents sur le même groupe de liaison, nous avons déduits que l'allèle apporté par K59 est responsable de la diminution d'effectifs.

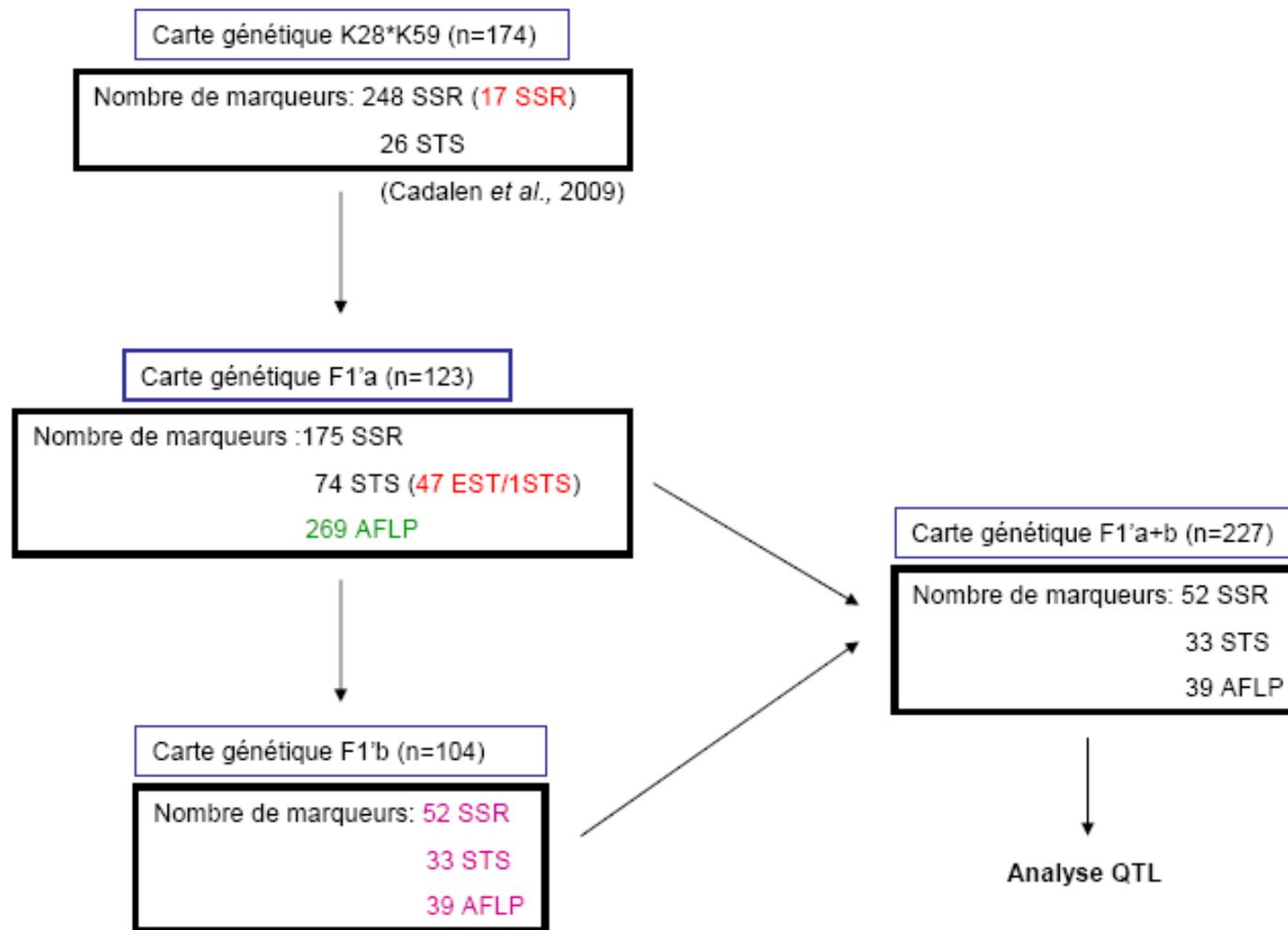
Carte Génétique

**Tableau 3** : Fréquence de distribution des génotypes par classe pour les marqueurs du groupe de liaison 7 de la sous-descendance F1'b. Les marqueurs en distortion sont suivis par un symbole (\*) représentant la probabilité associée au test X<sup>2</sup> (\* p<0.05, \*\* p<0.01, \*\*\* p<0.005, \*\*\*\* p<0.001, \*\*\*\*\* p<0.0005, \*\*\*\*\* p<0.0001, \*\*\*\*\* p<0.00005). L'effectif attendu pour chaque combinaison allélique est indiqué en bleu et en italique. Les classes présentant un effectif faible par rapport à l'effectif attendu sont indiquées en rouge. Le nombre de données manquantes pour chaque marqueur est indiqué par un symbole (- -).

Marqueurs	Type de ségrégation <K59xK28>	Génotypes possibles																	Probabilité associée au test X <sup>2</sup>		
		ac	ad	bc	bd	ee	ef	eg	fg	hh	hk	kk	h-	k-	ll	lm	nn	np	--	descendance F1'b (n=104)	
EU0003N	<nnxnp>															<i>51</i>	<i>51</i>	1	*		
EU0023H5	<abxcd>	<i>26</i> 33	<i>26</i> 33	<i>26</i> <b>12</b>	<i>26</i> 26														0	**	
EU05G01H4	<hkxhk>(h- ; kk)										<i>23</i> 13	<i>80</i> 90							1	****	
tgct200E	<lmxll>													<i>50</i> <b>31</b>	<i>50</i> 69				4	*****	
OC1821N	<nnxnp>															<i>51</i> 60	<i>51</i> 43	1	*		
tacg286E	<lmxll>													<i>50</i> <b>22</b>	<i>50</i> 78				4	*****	
OC0344E	<lmxll>													<i>52</i> <b>21</b>	<i>52</i> 83				0	*****	
ttat134N	<nnxnp>															<i>50</i> 52	<i>50</i> 48	4			
EU10H05H5	<abxcd>	<i>26</i> <b>10</b>	<i>26</i> <b>4</b>	<i>26</i> 40	<i>26</i> 49														1	*****	
cgat155N	<nnxnp>															<i>51</i> 49	<i>51</i> 54	1			
EU02E02N	<nnxnp>															<i>52</i> 50	<i>52</i> 54	0			
EU05D03H3	<efxeg>					<i>26</i> 39	<i>26</i> <b>10</b>	<i>26</i> 47	<i>26</i> <b>7</b>											1	*****
OC1517H1	<hkxhk>									<i>26</i> <b>7</b>	<i>52</i> 55	<i>26</i> 41							1	*****	

## Carte Génétique

Pour déterminer l'origine de la distorsion de ségrégation (zygotique ou gamétique), nous avons appliqué le même test (Lorieux *et al.*, 1995) que celui présenté dans l'article Cadalen et ses collaborateurs (2009) : deux tests  $\chi^2$  successifs réalisés à partir des marqueurs codominants (OC1517H1, EU05D03H3 et EU10H05H5) présents sur le groupe de liaison 7. Le résultat des tests pour chaque marqueur analysé montre que les facteurs de sélection gamétique sont responsables de la distorsion observée sur le groupe de liaison 7.



**Figure 9 :** Bilan du travail réalisé pour la cartographie de séquences polymorphes et la construction de cartes génétiques.

Au cours de ma thèse, la détection de polymorphisme et la cartographie des séquences EST et SSR, indiquées en rouge, ont été réalisées sur la descendance F1'a. Les marqueurs AFLP été déjà génotypés, ils ont été inclus à la carte génétique F1'a. Une carte génétique F1'b a été réalisée à partir des marqueurs sélectionnés sur la carte génétique F1'a. Les données des cartes génétiques F1'a et F1'b, nous ont permis de construire une carte génétique F1'a+b pour l'analyse QTL.

### IV. Discussion

Cette étude a eu pour but d'inclure 17 marqueurs SSR dans la carte génétique K28\*K59 et de construire une carte génétique F1'a (Figure 9) à partir de certains marqueurs déjà existants sur la carte génétique K28\*K59 (Cadalen *et al.*, 2009) (Tableau 4) et, en incluant de nouveaux marqueurs STS (marqueurs codominants) et AFLP (marqueurs dominants) (Tableau 5). Cette démarche a permis d'augmenter le nombre de marqueurs et d'améliorer la densité moyenne de marqueur pour chaque groupe de liaison.

Dans un premier temps, la recherche de polymorphisme au sein des séquences EST et SSR a été entreprise afin de cartographier les marqueurs. Les différentes démarches dans la recherche de polymorphisme au sein des séquences EST et SSR ont permis de révéler au final 71 marqueurs polymorphes et de cartographier sur la carte génétique F1'a, 64 nouveaux marqueurs répartis sur les 9 groupes de liaisons.

L'ajout de 269 marqueurs AFLP dans la carte F1'a, nous a permis d'augmenter de manière significative le nombre de marqueurs, la densité moyenne de marqueurs par groupe de liaison mais aussi de couvrir des zones pour lesquelles aucun marqueur SSR et ou STS n'était assigné dans la carte génétique K28\*K59 (Cadalen *et al.*, 2009). Ainsi, le nombre de marqueurs est plus élevé dans la carte F1'a avec 518 marqueurs contre 274 marqueurs (248 SSR et 26 STS) pour la carte génétique K28\*K59 et de ce fait, la distance moyenne entre les marqueurs adjacents pour la carte génétique F1'a est réduite à 1,5 cM par rapport à la carte génétique K28\*K59 (2,7 cM). De manière similaire à la carte génétique K28\*K59 (Cadalen *et al.*, 2009), le nombre de marqueurs est élevé pour les groupes de liaison LG1 et LG8 avec plus de 98 marqueurs cartographiés. Les groupes de liaison 4 et 5 ont un nombre de marqueurs plus faible avec 33 et 34 marqueurs, respectivement.

## Carte Génétique

**Tableau 4** : Nombre, type et densité moyenne de marqueurs pour chaque groupe de liaison pour la carte K28\*K59 (Figure 7).

	LG1	LG2	LG3	LG4	LG5	LG6	LG7	LG8	LG9	Total
<b>SSR</b>	15	19	23	10	13	14	17	39	10	160
<b>STS</b>	3	2	3	3	2	4	1	6	4	28
<b>N. total de marqueurs</b>	18	21	26	13	15	18	18	45	14	188
<b>Longueur (cM)</b>	95	69	89	58	70	94	70	98	112	755
<b>Densité moyenne (cM)</b>	5.3	3.3	3.4	4.2	4.6	5.2	3.9	2.2	6.9	4.3

**Tableau 5** : Nombre, type et densité moyenne de marqueurs pour chaque groupe de liaison pour la carte F1'a (Figure : 8) après cartographie des nouveaux marqueurs SSR et l'ajout de marqueurs AFLP et STS.

	LG1	LG2	LG3	LG4	LG5	LG6	LG7	LG8	LG9	Total
<b>SSR</b>	16	22	23	11	14	16	18	43	12	175
<b>STS</b>	10	8	9	7	4	5	7	16	8	74
<b>AFLP</b>	23	27	65	15	16	35	24	41	23	269
<b>N. total de marqueur</b>	49	57	97	33	34	56	49	100	43	518
<b>Longueur (cM)</b>	92	73	81	52	71	94	63	92	112	730
<b>Densité moyenne (cM)</b>	1.9	1.2	0.9	0.6	2.1	1.7	1.4	1	2.6	1.5

Dans le but d'une l'analyse QTL liée à l'embryogenèse somatique, nous avons entrepris de rechercher du polymorphisme aux sein des séquences EST exprimées différemment (analyse microarray et/ou PCR en temps réel) au cours de l'embryogenèse somatique, sur la carte génétique F1'a. Ainsi, 47 gènes ont pu être localisés.

Cependant, le nombre de marqueurs EST révélés polymorphes est relativement faible. De nombreux gènes candidats révélés par analyse microarray au 4<sup>ème</sup> jour de culture n'ont pu être cartographiés. La recherche de polymorphisme des séquences EST a été réalisée par migration des produits PCR par électrophorèse (SSCP). Cette méthode d'analyse est peut-être limitante et ne permet pas toujours de détecter tous les SNP et les InDel.

Pour 14 OC analysés, nous avons pu détecté la présence de polymorphisme *in silico* due à la présence de SNP. Cependant, le polymorphisme identifié *in silico* n'a été détecté que pour 9 OC après séparation des produits PCR en condition non dénaturantes et n'a pas été retrouvé pour toutes les séquences après migration des produits PCR.

Les alignements entre les séquences d'*Arabidopsis* et les séquences EST de chicorée correspondantes, nous ont permis d'identifier des régions introniques. A partir de 98 couples d'amorces définis autour des introns et générant un produit PCR, 87 régions introniques ont pu être amplifiées. Seulement, 31 régions ont révélé du polymorphisme.

Nous avons fait le choix de définir des amorces à chaque extrémité des séquences EST pour lesquelles nous n'avons pas pu détectées la présence de polymorphisme *in silico* ou identifier des régions introniques, Cette démarche empirique, nous a permis de révéler du polymorphisme pour 9 séquences EST.

Pour augmenter le nombre de gènes cartographiés, une autre méthode plus sensible pour détecter les variations génétiques (SNP, mutation et méthylation) pourrait être employé. Pour détecter les variations à partir de petit amplicon, une nouvelle méthode rapide et moins coûteuse, appelée HRM (High Resolution Melting), basée sur des courbes de fusion a été décrite dans la littérature (Liew *et al.*, 2004 ; Reed *et al.*, 2004). Cette technique fait intervenir un marqueur ADN fluorescent qui permet de saturer rapidement le double brin d'ADN. Les échantillons sont analysés selon leur courbe de fusion. Les hétérozygotes sont distingués par une courbe de

fusion modifiée alors que les différents homozygotes sont différenciés par un changement dans la température de fusion.

Au final, 63 gènes candidats impliqués dans l'embryogenèse somatique sont localisés sur les 9 groupes de liaison de la carte génétique F1'a et/ou K28\*K59 (Tableau 6). En effet, en plus des 47 gènes différentiellement exprimés lors de l'ES et cartographiés, 15 gènes, certains détaillés dans l'analyse bibliographique, ont été déjà cartographiés sur la carte génétique K28\*K59. De plus, le gène *STM* connu pour son implication dans l'embryogenèse somatique chez la chicorée (Da Silva, 2004) a été cartographié sur le groupe de liaison 6.

Outre le fait d'augmenter le nombre de marqueurs dans la carte F1'a, les marqueurs SSR et STS très informatifs peuvent s'avérer très efficaces pour établir des relations (synthénie) entre les groupes de liaisons d'autres cartes de la même espèce. De plus, une densité de marqueurs et une couverture du génome élevée pour l'ensemble des groupes de liaisons de la carte F1'a favorisent la détection de QTL. En effet, ce nombre élevé de marqueurs sur l'ensemble des groupes de liaison, nous a permis de sélectionner 124 marqueurs de la carte F1'a, répartis de manière plus ou moins homogène sur chaque groupe de liaison, de génotyper ces marqueurs sur 104 nouveaux individus de la sous-descendance F1'b et ainsi d'élaborer une carte génétique F1'a+b (Figure 9) robuste pour la détection de QTL lié à l'embryogenèse somatique et à l'organogenèse.

## Carte Génétique

**Tableau 6 :** Gènes candidats potentiellement impliqués dans le processus de l'embryogenèse somatique cartographiés sur la carte génétique K28\*K59 et/ou F1'a. En gras, les marqueurs cartographiés sur la carte génétique K28\*K59 et/ou F1'a. En rouge, les EST différentiellement exprimés au 4<sup>ème</sup> jour de l'embryogenèse somatique mis en évidence par analyse microarray et/ou PCR en temps réel. En vert, le gène *STM* identifié par les travaux de Sophie Da Silva et Brigitte Huss. Deux gènes candidats indiqués par le symbole (\*) n'ont pas révélé de polymorphisme dans la carte génétique K28\*K59. Au contraire, ils ont pu être cartographiés dans la carte génétique Rubis118.

Gènes cartographiés	Nature de la protéine déduite	Groupe de liaison
<b>CL022L1R1*</b>	<b>protéine kinase SERK</b>	<b>1</b>
<b>GST5'-RACE4</b>	<b>Glutathione-S-transferase</b>	<b>1</b>
OC0081	putative chloroplast outer membrane protein [ <i>Oryza</i> ]	1
OC0185	putative translation factor [ <i>Pinus pinaster</i> ]	1
OC0311	acidic ribosomal protein P0 [ <i>G. max</i> ]	1
OC0378_a	60S ribosomal protein L5 [ <i>Solanum melongena</i> ]	1
OC0581	S6 ribosomal protein [ <i>Asparagus officinalis</i> ]	1
OC1104	zinc finger protein [ <i>Pinus radiata</i> ]	1
OC1394	salt tolerance protein 5 [ <i>Beta vulgaris</i> ]	1
OC1531_a	putative ribosomal protein S10 [ <i>Arabidopsis thaliana</i> ]	1
OC0082_b	60s acidic ribosomal protein [ <i>Hyacinthus orientalis</i> ]	2
OC0686	unknown [ <i>Arabidopsis thaliana</i> ]	2
OC1359_a	60S ribosomal protein L17 [ <i>Oryza sativa (japonica cultivar-group)</i> ]	2
OC1528	NADH-glutamate dehydrogenase [ <i>Nicotiana plumbaginifolia</i> ]	2
OC2006	putative leaf development protein Argonaute [ <i>Arabidopsis thaliana</i> ]	2
OC2141	TSB2 tryptophan synthase beta-subunit [ <i>Arabidopsis thaliana</i> ]	2
<b>ESTE002C9</b>	<b>haloacid dehalogenase-like hydrolase family protein [<i>Oryza sativa (japonica cultivar-group)</i>]</b>	<b>3</b>
<b>ESTE002B3</b>	<b>GTP binding protein [<i>Gossypium hirsutum</i>]</b>	<b>3</b>
OC0026_a	glyceraldehyde 3-phosphate dehydrogenase [ <i>Solanum tuberosum</i> ]	3
OC0206	hydroxy-methylglutaryl-coenzyme A reductase [ <i>Nicotiana tabacum</i> ]	3
OC0347b	poly(ADP-ribose) polymerase [ <i>Zea mays</i> ]	3
OC1138	3-hydroxy-3-methylglutaryl-coenzyme A reductase [ <i>Hevea brasiliensis</i> ]	3
OC1952	NADPH-ferrihemoprotein reductase [ <i>Helianthus tuberosus</i> ]	3
OC2147	putative glycerol-3-phosphate dehydrogenase [ <i>Arabidopsis thaliana</i> ]	3
<b>MADS box</b>	<b>MADS box protein [<i>Cichorium intybus</i>]</b>	<b>4</b>
<b>ESTE002B7</b>	<b>amino acid permease 6 [<i>Arabidopsis thaliana</i>]</b>	<b>4</b>
OC0390	40S ribosomal protein S2 [ <i>Arabidopsis thaliana</i> ]	4
OC1412	unknown [ <i>Arabidopsis thaliana</i> ]	4
OC1796_b	vegetative storage protein, VSP [ <i>Cichorium intybus</i> ]	4
OC1833_b	ribosomal protein L3A [ <i>Nicotiana tabacum</i> ]	4
<b>SK95UTR9L19R6</b>	<b>sérine protéine kinase SERK</b>	<b>5</b>
OC0839	dormancy-associated protein [ <i>Codonopsis lanceolata</i> ]	5
OC1635_a	lipid transfer protein (LTP) family protein [ <i>Arabidopsis thaliana</i> ]	5
<b>GTPB L3R1</b>	<b>GTP binding protein (GTP1)</b>	<b>6</b>
<b>3262L1R1</b>	<b>dehydrine</b>	<b>6</b>
CiSTM	shootmeristemless	6
OC0687	arabinogalactan protein [ <i>Daucus carota</i> ]	6
OC0344	Unknow protein	7
OC1145	delta-12 fatty acid desaturase [ <i>Borago officinalis</i> ]	7
OC1517	putative trypanothione-dependent peroxidase [ <i>Oryza sativa</i> ]	7
OC1696	P protein [ <i>Flaveria pringlei</i> ]	7

## Carte Génétique

OC1821	40S ribosomal protein S11 [ <i>Euphorbia esula</i> ]	7
OC1932	hybrid proline-rich protein [ <i>Catharanthus roseus</i> ]	7
3142L2R1*	<b>hydroxy-methylglutaryl-coenzyme A synthase</b>	<b>8</b>
3704L2R1	<b>lipid Transfer Protein 2</b>	<b>8</b>
3361L1R1	<b>lipid Transfer Protein 1</b>	<b>8</b>
<b>Tubulin L2R2</b>	<b>B-tubulin</b>	<b>8</b>
OC0023_a	glutathione transferase [ <i>H. muticus</i> ]	8
OC0031	peroxidase 3 [ <i>Scutellaria baicalensis</i> ]	8
OC0347a	poly(ADP-ribose) polymerase [ <i>Zea mays</i> ]	8
OC0565_a	acetoacetyl-CoA thiolase [ <i>Arabidopsis thaliana</i> ]	8
OC0748	ribosomal protein L7 [ <i>Solanum tuberosum</i> ]	8
OC0812	RNA Binding Protein 47 [ <i>Nicotiana plumbaginifolia</i> ]	8
OC0818	glutamate synthase (ferredoxin) [ <i>Nicotiana tabacum</i> ]	8
OC1822_b	putative ribosomal protein L10a [ <i>Oryza sativa (japonica cultivar-group)</i> ]	8
OC1976	delta 12 fatty acid desaturase [ <i>Crepis palaestina</i> ]	8
OC2060	lipid transfer protein [ <i>Helianthus annuus</i> ]	8
<b>GTPB L2R2</b>	<b>GTP binding protein (GTP1)</b>	<b>9</b>
<b>3474 L1R1</b>	<b>hemoglobine 1</b>	<b>9</b>
OC1929	G protein beta subunit [ <i>N. plumbaginifolia</i> ]	9
OC0183	salt tolerance protein 5 [ <i>Beta vulgaris</i> ]	9
OC0603_a	cobalamine-independent methionine synthase [ <i>S. scutellarioides</i> ]	9
OC1580	40S ribosomal protein S6 [ <i>Cicer arietinum</i> ]	9

## Chapitre III

### Détections de QTL liées à la régénération *in vitro*

## I. Introduction

La grande majorité des connaissances actuelles concernant le programme génétique induit au cours de l'embryogenèse zygotique chez les végétaux supérieurs provient de l'étude des mutants (Mayer *et al.*, 1993 ; Shevell *et al.*, 1994 ; Hardke *et al.*, 1998 ; Steinmann *et al.*, 1999 ; Schrick *et al.*, 2000). Si à ce jour, aucun mutant spécifique de l'embryogenèse somatique (ES) n'est connu, l'étude des mutants zygotiques a contribué à l'augmentation des connaissances sur l'embryogenèse somatique. Mordhost et ses collaborateurs (Mordhost *et al.*, 1997, 1998) ont étudié l'embryogenèse somatique à partir de plusieurs mutants zygotiques d'*Arabidopsis*. Lorsque des graines du mutant *pt* (*primordia timing*) sont placées sur un milieu de germination riche en 2,4-D, la formation d'embryons somatiques a été observée. Les embryons somatiques présentent un méristème apical caulinaire (SAM) élargi et un profil caractéristique des embryons zygotiques malgré un nombre élevé et plus variable de cellules. Des clusters de cellules embryogènes ont été obtenus chez d'autres mutants comme *clavata* (*clv*), avec un méristème apical caulinaire également élargi. Chez le double mutant, *ptclv*, la fréquence de plantules embryogènes est très élevée. Alors un nombre de cellules élevées en division et des cellules méristématiques non engagées dans le méristème apical caulinaire faciliteraient l'établissement d'embryons somatiques chez *Arabidopsis*.

Au contraire, les mêmes auteurs ont montré que des cultures de cellules embryogènes pouvaient être obtenues à partir de mutants sans méristème apical caulinaire embryogène ; *shoot meristemless* (*stm*), *wushel* (*wus*) et *zwill/pinhead* (*zll/pinh*). Un SAM fonctionnel n'est peut-être pas nécessaire pour l'ES chez *Arabidopsis*.

Le mutant *pickle* (*pkl*) d'*Arabidopsis*, présente aussi des caractères embryogènes après germination (Ogas *et al.*, 1997, 1999). Le clonage et la caractérisation du gène *PKL* ont révélé que ce gène code pour un facteur de remodelage de la chromatine CHD3. Ce gène est impliqué dans la voie de développement modulée par l'acide gibbérellique qui réprime l'identité embryogène des cellules durant la germination en inhibant l'expression de certains gènes. Chez le mutant *pickle*, un gène *LEC1* codant pour un facteur de transcription favorisant l'identité embryogène des cellules est déréprimé.

Des facteurs de remodelage de la chromatine semblent nécessaires pour établir un nouveau profil d'expression de certains gènes et ainsi favoriser le développement d'embryons somatiques.

Des méthodes de transcriptomique, permettant de visualiser l'expression différentielle des gènes au cours de l'ES ont été utilisées afin de caractériser des gènes clés du développement embryonnaire (Thibaud-Nissen *et al.*, 2003; Legrand *et al.*, 2007). Cette approche qui fournit une idée globale sur l'expression des gènes pendant le processus de l'ES chez les végétaux est une source d'information importante pour la sélection des gènes candidats. En effet, l'objectif d'une analyse QTL est d'identifier les régions chromosomiques puis les gènes ayant un effet significatif sur les caractères d'intérêts. Les analyses transcriptomiques réalisées, ont permis de révéler de nombreux candidats pour lesquels il sera possible de tester leur effet sur la variation du caractère étudié.

L'avancé des marqueurs moléculaire a fourni une nouvelle opportunité pour l'analyse des variations qualitatives et quantitatives. Par les études d'association marqueurs-caractères, l'analyse QTL facilite la détection des facteurs génétiques influençant le caractère quantitatif et permet de les localiser sur des régions spécifiques du génome. De plus, cette analyse fournit l'effet de chaque QTL et leur contribution dans la variation du caractère étudié (Vreugdenhil *et al.*, 2006).

Plusieurs auteurs se sont intéressés à identifier des QTL liés à la régénération *in vitro* (revue ; Bolibok *et al.*, 2006). La capacité de régénération des végétaux est majoritairement un caractère quantitatif et héritable. La recherche de QTL a été réalisée à partir de différents explants et les caractères analysés ont porté sur la capacité à initier des cals, à régénérés des bourgeons mais aussi à former des embryons somatiques (Bolibok *et al.*, 2006). Contrairement aux caractères liés à la callogenèse et à l'organogenèse, la recherche de QTL liée à l'embryogenèse somatique a fait l'objet de peu d'études. Des locus impliqués dans l'ES ont été identifiés chez le maïs (Armstrong *et al.*, 1992), la tomate (Koornneef *et al.*, 1993), la luzerne (Yu and Pauls, 1993), l'orge (Mano *et al.*, 1996), le tournesol (Florres Berrios *et al.*, 2000b ; Huang *et al.*, 2007), le blé (Hayian *et al.*, 2007) et le seigle (Bolibok *et al.*, 2007) mais aucun gène majeur impliqué dans l'ES a été identifié. Au contraire, chez le riz l'approche QTL a permis d'isoler un gène majeur impliqué dans la régénération *in vitro* (Nishimura *et al.*, 2005). Ce gène code pour une enzyme clé (NiR, nitrite reductase) de l'assimilation des nitrates. Des cals transformés avec le

gène NiR chez le parent peu organogène a permis d'augmenter de manière significative le nombre de bourgeons formés.

La mise en place d'une méthode simple pour comptabiliser les embryons somatiques a permis d'évaluer le caractère embryogène des sous-descendances F1'a et F1'b issues des parents K28\*K59 (Chapitre I). Nous avons considéré que les plantules formées (PL) après induction d'explants racinaires en condition d'embryogenèse somatique étaient issues du développement des embryons somatiques. Lors de l'induction, des structures chlorophylliennes unipolaires (SH) ont également été observées. Ne connaissant pas l'origine de ces structures, nous avons préféré les distinguer des plantules formées. Enfin, nous avons considéré l'ensemble des structures formées (ET) après induction en condition d'embryogenèse somatique ; les plantules et les structures chlorophylliennes unipolaires.

Nous nous sommes également intéressés au potentiel organogène des génotypes de la sous-descendance F1'b (Chapitre I). Pour cette étude, le nombre de calcs initié et le nombre de bourgeons régénérés à partir de fragments de pétioles pour chacun des génotypes ont été comptabilisés. La collecte de ces données avait pour but de déterminer la présence ou non de QTL communs pour la formation des plantules et des structures chlorophylliennes unipolaires issues de l'embryogenèse somatique et, la formation des calcs et des bourgeons issues de l'organogenèse.

Les distributions des valeurs phénotypiques pour chaque caractère (PL, SH, bg et Cal) ont révélé une distribution normale et continue. Les caractères liés à PL et SH sont quantitatifs et hautement héritables, avec une héritabilité ( $h^2_L$ ) supérieure à 70% et 63% pour les sous-descendances F1'a et F1'b, respectivement (Chapitre I). L'héritabilité liée à la régénération des bourgeons est proche de celle pour les caractères PL et SH avec 65%. Au contraire, l'héritabilité liée à la callogenèse est inférieure ( $h^2_L = 35\%$ ) et, est expliquée par un effet génotypique plus faible (Chapitre I).

Un outil essentiel pour une analyse QTL est de disposer d'une carte génétique. La recherche de marqueur moléculaire au sein du laboratoire a permis de construire une carte de référence pour la chicorée (*Cichorium intybus* L.) (Cadalen *et al.* 2009). Nous avons augmenté la densité de marqueurs par groupe de liaison dans la carte génétique K28\*K59 en ajoutant 269 marqueurs AFLP et 17 marqueurs SSR, constituant ainsi une carte génétique F1'a. De plus, 47 gènes candidats, mis

en évidence par analyse transcriptomique, au 4<sup>ème</sup> jour d'induction de l'ES, chez le génotype embryogène (K59) et non embryogène (C15) ont été cartographiés. A partir de la carte F1'a, nous avons sélectionnés 124 marqueurs (52 SSR, 33 STS, 39 AFLP) que nous avons génotypé sur les ADN de la sous-descendance F1'b (n=104). Puis, une carte génétique F1'a+b a été construite à partir des données génotypiques des sous-descendances F1'a et F1'b pour ces marqueurs afin de réaliser une analyse QTL (Chapitre 2).

Dans ce chapitre, nous présentons l'identification des régions chromosomiques chez la chicorée associée au nombre de plantules (PL) issues des embryons somatiques et au nombre de structures chlorophylliennes unipolaires (SH) obtenues à partir d'explants racinaires dans la descendance F1'a+b. De plus, l'analyse QTL a été réalisée afin d'identifier les régions liées à la formation des cals (Cal) et des bourgeons (Bg) dans les explants foliaires de la sous-descendance F1'b. La différence phénotypique observée entre les sous-descendances F1'a et F1'b (Chapitre I et II), nous a conduit à réaliser une analyse QTL sur chaque sous-descendance. Ensuite, l'analyse QTL a été conduite à partir de la descendance F1'a+b.

## II. Matériels et méthodes

### 1. Détection des QTL

Le logiciel MapQTL® 4.0 (Van Ooijen et *al.*, 2002) a été employé pour la recherche de QTL. Les outils statistiques utilisés ont été, dans l'ordre d'application :

- Le test non paramétrique de Kruskal-Wallis (KW) (Lehmann, 1975) pour obtenir une idée générale de la présence et de la localisation des QTL. Ce test est l'équivalent non paramétrique d'une analyse de variance. Le modèle correspond à la recherche d'un effet individuel au marqueur.
- La cartographie par intervalle (Interval Mapping ou IM) (Stam, 1993 ; Jansen and Stam, 1994) qui a donné une meilleure précision dans la localisation des QTL et permis de sélectionner le marqueur le plus proche des QTL comme cofacteurs pour l'analyse suivante.

- Enfin, l'analyse par le modèle des QTL multiples (multiple QTL Model ou MQM ; Van Ooijen et *al.*, 2002) après la sélection d'un jeu de cofacteurs adéquats.

Ce modèle inclut les marqueurs proches des QTL identifiés dans le modèle simple de cartographie d'intervalle dans l'analyse QTL comme cofacteurs et ainsi permet d'identifier des QTL à effets mineurs non détectables dans le modèle simple. La sélection de cofacteurs permet "d'absorber" l'effet d'un QTL majeur pour identifier les QTL à effets mineurs. La sélection du cofacteur pour chaque QTL a été réalisée avec la fonction "*Automatic cofacteur selection*" du logiciel MapQTL et consiste à sélectionner le marqueur le plus lié significativement au QTL.

L'intervalle d'analyse choisi pour l'IM et le MQM a été de 2 cM et 4 cM, respectivement.

Des tests de permutation (Doerge and Churchill, 1996) ont permis de fixer le seuil de détection des QTL à 3,1. Dans ces tests, les données sont échantillonnées un grand nombre de fois en gardant les informations aux locus mais en redistribuant au hasard les valeurs phénotypiques. Deux cent tests de permutation ont été réalisés et les valeurs obtenues pour chaque test ont été moyennées afin de déterminer le seuil de détection des QTL.

La région chromosomique avec la valeur de LOD score la plus élevée a été considérée comme la meilleure position des QTL et un intervalle de confiance d'une valeur LOD égale 2 a été défini autour de cette valeur. La proportion de variation phénotypique expliquée par chaque QTL a été déterminée par le coefficient de détermination ( $R^2$ ). Chaque QTL détecté a été codé par l'abréviation du caractère (PL : plantules issues du développement d'embryons somatiques, SH : structures chlorophylliennes unipolaires et ET : événements totaux), puis le paramètre d'analyse utilisé ( $_{90}$ , intensité moyenne de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires et  $_{50}$ , intensité maximale de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires) et enfin, le numéro du groupe de liaison auquel il a été assigné. Par exemple, le QTL identifié sur le groupe de liaison 2 pour le caractère plantule avec les données représentant l'intensité moyenne de réponse liée à l'embryogenèse somatique a été codé : PL $_{90}$ 2

Si un QTL a été détecté quel que soit le paramètre d'analyse (PL<sub>902</sub> et PL<sub>502</sub>), il a été codé seulement par l'abréviation du caractère et le numéro du groupe de liaison auquel il a été assigné (PL2).

De la même manière, les QTL liés à la formation de cals et de bourgeons ont été codés par l'abréviation du caractère (Cal pour cal et Bg pour bourgeon) et le numéro du groupe de liaison auquel il a été assigné.

## 2. Calcul des Interactions

Le calcul des interactions a été basé sur un modèle d'analyse de variance à 2 facteurs, en prenant comme premier facteur le marqueur présent sur un groupe de liaison et en considérant tous les autres marqueurs présents sur les autres groupes de liaison de la carte comme deuxième facteur, permettant ainsi de tester les effets d'épistasie.

Le modèle s'écrit :  $p_{ijk} = \mu + M_j + M_k + M_j M_k + \sum_{ijk}$

$P_{ijk}$  = valeur phénotypique du descendant  $i$  en considérant dans le modèle les marqueurs  $j$  et  $k$ .

$\mu$  = moyenne phénotypique

$M_j$  = valeur génotypique au marqueur  $j$

$M_k$  = valeur génotypique au marqueur  $k$

$\sum_{ijk}$  = résiduelle du modèle

Le calcul d'interaction a été réalisé à l'aide du programme SAS (procédure GLM, SAS 9.1 2002-2003) pour chaque sous-descendance F1'a et F1'b. Avec 11574 interactions, un seuil minimal de  $P < 0.0005$  a été considéré comme significatif.

## III. Résultats

1. Analyse des QTL liée à la formation des plantules, des structures chlorophylliennes unipolaires et aux évènements totaux

L'analyse des données phénotypiques (PL, SH et ET) (Chapitre 1) mais également la construction de la carte génétique (Chapitre 2) pour la détection de

QTL ont révélé des différences entre les deux sous-descendances étudiées F1'a et F1'b. Nous avons donc réalisé une analyse QTL d'une part avec les données génotypiques et phénotypiques de la sous-descendance F1'a puis ensuite avec ceux de la sous-descendance F1'b et ceci pour chaque caractère (PL, SH et ET) et pour chaque paramètre (PL<sub>90</sub>, SH<sub>90</sub>, ET<sub>90</sub> et PL<sub>50</sub>, SH<sub>50</sub>, ET<sub>50</sub>) étudiés.

Les résultats IM et MQM ont été présentés sous forme de graphiques et regroupés dans des tableaux. Seules les données MQM sont présentées dans le chapitre, les données IM ont été placées en annexes. Les marqueurs significatifs aux tests de Kruskal-Wallis sont indiqués par un symbole (◆) au niveau des graphiques.

### 1.1. Analyse des QTL dans la sous-descendance F1'a

#### 1.1.1. Détections des QTL liées à la formation de plantules

Les tests de Kruskal Wallis suggèrent l'existence de QTL liés à la formation de plantules sur 4 groupes de liaisons (LG1, LG2, LG6 et LG8). Les marqueurs compris entre 58 et 93 cM sur le groupe de liaison 1, montrent un effet significatif. Seuls quatre marqueurs (OC2006E, OC0082E, OC1359 et EU05B10H5) montrent un effet significatif pour le groupe de liaison 2 (LG2). Dans le cas du groupe de liaison 6, 6 marqueurs (OC0687R2L2, B12CL1R1E, CiSTM, tgct75E, tgta165E et EU05B04H3) sont significativement associés avec la formation des plantules. De même, 4 marqueurs (OC0812E, OC1976, E07AL1R1 et OC0023E) situés dans la région proximale du groupe de liaison 8 montrent un effet très significatifs ( $p < 0.005$ ).

L'analyse IM (Annexes 10 et 11), a permis d'identifier 2 QTL (PL1 et PL6) pour les paramètres PL<sub>90</sub> et PL<sub>50</sub>. Sur le groupe de liaison 1, un QTL (PL1) a été détecté expliquant environ 16% de variation phénotypique. Un autre QTL localisé sur le groupe de liaison 6 (PL<sub>90</sub>, LOD=3,31 et PL<sub>50</sub>, LOD=3,22) et expliquant environ 12% de variation phénotypique a été identifié. Sur le groupe de liaison 8, un QTL putatif avec un LOD score légèrement inférieur au seuil de détection (PL8, LOD=2,94) a été détecté dans la région comprise entre 64 et 92 cM.

L'analyse MQM (Tableau 1 et Figure 2) a été réalisée en sélectionnant comme jeu initial 3 cofacteurs (PPX1R3L3 (LG1), sw4D08H3 (LG1) et CiSTM (LG6))

positionnés au niveau des QTL détectés lors de l'analyse IM. La sélection de ces cofacteurs, nous a permis de confirmer la présence du QTL (PL2) pour le groupe de liaison 2 avec un LOD score égale à 4,29 (PL<sub>90</sub>2) et 3,6 (PL<sub>50</sub>2) expliquant chacun environ 10% de la variation phénotypique. Le marqueur OC0082E (LG2) significativement lié à ce QTL a été ajouté au jeu de cofacteurs initial et une deuxième analyse a été lancée. Aucun autre QTL n'a été détecté parmi les 9 groupes de liaison.

Le QTL PL<sub>90</sub>1 identifié par analyse IM n'est plus significative (LOD = 2,2) après la sélection des cofacteurs alors qu'au contraire le QTL PL<sub>50</sub>1 (LOD = 3,62) a été confirmé. Cependant, même si le QTL n'est plus significatif, l'allure de la courbe définie à partir des valeurs LOD pour PL<sub>90</sub>1 est très similaire à celle de PL<sub>50</sub>1. Ce résultat montre l'importance de l'analyse des données phénotypiques dans la détection de QTL.

Les 3 QTL (PL<sub>50</sub>1, PL<sub>50</sub>2 et PL<sub>50</sub>6) liés à la formation des plantules expliquent au total environ 36% de la variation phénotypique observée. Pour chaque QTL identifié, l'allèle parental intervenant positivement sur la variation du caractère a été déterminé à partir du marqueur sélectionné comme co-facteur. Les données des effets alléliques ont été regroupées en annexe 14. Pour le QTL PL1, c'est la combinaison entre les allèles parentaux K59 (*f*) et K28 (*g*) qui semble favorable pour augmenter le nombre de plantules. Pour le QTL PL2, c'est la présence de l'allèle (*m*) de K59 qui contribue à augmenter la capacité embryogène des individus.

Pour le QTL PL6, la classe génotypique // du marqueur CiSTM révèle une valeur phénotypique plus élevée par rapport à la classe génotypique *Im*. Pour vérifier dans la classe génotypique //, si c'est l'allèle *l* apporté par K59 ou K28 qui est responsable de l'augmentation de la capacité embryogène, nous nous sommes intéressés aux autres marqueurs (B12CL1R1Ef et OC0687R1L1E) flanquants le marqueur CiSTM et significativement liés au QTL PL6.

## Analyse QTL

Marqueurs	cM	K59		K28		1	2	3	4	5							
tgct410Ef	0	l	m	l	l	ll	l	l	lm	m	l	lm	m	l	ll	l	l
FDCB39H3f	6	e	f	e	g	ee	e	e	ef	f	e	fg	f	g	fg	f	g
tacg118Nf	7	n	n	p	n	np	n	p	np	n	p	nn	n	n	nn	n	n
OC0687R2L2E	14	l	m	l	l	ll	l	l	lm	m	l	lm	m	l	lm	m	l
B12CL1R1Ef	14	m	l	l	l	lm	m	l	ll	l	l	ll	l	l	ll	l	l
CiSTM	16	m	l	l	l	lm	m	l	ll	l	l	ll	l	l	ll	l	l
3262L1R1H1H5f	20	a	b	c	d	ac	a	c	bc	b	c	bd	b	d	bd	b	d
tgct75Ef	29	m	l	l	l	lm	m	l	ll	l	l	ll	l	l	ll	l	l
tgta165Ef	36	l	m	l	l	ll	l	l	ll	l	l	lm	m	l	ll	l	l
EU05B04-H3	42	e	f	g	e	eg	e	g	eg	e	g	ef	f	e	ee	e	e
ttat177Ef	54	l	m	l	l	ll	l	l	ll	l	l	lm	m	l	ll	l	l
FDCA134H5f	63	b	a	d	c	bd	b	d	bd	b	d	ac	a	c	bc	b	c
tacg191Ef	63	m	l	l	l	lm	m	l	lm	m	l	ll	l	l	lm	m	l
EU05C06-H5	82	a	b	c	d	ac	a	c	ac	a	c	bd	b	d	ad	a	d
EU04B02-H1	90	k	h	h	k	hk	k	h	hk	k	h	hk	h	k	kk	k	k
atgc262Nf	98	n	n	p	n	np	n	p	np	n	p	nn	n	n	nn	n	n

**Figure 1:** Génotypes graphiques multilocus construits pour les individus 1, 2, 3, 4 et 5 à partir des données génotypiques de K59 et K28 pour les marqueurs du groupe de liaison 6. Les allèles indiqués dans des cases bleu ciel et turquoise sont apportés par K59, et en rose et mauve par K28.

Si l'on observe le génotype graphique (Figure 1), le marqueur B12CL1R1 est dans la même phase que le marqueur CiSTM et ne nous permet pas de conclure quand à l'origine de l'allèle *l* responsable d'une augmentation de la valeur phénotypique. Pour le marqueur OC0687R2L2E, c'est l'allèle *m* apporté par K59 qui contribue à augmenter le nombre d'embryons formés. Si l'on observe le génotype graphique, on peut voir que le marqueur OC0687R2L2E est en répulsion avec les deux autres marqueurs (CiSTM et B12CL1R1Ef) et l'allèle *m* apporté par K59 du marqueur OC0687R2L2 est associé à l'allèle *l* apporté par K59 pour les individus 2, 3 et 4. Alors, l'allèle *l* du marqueur CiSTM apporté par K59 est responsable de l'augmentation de la capacité embryogène des individus pour le QTL PL6.

Une association significative a été observée entre la capacité embryogène des individus et 5 marqueurs liés sur le groupe de liaisons 8 par le test de Kruskal-Wallis. Cependant, l'association caractère-marqueur n'était pas significative avec les autres analyses (IM et MQM), même si la couverture du génome sur ce groupe de liaison est très élevée. Dans ce cas, c'est peut-être l'effectif de la sous-descendance F1'a ( $n=123$ ) qui ne permet pas de détecter le QTL.

### 1.1.2. Détections des QTL liées à la formation des structures chlorophylliennes unipolaires

Les tests non paramétrique (Kruskal-Wallis) suggèrent l'existence de plusieurs QTL liés aux structures chlorophylliennes unipolaires utilisant un seuil  $p < 0.01$  (\*\*) de détection des QTL.

La méthode de cartographie par intervalle a permis d'identifier au moins 3 régions impliquées dans la formation des structures chlorophylliennes unipolaires (Annexe 10 et 11). Le QTL SH1, situé au niveau proximal du groupe de liaison 1, explique environ 20% de la variation phénotypique totale. De même, la présence de QTL sur le groupe de liaison 6 (SH<sub>90</sub>6, LOD=4,04 ; SH<sub>50</sub>6, LOD=3,30) et 8 (SH<sub>90</sub>8, LOD=3,41 ; SH<sub>50</sub>8, LOD=4,62) ont pu être identifiés. Ces QTL expliquent chacun plus de 12% de variation phénotypique observée (SH<sub>90</sub>6, 14.7%; SH<sub>50</sub>6, 12.2%; SH<sub>90</sub>8, 17,8%; SH<sub>50</sub>8, 18,2%).

Ensuite, 4 cofacteurs ont été sélectionnés comme jeu initial de marqueurs avant d'appliquer la méthode des QTL multiples (Tableau 1 et Figure 2). La sélection des cofacteurs, PPX1R3L3 (LG1), SW4D08H3 (LG1), C/STM (LG6) et OC1976H1 (LG8), nous a permis d'identifier une région supplémentaire liée à formation des structures chlorophylliennes unipolaires située sur le groupe de liaison 2, qui explique environ 10% de la variation phénotypique observée. Ensuite, l'ajout du cofacteur OC0082E (LG2), au jeu initial de cofacteur a permis de détecter un dernier QTL (SH<sub>50</sub>4 et SH<sub>90</sub>4) sur le groupe de liaison 4. Cependant, le QTL SH<sub>90</sub>1 du groupe de liaison 1 n'a pu être détecté suite à la sélection des cofacteurs alors qu'avec les données correspondantes à l'intensité maximale de réponse liée la formation des structures chlorophylliennes, le QTL SH<sub>50</sub>1 avec un LOD de 3,09 proche du seuil de détection a pu être identifié. Les 5 QTL (SH<sub>50</sub>1, SH<sub>50</sub>2, SH<sub>50</sub>6 et SH<sub>50</sub>8) liés à la formation des structures chlorophylliennes unipolaires expliquent 52,5% (SH<sub>50</sub>) de la variation phénotypique observée.

L'allèle favorable porté par les QTL SH2, SH6, SH8 provient du parent K59. Pour le QTL SH4, s'est l'allèle de K28 (*l*) qui est favorable à l'augmentation de la formation des structures chlorophylliennes unipolaires.

### 1.1. 3. Détections des QTL liées aux évènements totaux

Dans les analyses précédentes, plusieurs régions présentent sur les groupes de liaison 1, 2, 4, 6 et 8 montrent un effet significatif pour les caractères PL et SH. De plus, les courbes de LOD, obtenues après analyse IM et MQM, pour ces mêmes caractères sont très similaires quel que soit le QTL détecté et aucun QTL spécifique pour le caractère SH et PL n'a été détecté.

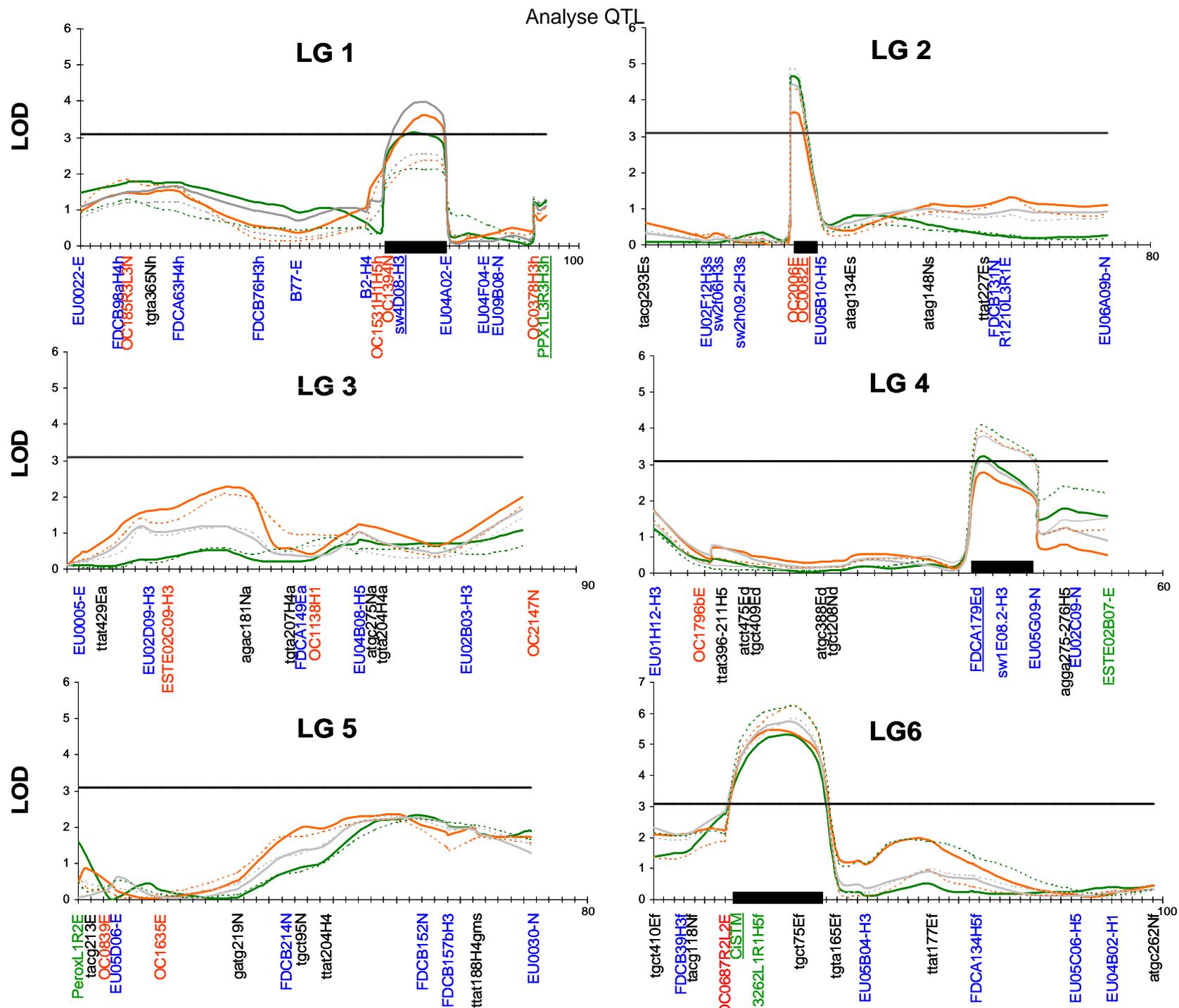
Nous avons alors entrepris de réaliser une analyse QTL à partir des événements totaux, correspondant à l'ensemble des données PL et SH pour un génotype donné. Ainsi, 5 régions identiques à celles détectées lors de l'analyse des données PL et SH ont été identifiées. Les QTL (ET1, ET2, ET4, ET6 et ET8) expliquent au total plus de 45% de variation phénotypique (Tableau 1 et Figure 2). La valeur phénotypique totale est proche de celle observée pour le caractère SH (SH<sub>90</sub>, 43,9% et SH<sub>50</sub>, 53,5%) et par contre, plus élevée que les valeurs phénotypiques totales obtenues pour le caractère PL (PL<sub>50</sub>, 27,3%; PL<sub>90</sub>, 36,2%).

**Tableau 1:** Analyse MQM (Multiple QTL Mapping). Identification des QTL liés à la formation des plantules (PL), des structures chlorophylliennes unipolaires (SH) et des événements totaux (ET) dans les sous-descendances F1'a, F1'b et la descendance F1'a+b. L'analyse a été réalisée pour chaque paramètre d'analyse (<sub>90</sub>, intensité moyenne de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires et <sub>50</sub>, intensité maximale de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires). La position du QTL (cM), le LOD score (LOD) et le pourcentage de variation expliquée ( $R^2$ ) sont indiqués pour chaque QTL. Les groupes de liaison non liés significativement à un QTL sont indiqués par –

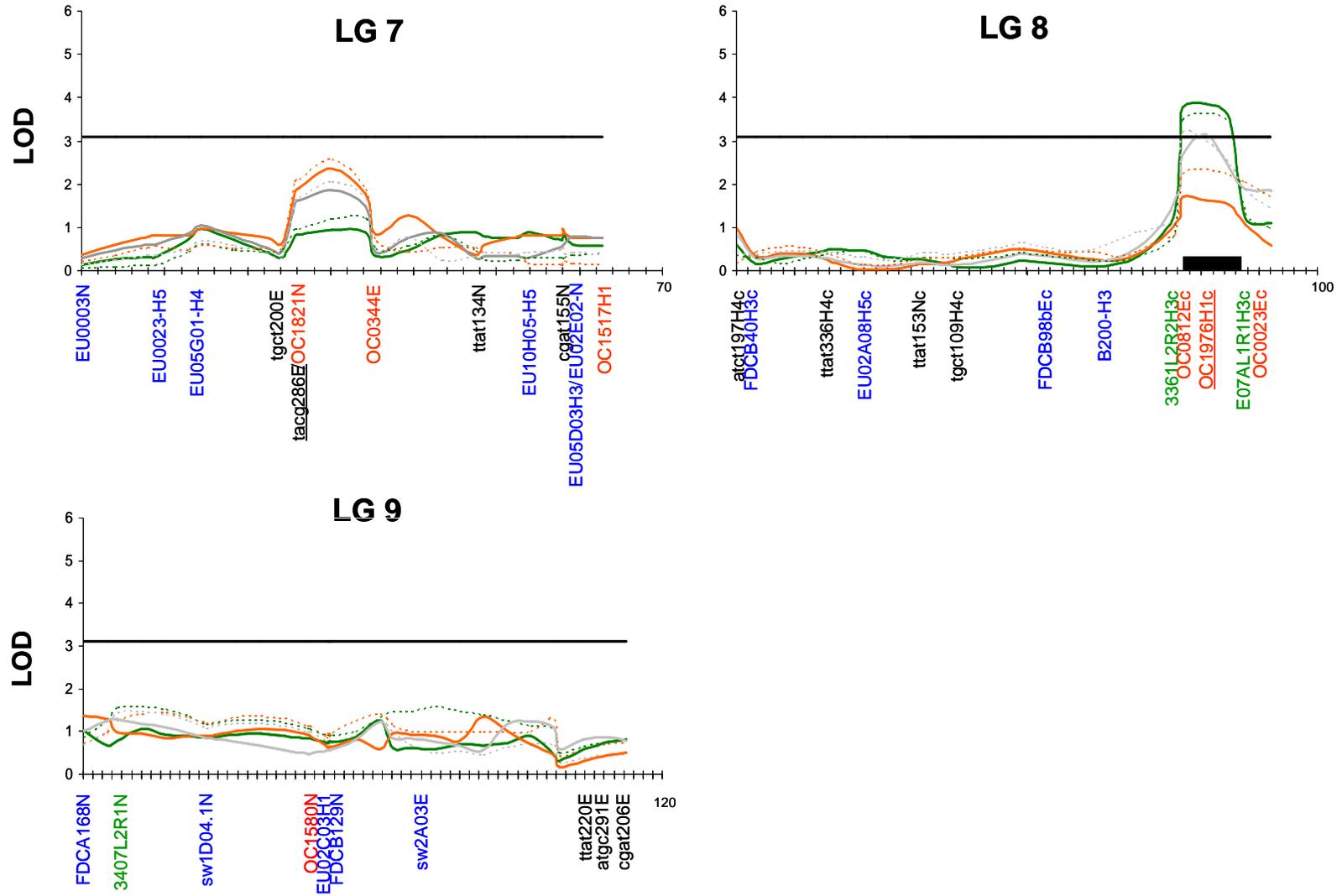
## Analyse QTL

QTL	LG	F1'a			F1'b			F1'a+b			
		POS. (cM)	LOD	R <sup>2</sup> %	POS. (cM)	LOD	R <sup>2</sup> %	POS.(cM)	LOD	R <sup>2</sup> %	
PL <sub>90</sub>	1		-	-		-	-		-	-	
	2	PL <sub>90</sub> 2	22-25 (23)	4.29	10.3	22-25 (23)	3.85	8.8	23-25 (23)	6.17	6.7
	3	PL <sub>90</sub> 3		-	-	36-44 (44)	4.88	12.3	37-40 (39)	3.85	8.3
	4			-	-						
	6	PL <sub>90</sub> 6	13-33 (22)	6.15	17		-	-	13-33 (22)	5.55	9.1
	7			-	-		-	-		-	-
	8	PL <sub>90</sub> 8		-	-	76-85 (81)	5.46	13.3	76-85 (85)	5.48	8
					<b>27.3</b>			<b>34.4</b>			<b>32.1</b>
PL <sub>50</sub>	1	PL <sub>50</sub> 1	64-73 (69)	3.62	11.2		-	-		-	-
	2	PL <sub>50</sub> 2	22-25 (23)	3.6	9.1		-	-	23-25 (23)	5.1	5.7
	3			-	-	36-44 (44)	4.7	12	37-40 (39)	3.84	8.8
	4			-	-						
	6	PL <sub>50</sub> 6	13-33 (22)	5.48	15.9		-	-	13-33 (22)	5.29	9.1
	7			-	-		-	-		-	-
	8	PL <sub>50</sub> 8		-	-	76-85 (81)	4.49	11	76-85 (85)	4.44	6.9
					<b>36.2</b>			<b>23</b>			<b>30.5</b>
SH <sub>90</sub>	1			-	-		-	-		-	-
	2	SH <sub>90</sub> 2	22-25 (23)	4.65	10	22-25 (23)	7.78	15.2	23-25 (23)	6.17	7.5
	3	SH <sub>90</sub> 3		-	-	36-44 (44)	6.65	13.7		-	-
	4	SH <sub>90</sub> 4	37-41 (39)	3.95	9.4		-	-		-	-
	6	SH <sub>90</sub> 6	13-33 (22)	6.15	14.5	13-33 (22)	4.8	9.7	13-33 (22)	8.37	11.6
	7			-	-		-	-		-	-
	8	SH <sub>90</sub> 8	76-85 (41)	3.61	10	76-85 (81)	6.08	14.3	76-85 (85)	8.8	12.6
					<b>43.9</b>			<b>52.9</b>			<b>31.7</b>
SH <sub>50</sub>	1	SH <sub>50</sub> 1	62-73 (69)	3.09	7.9		-	-		-	-
	2	SH <sub>50</sub> 2	22-25 (23)	4.67	10.6	22-25 (23)	6.34	12.9	23-25 (23)	5.41	6.7
	3	SH <sub>50</sub> 3		-	-	36-44 (44)	5.72	12.4		-	-
	4	SH <sub>50</sub> 4	37-41 (39)	3.03	8.1		-	-		-	-
	6	SH <sub>50</sub> 6	13-33 (22)	5.15	13.3	13-33 (22)	5.28	11.3	13-33 (22)	7.37	8.6
	7			-	-		-	-		-	-
	8	SH <sub>50</sub> 8	76-85 (81)	4.51	12.6	76-85 (81)	6.67	16.9	76-85 (85)	8.68	12.4
					<b>52.5</b>			<b>53.5</b>			<b>27.7</b>
ET <sub>90</sub>	1			-	-		-	-		-	-
	2	ET <sub>90</sub> 2	22-25 (23)	4.84	10.8	22-25 (23)	5.36	14.7	23-25 (23)	5.62	7
	3	ET <sub>90</sub> 3		-	-	36-44 (44)	3.86	9.9		-	-
	4	ET <sub>90</sub> 4	37-41 (39)	3.64	9.4		-	-		-	-
	6	ET <sub>90</sub> 6	13-33 (22)	5.78	15.1	13-33 (22)	3.6	8.2	13-33 (22)	7.52	11.1
	7			-	-		-	-		-	-
	8	ET <sub>90</sub> 8	76-85 (81)	3.17	10	76-85 (81)	3.92	11.4	76-85 (85)	6.05	9.8
					<b>45.3</b>			<b>44.2</b>			<b>27.9</b>
ET <sub>50</sub>	1	ET <sub>50</sub> 1	64-73 (69)	3.97	11.1		-	-		-	-
	2	ET <sub>50</sub> 2	22-25 (23)	4.37	10.5	22-25 (23)	6.9	12.7	23-25 (23)	6.66	7.9
	3	ET <sub>50</sub> 3		-	-	36-44 (44)	5.76	13.2		-	-
	4	ET <sub>50</sub> 4		-	-						
	6	ET <sub>50</sub> 6	13-33 (22)	5.68	15.7	13-33 (22)	4.14	8.8	13-33 (22)	7.33	10.7
	7			-	-		-	-		-	-
	8	ET <sub>50</sub> 8	76-85 (81)	3.14	10.7	76-85 (81)	5.55	11.6	76-85 (85)	7.05	11.1
					<b>48</b>			<b>46.3</b>			<b>29.7</b>

**Figure 2:** Courbes LOD score obtenues après analyse MQM des données plantules (PL ; en orange), structures chlorophylliennes unipolaires (SH ; en vert) et événements totaux (ET ; en gris) dans la sous-descendance F1'a. L'analyse a été réalisée pour chaque paramètre d'analyse ( $\mu_{90}$ , intensité moyenne de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait en pointillé) et  $\mu_{50}$ , intensité maximale de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait plein). Les régions liées aux QTL PL, SH et ET sont indiquées par un cadre noir. Les co-facteurs utilisés lors de l'analyse MQM sont soulignés.



Analyse QTL



## 1.2. Analyse des QTL dans la sous-descendance F1'b

### 1.2.1. Détections des QTL liées à la formation de plantules

La recherche de QTL liés aux nombres de plantules mais aussi aux nombres de structures chlorophylliennes unipolaires formées a été réalisée dans la sous-descendance F1'b par la même approche que dans la sous-descendance F1'a afin de comparer les analyses faites entre les deux sous-descendances.

Les tests de Kruskal Wallis ont révélé 4 régions significatives (LG2, LG3, LG6 et LG8) impliquées dans la formation des plantules avec une probabilité  $p < 0.01$  (\*\*) de détection des QTL (Annexe 12). Les régions LG2, LG6 et LG8 sont identiques à celles observées dans la sous-descendance F1'a et semblent liées aux QTL pour la capacité embryogène. De plus, une région située entre 36 et 48 cM sur le groupe de liaison 3 s'est révélée significative dans la sous-descendance F1'b par le test non paramétrique. Cette région n'avait pas été détectée dans la sous-descendance F1'a contrairement aux régions situées sur les groupes de liaison 2, 6 et 8.

Des tests de permutation ont aussi été réalisés pour la sous-descendance F1'b afin de calculer un seuil de détection des QTL. Ce seuil prenant en compte les données génotypiques et phénotypiques de chaque individu, il peut varier d'une sous-descendance à l'autre. Malgré les différences observées entre les deux sous-descendances, F1'a et F1'b, le seuil de détection des QTL est resté le même avec une valeur de 3,1.

L'analyse par cartographie d'intervalle confirme la présence de QTL sur les groupes de liaison 3 et 8 (Annexe 10 et 11) pour les deux paramètres  $PL_{50}$  et  $PL_{90}$ , alors que seules les données du paramètre  $PL_{90}$  ont permis d'identifier un QTL sur le groupe de liaison 2 ( $PL_{90}2$ ). Pour l'analyse MQM, 3 cofacteurs (OC0082E (LG2), FDCA149E (LG3) et OC1976H1 (LG8)) ont été sélectionnés pour les paramètres  $PL_{50}$  et  $PL_{90}$ , respectivement (Tableau 1 et Figure 3). L'analyse n'a pas confirmé, ni pour  $PL_{90}$  ni pour  $PL_{50}$ , la présence du QTL sur le groupe de liaison 6. Le QTL  $PL2$  détecté à partir des données  $PL_{90}$  n'a pas été retrouvé avec les données  $PL_{50}$ , même si cette région était significative lors du test de Kruskal Wallis.

La somme de variation phénotypique observée est de 34,4% pour le paramètre d'analyse  $PL_{90}$  mais est plus faible pour l'autre paramètre  $PL_{50}$  (23%). La différence s'explique par la détection du QTL ES2 pour le paramètre  $PL_{90}$  mais pas

pour le paramètre PL<sub>50</sub>. Pour les 3 QTL (PL2, PL3 et PL8) détectés, c'est l'allèle porté par K59 qui semble favorable à l'augmentation de la capacité embryogène des individus (Annexe 14).

### 1.2.2. Détections des QTL liées à la formation des structures chlorophylliennes unipolaires

De même que pour les données PL, les régions présentes sur les groupes de liaison 2, 3, 6 et 8 montrent également un effet significatif avec l'analyse des données SH par le test de Kruskal Wallis. L'analyse par cartographie d'intervalle (IM) a confirmée la présence de QTL pour les groupes de liaison 2, 3 et 8 (Annexes 10 et 12) mais pas pour le groupe de liaison 6 comme il avait été observé avec les données PL. Cependant, les courbes de LOD score pour SH<sub>906</sub> et SH<sub>506</sub> montrent une allure très semblable à celle observée pour le caractère PL.

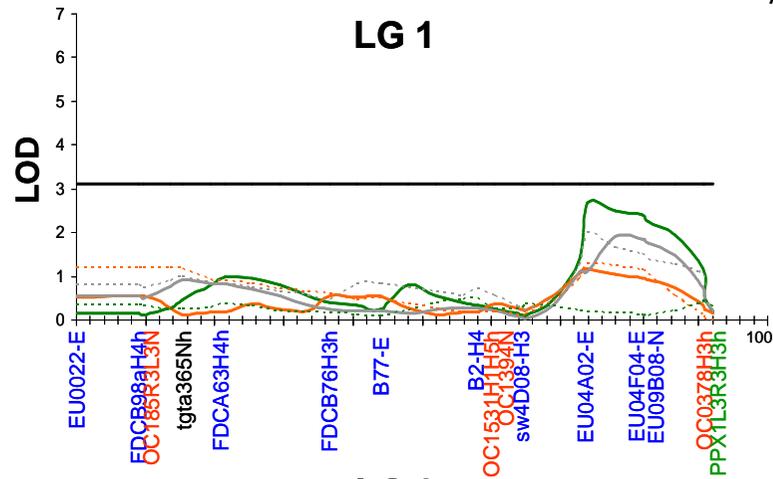
Trois cofacteurs ont été sélectionnés avec OC0082E (LG2), FDCA149E (LG3) et OC1976H1 (LG8) pour l'analyse des QTL multiples (Tableau 1 et Figure 3). Le choix de ces cofacteurs nous a permis de confirmer la présence du QTL (SH6) sur le groupe de liaison 6, déjà révélé par le test de Kruskal Wallis. Le QTL SH<sub>906</sub> et SH<sub>506</sub> explique 9,7% et 12,4% de variation phénotypique observée. Le marqueur significativement lié à ce QTL est le marqueur *CiSTM*. L'ajout de ce marqueur dans le jeu de cofacteur précédemment défini n'a pas permis de détecter d'autre QTL. L'ensemble des QTL, lié à la formation des structures chlorophylliennes unipolaires dans la sous-descendance F1'b, détecté explique entre 52,9 et 53,5% de variation phénotypique pour le caractère SH<sub>90</sub> et SH<sub>50</sub>, respectivement.

Les effets alléliques pour les QTL SH2, SH3 et SH8 (Annexe 14) sont identiques à ceux déterminés pour les données PL, à savoir un effet favorable de K59 sur l'augmentation de la formation des structures chlorophylliennes unipolaires.

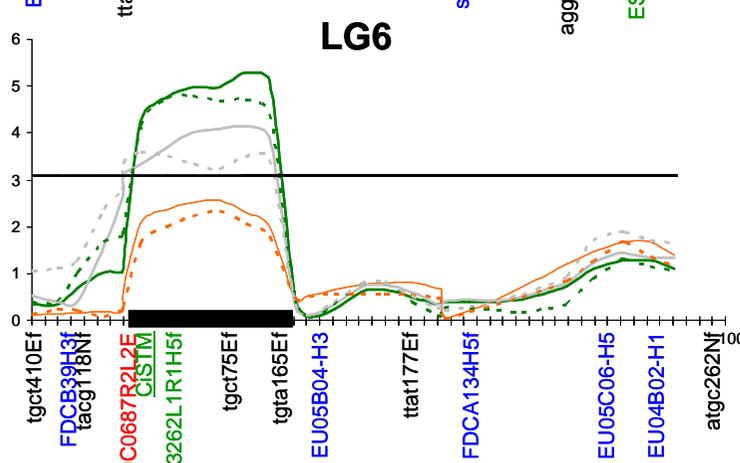
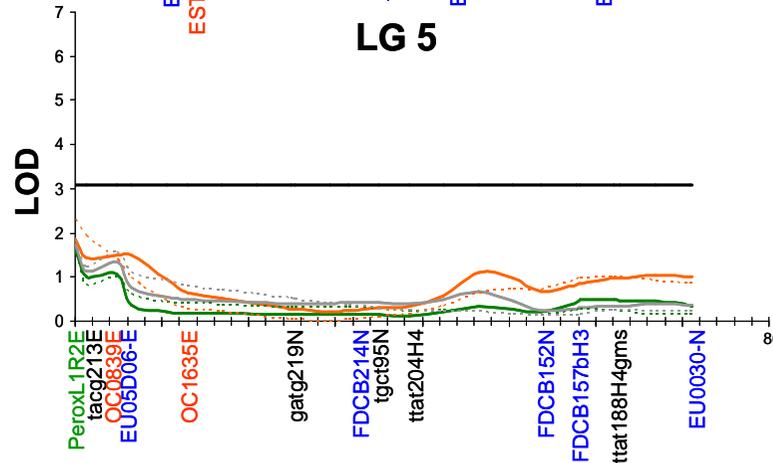
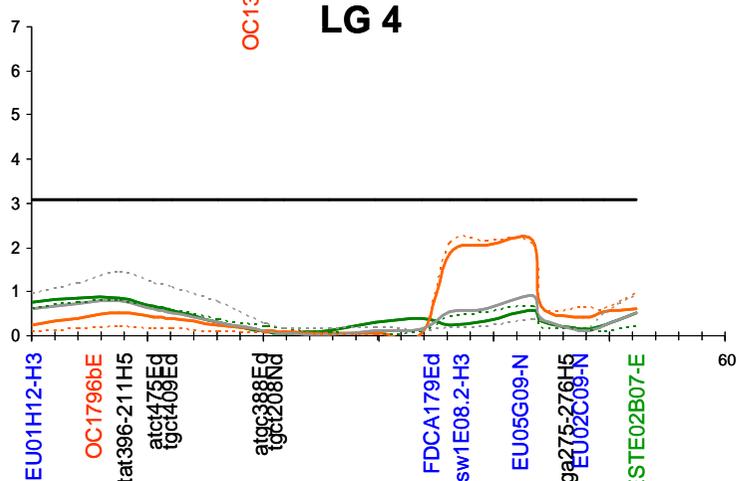
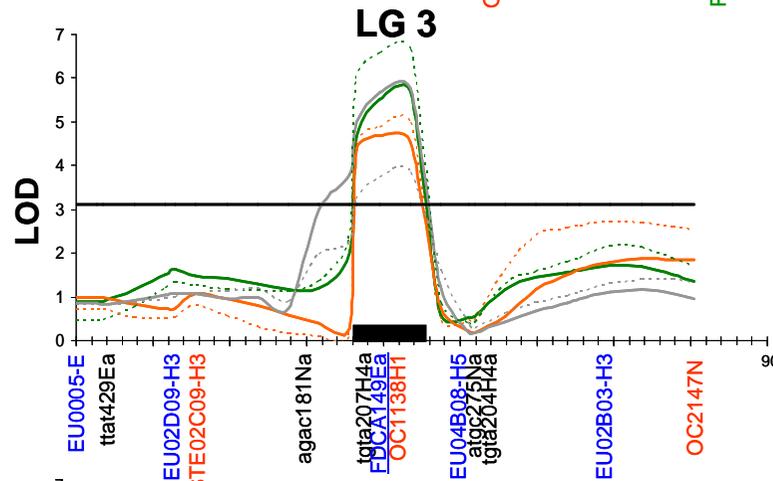
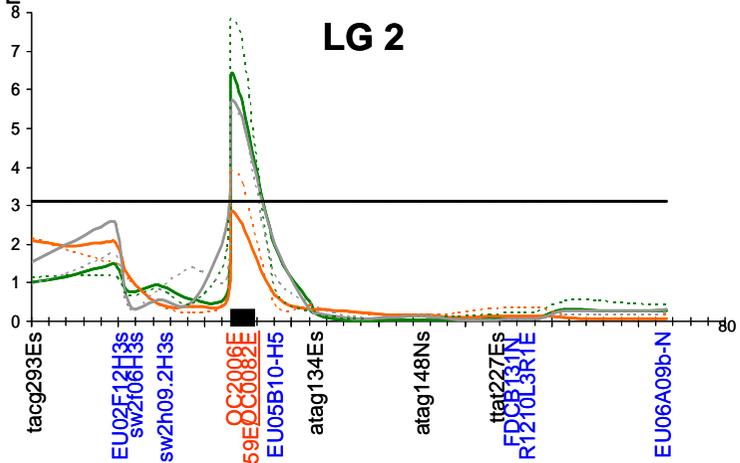
### 1.2.3. Détections des QTL liées aux évènements totaux

La recherche de QTL liée aux évènements totaux (Tableau 1) a été réalisée et nous a permis d'identifier 5 régions chromosomiques semblables à celles détectées lors de l'analyse des données PL et SH. Les QTL (ET1, ET2, ET4, ET6 et ET8) expliquent plus de 44% de variation phénotypique.

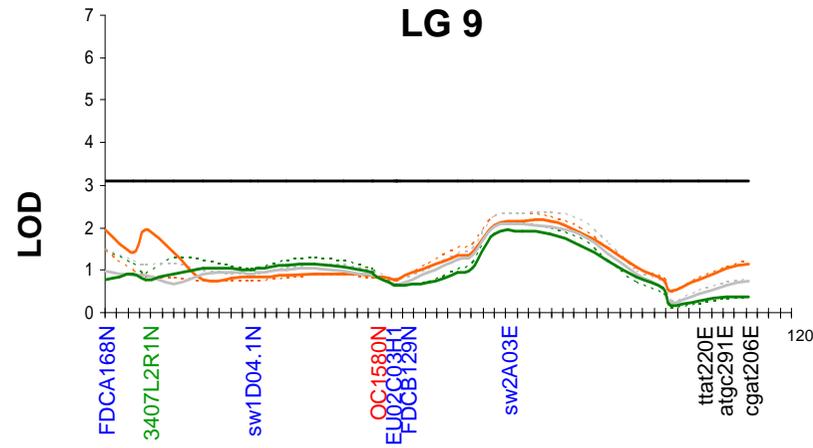
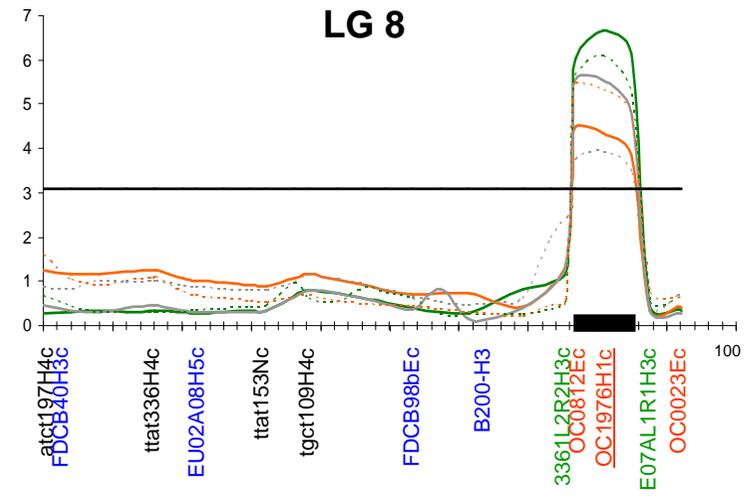
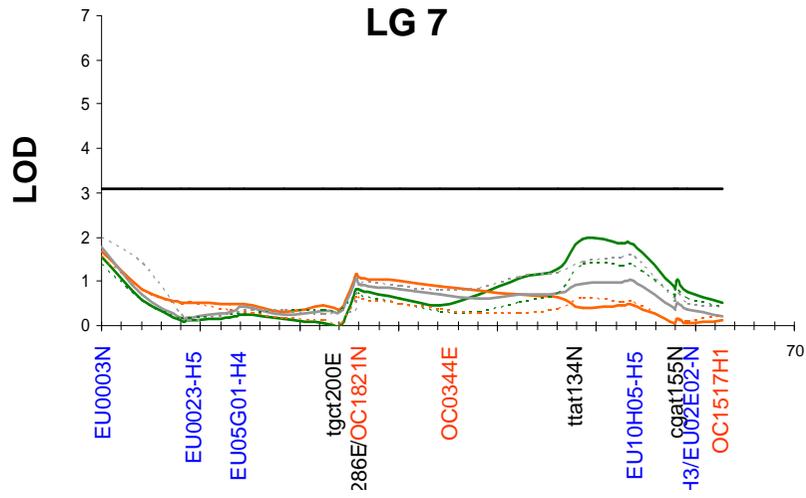
**Figure 3 :** Courbes LOD score obtenues après analyse MQM des données plantules (PL ; en orange), structures chlorophylliennes unipolaires (SH ; en vert) et événements totaux (ET ; en gris) dans la sous-descendance F1'b. L'analyse a été réalisée pour chaque paramètre d'analyse ( $\mu_{90}$ , intensité moyenne de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait en pointillé) et  $\mu_{50}$ , intensité maximale de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait plein). Les régions liées aux QTL PL, SH et ET sont indiquées par un cadre noir. Les co-facteurs utilisés lors de l'analyse MQM sont soulignés.



Analyse QTL



Analyse QTL



### 1.3. Analyse des QTL dans la descendance F1'a+b

L'observation des résultats de l'analyse MQM (Tableau 1) dans les deux sous-descendances F1'a et F1'b montre que 3 QTL majeurs sont retrouvés sur le groupe de liaison 2, 6 et 8 pour la capacité liée à la formation des structures chlorophylliennes (SH) et les événements totaux (ET) quelle que soit la sous-descendance étudiée. Au vu de ces résultats, nous avons appliqué une analyse QTL sur la descendance (F1'a+b). La même démarche appliquée pour les sous-descendances F1'a et F1'b, afin de détecter les QTL liés à l'embryogenèse somatique (PL), aux structures chlorophylliennes unipolaires (SH) et aux événements totaux (ET), a été entreprise. Ainsi, l'analyse par IM (Annexe 10 et 13) puis MQM (Tableau 1 et Figure 4) a permis de détecter 3 QTL sur les groupes de liaison 2, 6 et 8 pour la capacité lié aux PL, SH et ET. Ces QTL co-localisent avec les QTL identifiés dans les sous-descendances F1'a et F1'b. La part phénotypique observée pour chacun des caractères est en moyenne de 30%.

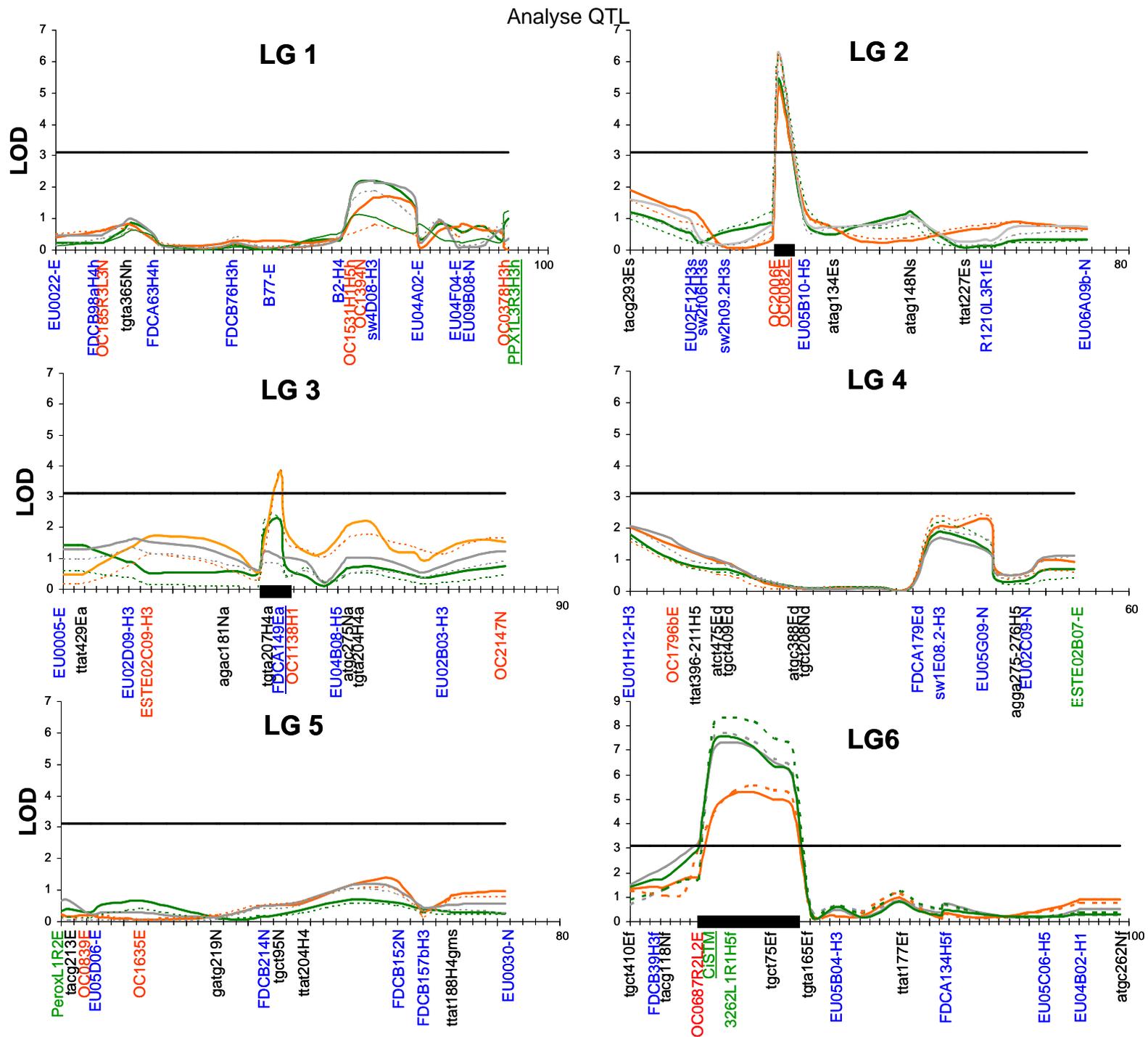
Cependant, les QTL identifiés sur les groupes de liaison 1 et 4 dans la sous-descendance F1'a n'ont pas été détectés dans la descendance F1'a+b. Lorsqu'on observe les effets alléliques (Annexe 14) des marqueurs sw4D08H3 et FDCA179 (liés significativement aux QTL PL1, SH1, ET1 et PL4, SH4, ET4, respectivement), dans la sous-descendance F1'b, les valeurs phénotypiques sont identiques entre les classes. Lorsque nous avons couplé les données des 2 sous-descendances F1'a et F1'b pour réaliser une analyse QTL pour la descendance F1'a+b, nous avons diminué l'effet observé dans la sous-descendance F1'a, ce qui explique que les QTL ne soient plus significatives dans la descendance F1'a+b. Les QTL détectés sur les groupes de liaison 1 et 4 semblent spécifiquement liés à la sous-descendance F1'a.

Au contraire, le QTL détecté sur le groupe de liaison 3 uniquement dans la sous-descendance F1'b a également été identifié dans la descendance F1'a+b. Les valeurs phénotypiques totales associées aux classes *lm* et *ll* du marqueur FDCA149 lié significativement au QTL du groupe de liaison 3 (Tableau 2), montrent un effet significatifs pour la classe *lm* quelle que soit la sous-descendance étudiée. Ainsi, il n'est pas très surprenant de détecter le QTL dans la descendance F1'a+b. Par contre, nous ne pouvons pas expliquer pourquoi aucun QTL sur le groupe de liaison 3 n'a été détecté dans la sous-descendance F1'a au cours de notre analyse, alors que la valeur phénotypique totale montre une valeur plus élevée pour la classe *lm*.

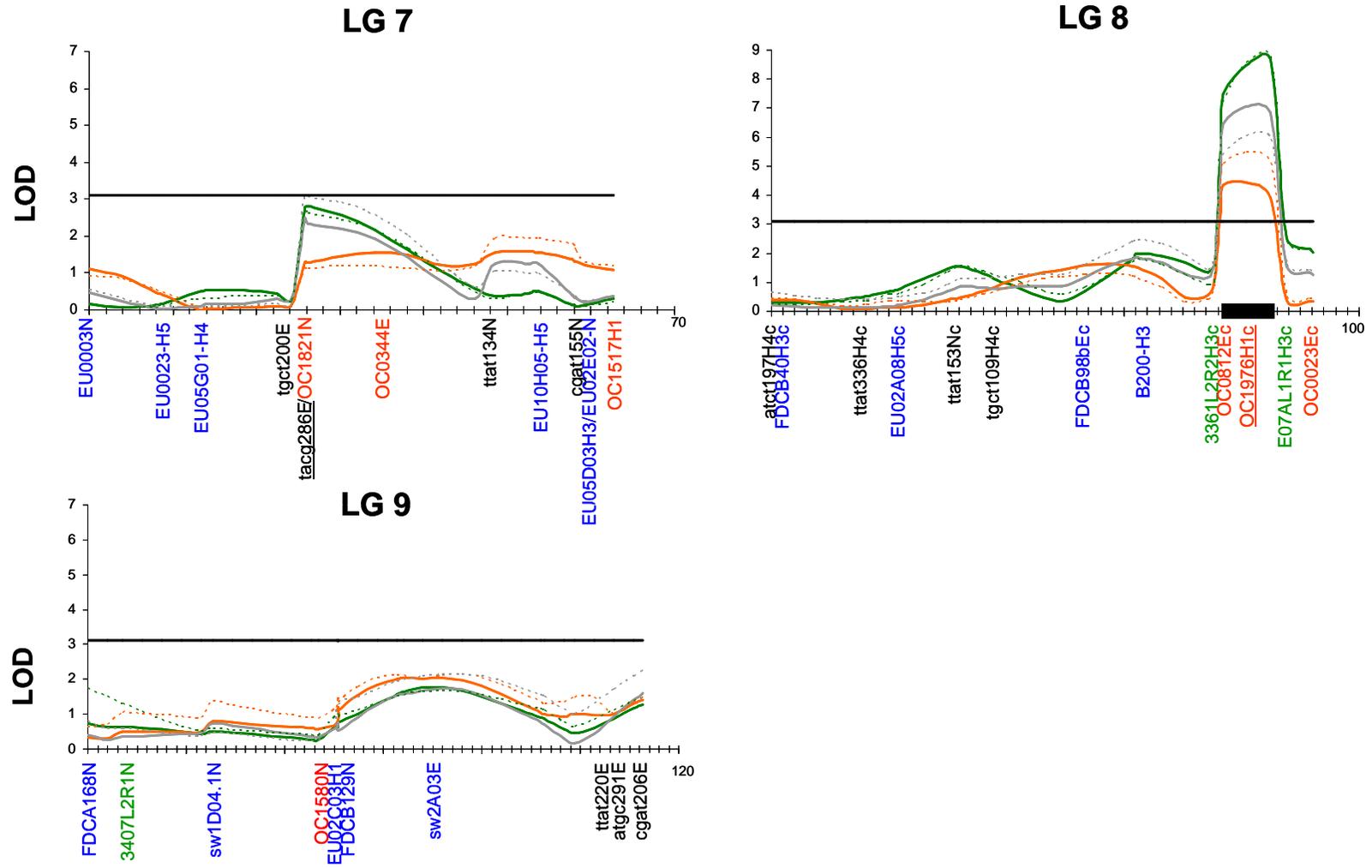
**Tableau 2** : Effet allélique pour le marqueur FDCA149 lié significativement au QTL PL3 dans les sous-descendances F1'a et F1'b.

F1'a				F1'b			
LG3 FDCA149	II	Im	Im-II	II	Im	Im-II	
PL50	1.35	1.61	0.26	1.16	2.04	0.88	
PL90	1.11	1.39	0.28	0.89	2.09	1.20	
SH50	1.20	1.50	0.30	1.80	2.10	0.30	
SH90	1.00	1.30	0.30	1.51	2.06	0.55	
TOT50	1.57	1.87	0.30	1.88	2.17	0.29	
TOT90	1.32	1.63	0.31	1.76	2.12	0.36	

**Figure 4:** Courbes LOD score obtenues après analyse MQM des données plantules (PL ; en orange), structures chlorophylliennes unipolaires (SH ; en vert) et événements totaux (ET ; en gris) dans la descendance F1'a+b. L'analyse a été réalisée pour chaque paramètre d'analyse ( $_{90}$ , intensité moyenne de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait en pointillé) et  $_{50}$ , intensité maximale de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait plein). Les régions liées aux QTL PL, SH et ET sont indiquées par un cadre noir. Les co-facteurs utilisés lors de l'analyse MQM sont soulignés.



Analyse QTL



## 2. Détections de QTL liées à la formation de cals et de bourgeons à partir d'explants foliaires

Dans la sous-descendance F1'b, la mise en culture des pétioles en condition d'induction d'organogenèse (Chapitre I) a été réalisée en parallèle avec l'induction des racines en condition d'embryogenèse somatique. Les pétioles ont été placés sur un milieu favorable à l'organogenèse, favorisant le développement des cals, puis après 10 jours de culture, les cals ont été transférés sur un milieu Heller permettant la régénération des bourgeons.

Pour chacun des génotypes étudiés, le nombre de cals et de bourgeons a été comptabilisé. La fréquence de distribution (Chapitre I) des données comptabilisées révèle que la capacité liée à la callogenèse et à la régénération des bourgeons *in vitro* est lié à un caractère quantitatif héritable.

L'analyse IM (Annexes 10 et 15) révèle un seul QTL pour les caractères liés à la callogenèse et à la régénération des bourgeons, situé sur les groupes de liaisons 8 (Cal8) et 3 (Bg3), respectivement (Tableau 3). Le QTL Cal8 a été retenu même s'il présente une valeur de LOD (3,07) légèrement inférieure au seuil de détection des QTL. Cal8 se situe au centre du groupe de liaison 8 entre 46 et 53 cM, et est lié à un marqueur FDCA98bEc. Le choix de ce marqueur comme co-facteur pour l'analyse MQM, confirme la présence de ce QTL (LOD=3,89, R<sup>2</sup>=18,7%) et permet de détecter un autre QTL (Cal2) situé sur le groupe de liaison 2 avec un LOD score de 2,99 (13,6%) (Tableau 3 et Annexe 16).

La région identifiée sur le groupe de liaison 8 semble également impliquée dans la variation du caractère lié à la formation des bourgeons. Ce QTL (Bg8) a été détecté après sélection du marqueur FDCA149 localisé sur le groupe de liaison 3 comme co-facteur dans l'analyse MQM.

La totalité de la variation phénotypique observée est de 32,3% et 42,1% et pour les caractères liés à la callogenèse et à la régénération des bourgeons, respectivement. Pour le QTL Cal8 et Bg8, l'allèle *m* de K59 semble défavorable à la formation de cals et de bourgeons. Pour le QTL Cal2, c'est l'allèle *g* apporté par K28 favorable à la formation de cals et au contraire c'est l'allèle *b* de K59 qui est favorable à la formation des bourgeons au niveau du QTL Bg3.

Lorsqu'on observe l'allure des courbes des QTL Cal8 et Bg8, sur les graphiques obtenues par analyse MQM (Annexe 16), on peut voir que ces QTL sont très proches et qu'ils couvrent une région génomique commune. Avec notre analyse, il est difficile de dire si ces deux QTL sont identiques ou alors s'ils sont distincts l'un de l'autre. La taille de l'effectif de la sous-descendance F1'b est peut-être un probablement trop faible ( $n=104$ ) pour déterminer de manière précise la position du QTL sur le groupe de liaison 8.

**Tableau 3** : Analyse MQM (Multiple QTL Mapping). Identification des QTL liés à la callogenèse et l'organogenèse dans la sous-descendance F1'b. La position des QTL (cM), le LOD et le pourcentage de variation expliquée ( $R^2$  %) sont indiqués pour chaque QTL.

LG	QTL	POSITION (cM)	LOD	$R^2$ %
2	Cal2	11-23 (19)	2.99	13.6
8	Cal8	38-50 (46)	3.89	18.7
				<b>32.3</b>
3	Bg3	51-64 (60)	4.78	24.2
8	Bg8	32-39 (33)	3.38	17.9
				<b>42.1</b>

### 3. Effets d'interaction

L'analyse QTL a été réalisée à partir des deux sous-descendances F1'a et F1'b et a révélé la présence de QTL identiques mais aussi différents entre les deux sous-descendances. Ces différences pourraient s'expliquer par un effet environnemental (Chapitre I) mais aussi peut-être par des effets d'interaction entre marqueurs. En effet, tous les QTL identifiés n'expliquent pas la totalité de la variation phénotypique observée ce qui pourrait être dû à des effets d'épistasies. La recherche d'interaction entre tous les marqueurs présents sur des groupes de liaisons différents a été entreprise pour les données PL, SH et ET.

Pour les données PL, SH et ET une interaction significative ( $P<0.0001$ ) entre les marqueurs sw4D08H3 (LG1) et OC1976H1 (LG8) uniquement dans la sous-descendance F1'a semble significative. Ces marqueurs sont liés aux régions chromosomiques identifiées précédemment par des analyses IM et MQM. C'est

l'association entre les allèles de K59 qui semble favorable pour augmenter le nombre de plantules et de structures chlorophylliennes unipolaires formées.

Une autre interaction ( $P < 0.0001$ ) a été observée entre les marqueurs présents au niveau du QTL localisé sur le groupe de liaison 1 avec des marqueurs du groupe de liaison 7 dans la sous-descendance F1'a. L'interaction a été observée entre les marqueurs EU09B08N, EU03H07N (LG1) et tacg200E, tacg286E (LG7). C'est l'association entre les allèles issus du même parent (K28) qui est à l'origine d'une augmentation du nombre d'évènements formés (PL et SH).

Cette interaction présente entre le groupe de liaison 1 et 7 dans la sous-descendance F1'a n'a pas été retrouvée dans la sous-descendance F1'b. Dans le chapitre précédent (Chapitre 2), une forte distorsion de ségrégation pour certains marqueurs assignés au groupe de liaison 7 a été observée. Les marqueurs tacg286E et tacg200E (LG7) ont une classe génotypique (//) diminuée dans la sous-descendance F1'b. Nous pouvons émettre l'hypothèse, que le faible effectif observé dans cette classe est lié avec la perte d'interaction entre les deux régions du groupe de liaison 1 et 7 et par conséquent, la perte du QTL dans la région du groupe de liaison 1. Cette hypothèse n'a pas été confirmée dans le cadre de ma thèse.

### III. Discussion

Lors de cette étude, nous avons détecté plusieurs QTL robustes dispersés sur le génome de la chicorée, associés aux données plantules (PL) et structures chlorophylliennes (SH), confirmant la complexité du caractère lié à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires. Trois QTL sur les groupes de liaison 2, 6 et 8 ont été retrouvés quelle que soit la sous-descendance analysée mais aussi lors de l'analyse des données de la descendance F1'a+b. Au contraire, les QTL SH1, PL1, SH4 et PL4 ont été identifiés dans la sous-descendance F1'a exclusivement expliquant en moyenne 20% de variation phénotypique totale. Quand aux QTL PL3, SH3, ils ont été détectés dans la sous-descendance F1'b et dans la descendance F1'a+b.

Lors de cette analyse, les QTL expliquent en moyenne entre 8,1 et 15% (F1'a) et entre 8,3 et 16,5% (F1'b) de variation phénotypique. La taille des 2 sous-descendances analysées (F1'a,  $n=123$  et F1'b,  $n=104$ ), ne nous permet pas

d'affirmer que les QTL à effets majeurs ou modérés mais surtout que ceux à effets mineurs ont été détectés par l'approche utilisée. On peut citer, par exemple, le QTL identifié au niveau du groupe de liaison 3 dans la sous-descendance F1'b et la descendance F1'a+b non détecté dans la sous-descendance F1'a.

Les analyses que nous avons réalisées ne permettent pas de comprendre pourquoi le nombre et la localisation génomique des QTL différent d'une sous-descendance à l'autre. La détection de QTL différent à partir d'une même descendance a également été observée chez le maïs (Krakowsky *et al.*, 2006). Chez cette espèce, une analyse QTL a été réalisée pour identifier les régions génomiques impliquées dans l'initiation des cals et la totipotence cellulaire. Cette étude menée à partir d'une descendance de 126 individus (RILS), a été réalisée à partir des données obtenues en 1997 puis en 1998. Six et 3 QTL ont été détectés pour les données relevées en 1997 et 1998, respectivement, et un seul QTL identifié est commun aux 2 années. Les auteurs ont moyenné les valeurs phénotypiques obtenues en 1997 et 1998 pour chacun des individus, et ils ont réalisé une nouvelle analyse QTL. L'ensemble des QTL détectés avec les valeurs phénotypiques en 1997 et 1998 ont été identifiés sauf un QTL identifié avec les données de 1997. Des effets d'interaction différents ont été identifiés avec les données obtenues en 1997, 1998 et la moyenne des 2 années.

Notre analyse QTL révèle une co-localisation de plusieurs QTL liée à la formation des plantules et des structures chlorophylliennes unipolaires dans un nombre limité de régions chromosomiques. En effet, les régions QTL spécifiques du caractère embryogène coïncident avec les régions liées aux QTL pour la formation des structures chlorophylliennes sauf pour le groupe de liaison 4. La co-localisation des QTL indique certainement l'existence d'une base génétique commune pour les caractères étudiés, résultant d'effets pléiotrope ou de locus étroitement liés. Ce résultat mais aussi la corrélation entre la formation des plantules et des structures chlorophylliennes unipolaires présentée dans le chapitre I, pourraient suggérer que ces deux caractères ont une origine commune ou alors que les structures chlorophylliennes unipolaires seraient issues d'un développement inachevé des embryons somatiques. Ces suggestions seront évoquées plus en détail dans la discussion générale (Chapitre V).

La part de variation phénotypique expliquée pour la capacité à former des plantules pour chacun des QTL détectés est proche de celle observée dans la littérature. Dans les deux sous-descendances F1'a et F1'b, le LOD des QTL détectés est compris entre 3,6 et 6,9 et, la variation phénotypique observée est de 8,2 à 15,1%. Des analyses QTL liées à l'ES réalisées chez d'autres espèces, le plus souvent à partir de lignées recombinantes (RILS), ont permis d'identifier de 1 à 7 QTL expliquant de 6,2 à 41,6 % de variation phénotypique observée (Tableau 2, Analyse bibliographique). Chez le maïs, 4 QTL expliquant entre 6,8 et 14,3% de variation phénotypique ont été identifiés après la mise en culture des embryons immatures (Armstrong *et al.*, 1992). Chez le tournesol (Florres-Berrios *et al.*, 2000b), 7 régions ( $R^2$  compris entre 7 et 15%) génomiques liées à la formation des embryons ont été identifiées expliquant au total 89%. Plus récemment, 4 QTL liés à la formation des embryons obtenus à partir de cals chez le blé (Haiyan *et al.*, 2007) expliquant chacun entre 11 et 22, 8% de variation ont été détectés. De même, le pourcentage total de variation phénotypique expliquée (supérieur à 45%) pour les QTL liés aux événements totaux est comparable à celui observé dans la littérature.

La part de variation phénotypique expliquée par 4 QTL liée à l'ES chez le seigle est une exception (Bolibok *et al.*, 2007). En effet, les QTL expliquent chacun une part de variation phénotypique très élevée de 21,5 à 41,6% par rapport à celle observée chez d'autres espèces.

L'analyse QTL a également été réalisée à partir des données de callogenèse et d'organogenèse dans la sous-descendance F1'b afin d'identifier les régions impliquées dans ces caractères. Au total, nous avons identifié 2 QTL pour chacun de ces caractères (Cal 2, Cal8 (32,3%) et Bg3, Bg8 (42,1%)).

Plusieurs auteurs ont également recherché les locus impliqués dans l'induction des cals puisque dans la plupart des cas la présence de cals est un bon indicateur pour le rendement final de plante régénérée (He *et al.*, 1998 ; Murigneux *et al.*, 1994). Dans notre analyse, une corrélation faible ( $r=0,54$ , Chapitre 1) entre la formation des cals et la formation des bourgeons a été observée. Ceci pourrait expliquer qu'aucun QTL commun entre les deux caractères n'a été détecté. Chez *Arabidopsis*, aucune corrélation entre le pourcentage de cals formés et la capacité de régénération n'a été identifiée dans les explants foliaires (Schiantarelli *et al.*,

2001). Dans cette étude, c'est la morphologie des cals qui semble être liée à la capacité de régénération des explants.

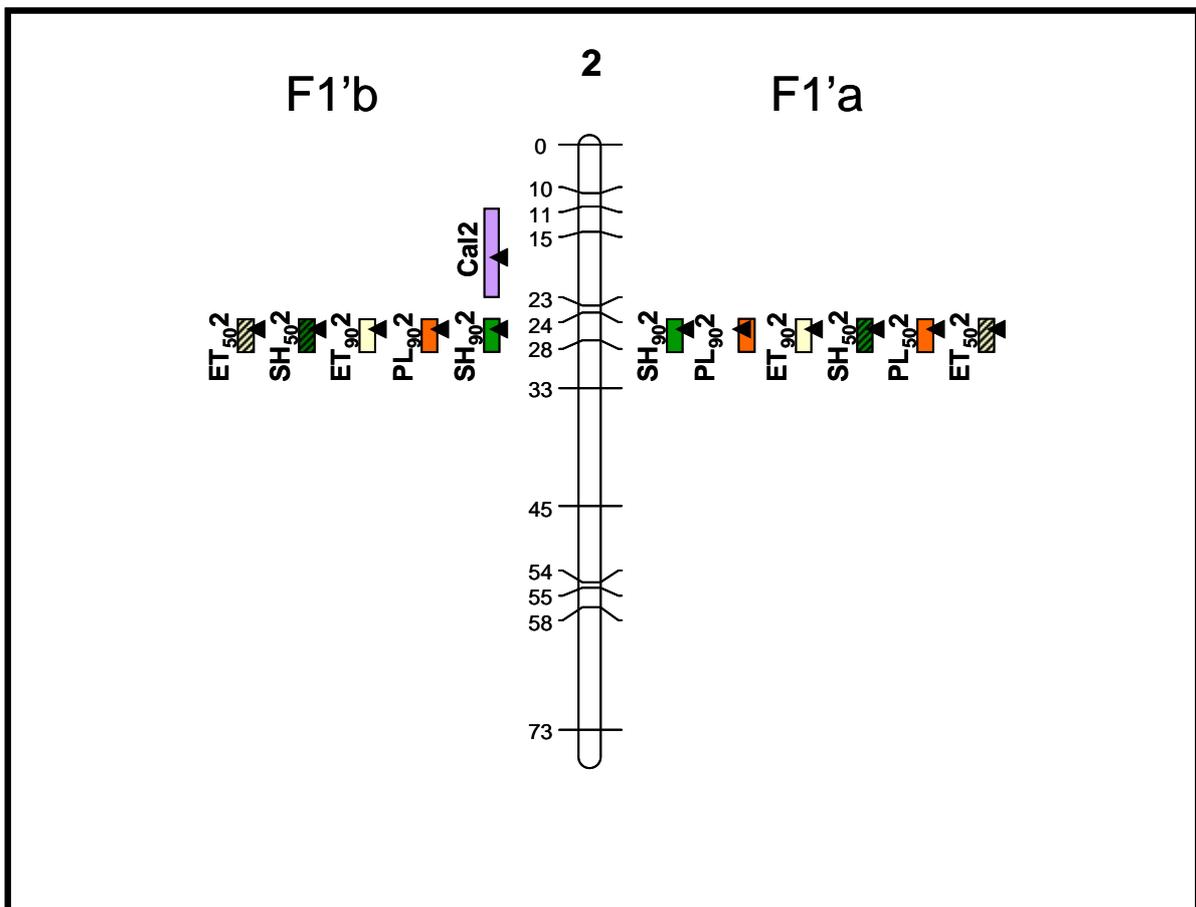
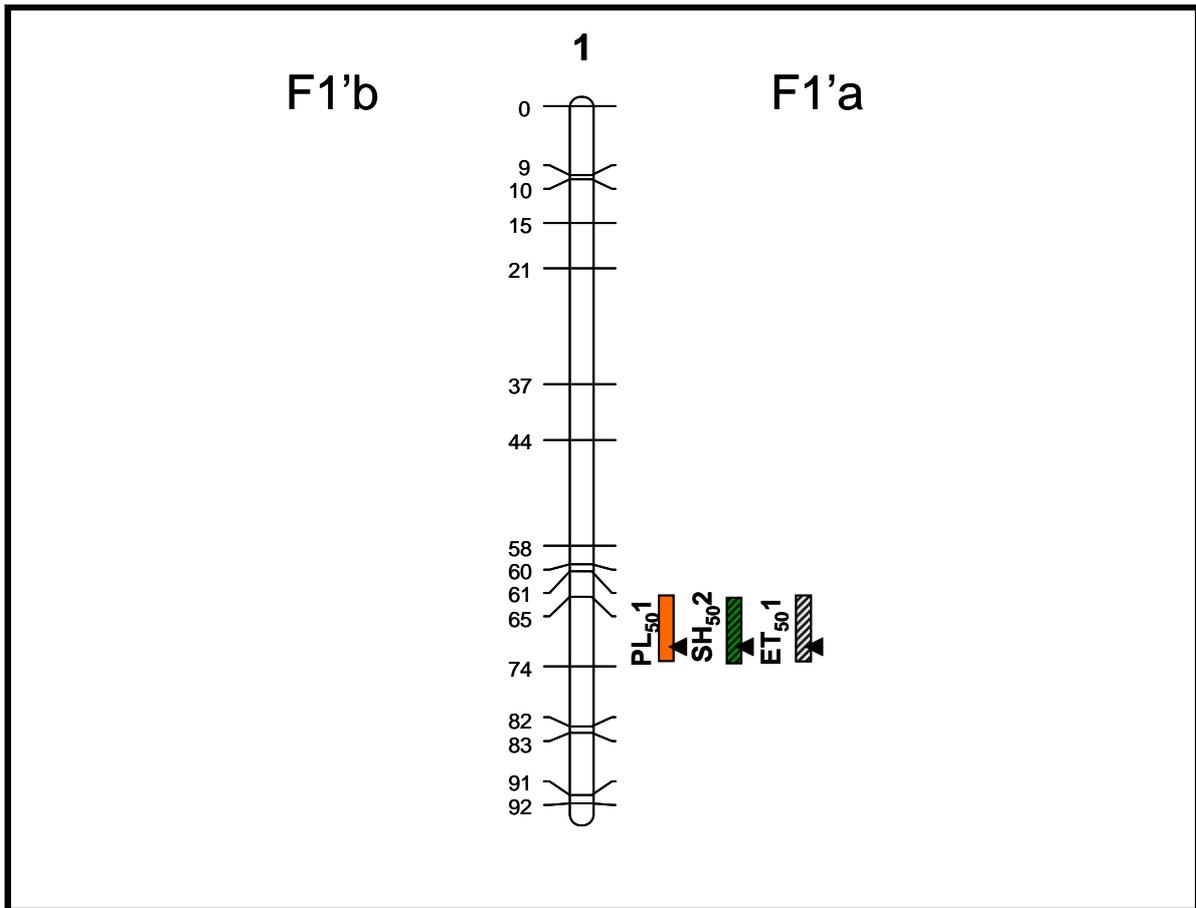
Le nombre de bourgeons régénérés à partir de cals a fait l'objet d'une analyse QTL chez le riz (Taguchi-Shiobara *et al.*, 1997), l'orge (Ben Amer *et al.*, 1997), *Arabidopsis* (Schiantarelli *et al.*, 2001), le brocoli (Holme *et al.*, 2004) et le blé (Haiyan *et al.*, 2007). L'analyse de ce caractère a permis de détecter entre 2 et 5 QTL. Le pourcentage total de variation phénotypique explique entre 28,8 et 38,5% chez le brocoli et le riz, respectivement.

Les QTL Cal et Bg sur les groupes de liaison 2, 3 et 8 diffèrent de ceux identifiés pour les caractères liés à la formation des plantules et des structures chlorophylliennes unipolaires. Les locus impliqués dans la formation des plantules (PL) et dans la formation des structures chlorophylliennes unipolaires (SH) ne semblent pas être identiques à ceux impliqués dans la callogenèse (Cal) et l'organogenèse (Bg). Ces résultats pourraient nous éclairer quand à l'origine des structures chlorophylliennes unipolaires. En effet, dans le chapitre I, nous nous interrogeons sur l'identité de ces structures. La détection de QTL identiques entre les caractères liés à PL et SH, et l'absence de co-localisation entre les QTL SH, Cal et Bg indique peut-être que les structures chlorophylliennes unipolaires seraient plutôt issues d'une voie de développement embryogène. Toutefois, pour confirmer cette hypothèse, il faudrait réaliser une recherche de QTL après analyse du potentiel organogène des explants racinaires puisque dans notre analyse celle-ci a été réalisée à partir de fragments de pétioles. Chez *Arabidopsis*, des QTL différents ont été identifiés à partir d'explants racinaires et foliaires pour le même caractère étudié (Schiantarelli *et al.*, 2001).

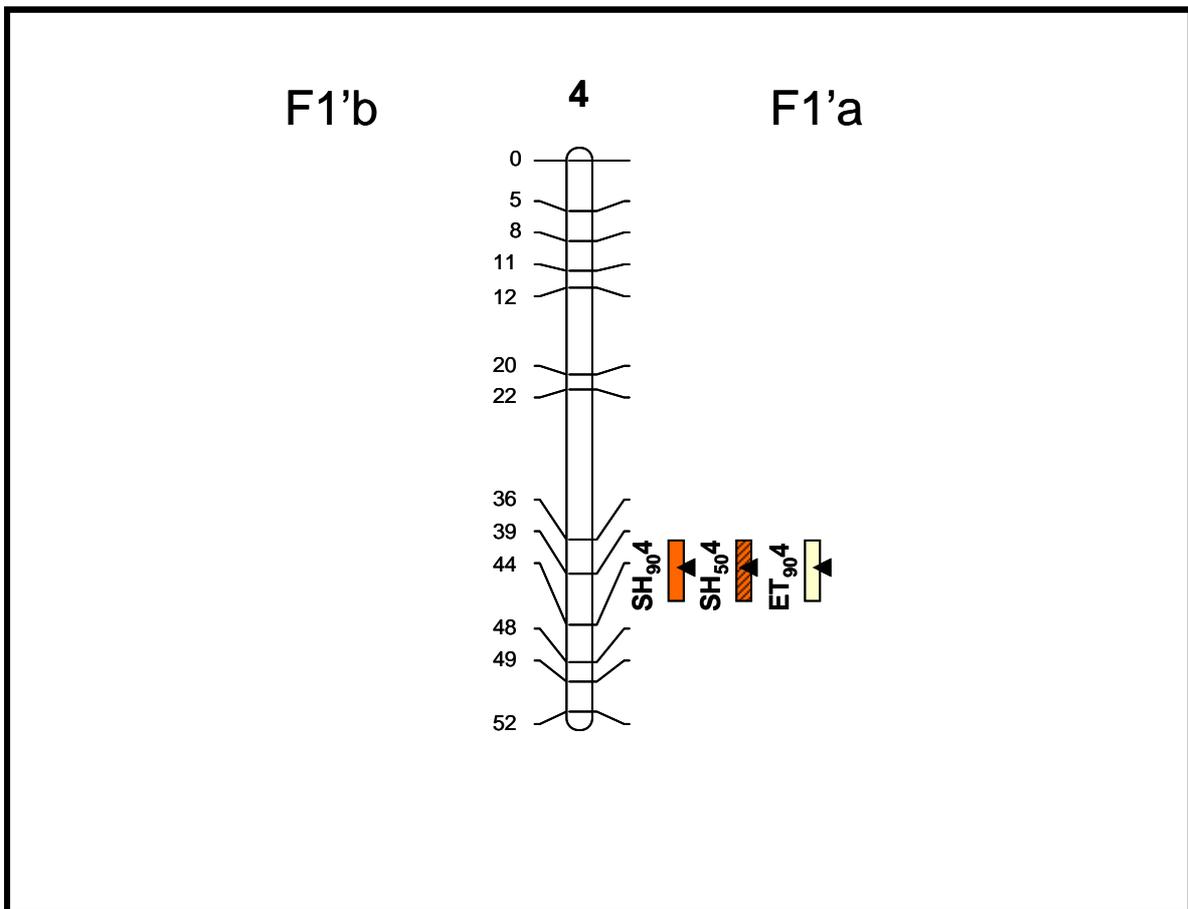
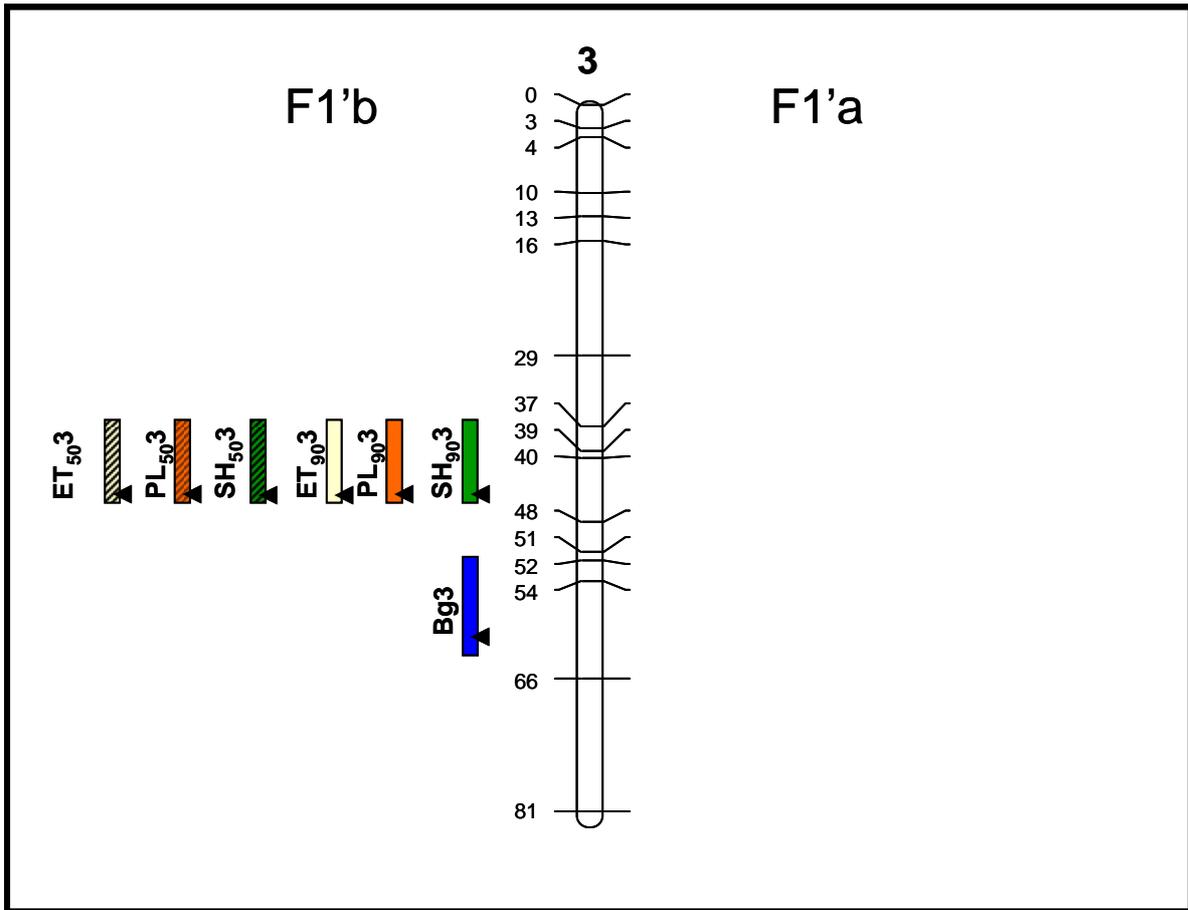
Dans cette étude, nous avons identifié plusieurs régions chromosomiques distribuées sur le génome et impliquées dans la régénération *in vitro* (Figure 5). Cette première étude sur les caractères liés à l'embryogenèse somatique et l'organogenèse chez la chicorée, représente une étape importante pour élucider les mécanismes qui gouvernent ces deux caractères. De plus, les marqueurs associés aux QTL sont des outils pour la sélection assistée par marqueur afin de générer des individus plus embryogènes ou organogènes.

**Figure 5 :** Bilan des QTL détectés dans les sous-descendances F1'a (à droite) et F1'b (à gauche). Les QTL ont été désignés par PL (Plantules), SH (structures chlorophylliennes unipolaires) et ET (événements totaux) suivis par le paramètre d'analyse utilisé ( $\mu_{90}$ , intensité moyenne de réponse liée à l'embryogenèse somatique et à la formation de structures chlorophylliennes unipolaires) et  $\mu_{50}$ , intensité maximale de réponse liée à l'embryogenèse somatique et à la formation de structures chlorophylliennes unipolaires) et par le numéro du groupe de liaison correspondant. Les QTL liés à la callogenèse (Cal) et l'organogenèse (Bg) dans la sous-descendance F1'b sont suivis par le numéro du groupe de liaison correspondant.

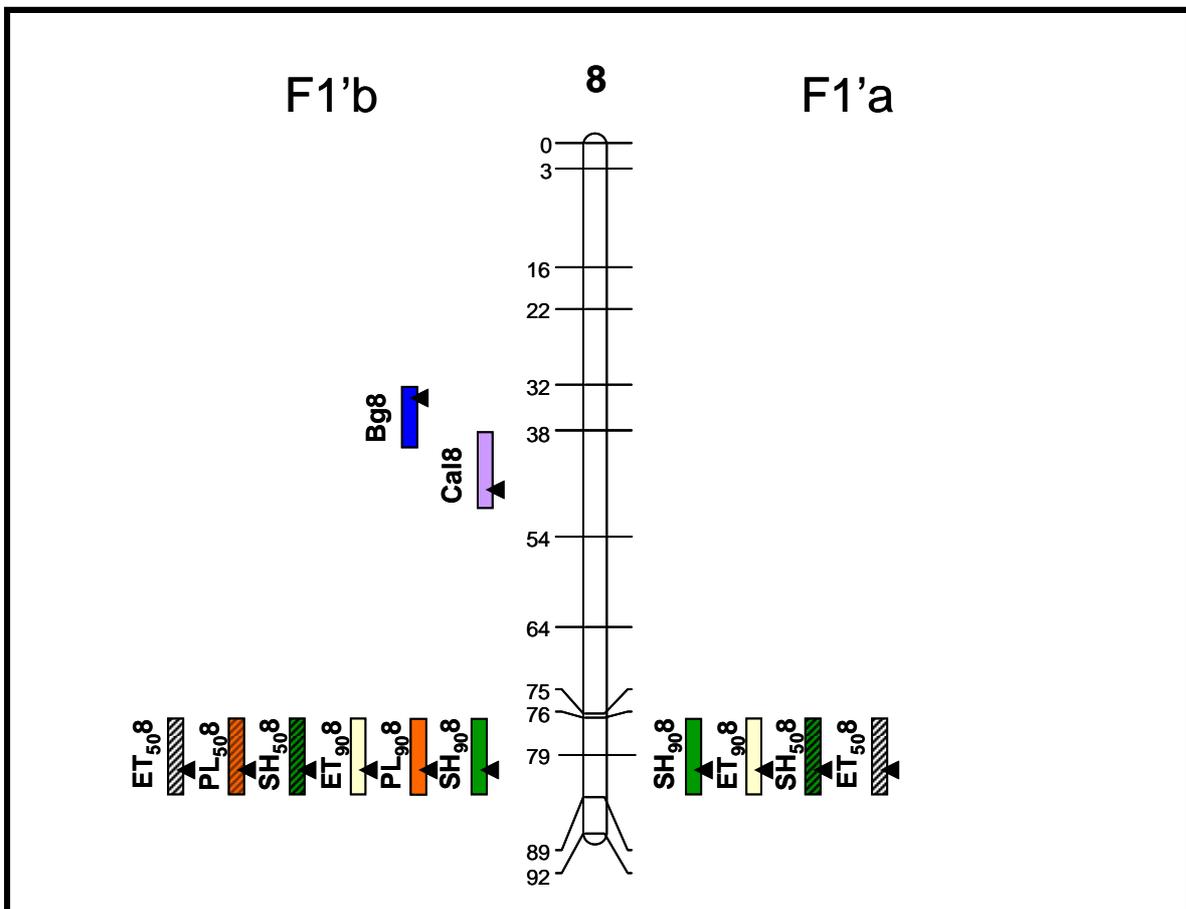
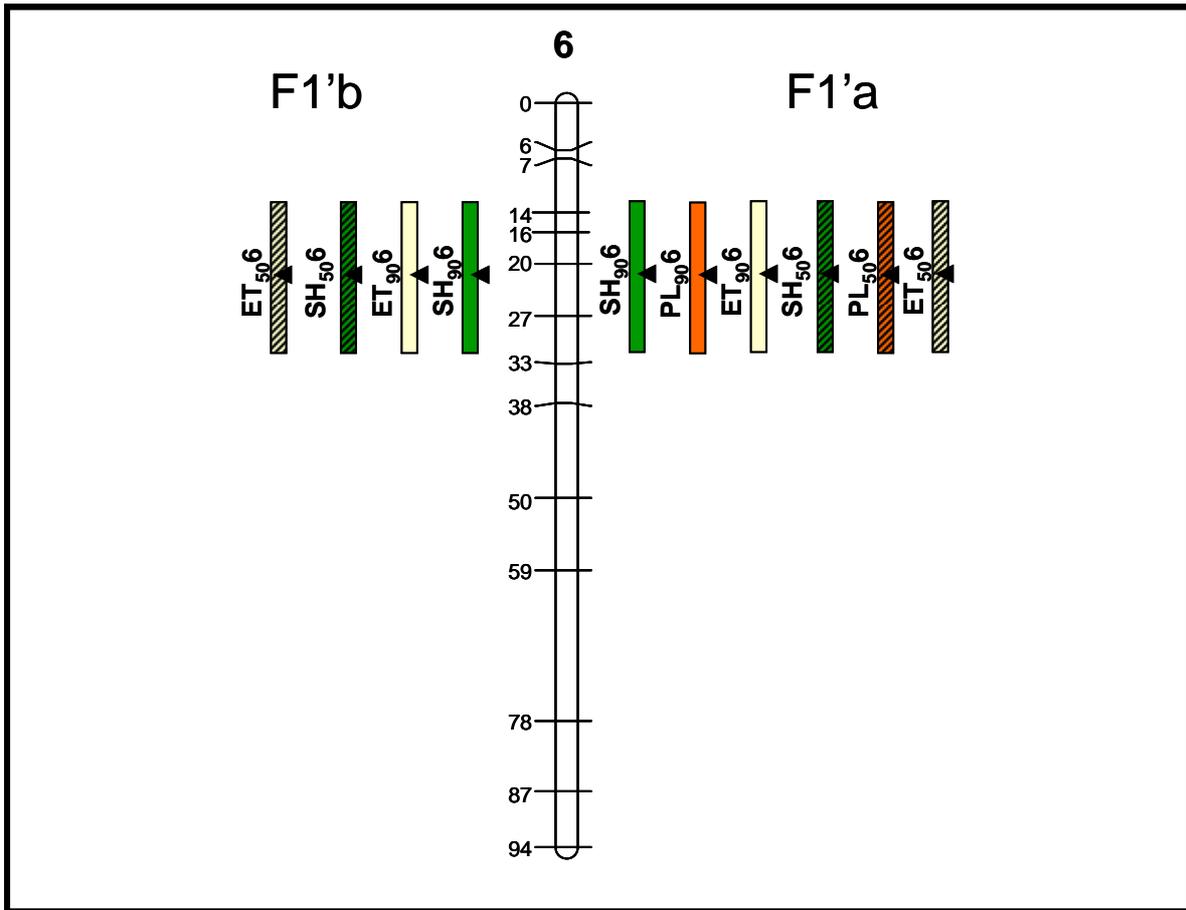
Analyse QTL



Analyse QTL



Analyse QTL



## Chapitre IV

### Des gènes candidats qui co-localisent avec les QTL

### I. Introduction

Le but ultime dans la détection des QTL est souvent d'identifier les gènes spécifiques encore appelés gènes candidats responsables de la variation phénotypique. L'hypothèse de travail assume que le polymorphisme moléculaire du gène candidat est lié à la variation phénotypique. L'approche gène candidat a été utilisée avec succès en génétique humaine et animale (Rothschild and Soller, 1997) et depuis 1990, chez les plantes (Byrne and McMullen 1996). Plusieurs caractères de l'agriculture sont hautement héréditaires et quantitatifs, lesquels résultent souvent de l'influence de gènes multiples. Parce que la multiplicité des gènes définit un caractère complexe, leur effet partiel sur la variation phénotypique et leur localisation imprécise sur la carte génétique, l'approche gène candidat est une alternative à la caractérisation des QTL par clonage positionnel ou la mutagenèse.

### II. Etude des gènes à fonction connue liés aux QTL détectés

Lorsque les QTLs ont été identifiés, nous avons recherché la présence de gènes à fonction connue dans les régions significatives liées à ces QTL (Tableau 1). Nous avons essayé d'établir une relation entre l'expression différentielle des gènes candidats à jour 4 de l'ES et la formation des plantules et/ou structures chlorophylliennes. Mais le jour 4 de l'ES (début des 1<sup>ères</sup> divisions des cellules embryogènes) est caractéristique des phases précoces du processus de l'ES alors que nos observations phénotypiques ont été réalisées au stade plantule, stade tardif du développement des embryons somatiques. Certains des QTL détectés au cours de nos analyses sont peut-être plus spécifiques des stades tardifs de l'ES. Ceci pourrait expliquer que certains QTL détectés co-localisent plus particulièrement avec des gènes exprimés tardivement dans l'ES.

Parmi les 63 gènes, cartographiés sur la carte génétique K28\*K59 et/ou F1'a, potentiellement liés au processus de l'embryogenèse somatique, 16 co-localisent avec les QTL (de 2 à 6 gènes par QTL détecté). Ces gènes candidats codent pour des protéines ribosomales (RPS10, RPS6, RPL10, RPL10a, RPL17), une protéine de tolérance au sel (STP), des protéines impliquées dans le développement du méristème caulinaire (STM et AGO1), des protéines de transfert lipidique (LTP, LTP1

Gènes candidats

et LTP2), une  $\beta$ -tubuline, une protéine arabinogalactane (AGP) et enfin des enzymes (HMG-CoA réductase et FAD). Sur ces 16 gènes, 11 ont été révélés par analyse microarray montrant une expression différentielle au 4<sup>ème</sup> jour d'induction de l'embryogenèse somatique (Legrand, 2006).

**Tableau 1** : Gènes liés à l'embryogenèse somatique et cartographiés sur la carte génétique F1'a et ou K28\*K59.

En gras sont indiqués les marqueurs déjà cartographiés sur la carte F1'a. En italique les séquences EST différentiellement exprimées chez le génotype embryogène K59 (E) ou non embryogène C15 (NE), révélées par analyse microarray et/ou RT-PCR au 4<sup>ème</sup> jour d'induction de l'embryogenèse somatique (Legrand, 2006 ; Legrand *et al.*, 2007) .

Gènes cartographiés	Code	Nature de la protéine traduite	Expression différentielle	Groupe de liaison	QTL
<b>CLO22L1R1*</b>					
<b>GST5'-RACE4</b>		<b>glutathione-S-transferase</b>		<b>1</b>	
<i>OC0081</i>		putative chloroplast outer membrane protein [ <i>Oryza</i> ]	NE	1	
<i>OC0185</i>		putative translation factor [ <i>Pinus pinaster</i> ]	NE	1	
<i>OC0311</i>		acidic ribosomal protein P0 [ <i>G. max</i> ]	E	1	
<i>OC0378_a</i>		60S ribosomal protein L5 [ <i>Solanum melongena</i> ]	E	1	
<i>OC0581</i>		S6 ribosomal protein [ <i>Asparagus officinalis</i> ]	E	1	
<i>OC1104</i>		zinc finger protein [ <i>Pinus radiata</i> ]	NE	1	
<i>OC1394</i>	STP	salt tolerance protein 5 [ <i>Beta vulgaris</i> ]	E	1	QTL1
<i>OC1531_a</i>	RPS10	putative ribosomal protein S10 [ <i>Arabidopsis thaliana</i> ]	E	1	QTL1
<i>OC0686</i>		unknown [ <i>Arabidopsis thaliana</i> ]		2	
<i>OC0082_b</i>	RPS6	60s acidic ribosomal protein [ <i>Hyacinthus orientalis</i> ]	E	2	QTL2
<i>OC1359_a</i>	RPL17	60S ribosomal protein L17 [ <i>Oryza sativa (japonica cultivar-group)</i> ]	E	2	QTL2
<i>OC1528</i>		NADH-glutamate dehydrogenase [ <i>Nicotiana plumbaginifolia</i> ]	NE	2	QTL2
<i>OC2006</i>	AGO1	putative leaf development protein Argonaute [ <i>Arabidopsis thaliana</i> ]	E	2	QTL2
<i>OC2141</i>		TSB2 tryptophan synthase beta-subunit [ <i>Arabidopsis thaliana</i> ]	NE	2	
<b>ESTE002C9</b>		<b>haloacid dehalogenase-like hydrolase family protein [<i>Oryza sativa (japonica cultivar-group)</i>]</b>		<b>3</b>	
<b>ESTE002B3</b>		<b>GTP binding protein [<i>Gossypium hirsutum</i>]</b>		<b>3</b>	
<i>OC0026_a</i>		glyceraldehyde 3-phosphate dehydrogenase [ <i>Solanum tuberosum</i> ]	E	3	
<i>OC0206</i>	HMG-CoA	hydroxy-methylglutaryl-coenzyme A reductase [ <i>Nicotiana tabacum</i> ]	NE	3	QTL3
<i>OC1138</i>	HMG-CoA	3-hydroxy-3-methylglutaryl-coenzyme A reductase [ <i>Hevea brasiliensis</i> ]	NE	3	QTL3
<i>OC0347b</i>		poly(ADP-ribose) polymerase [ <i>Zea mays</i> ]	E	3	
<i>OC1952</i>		NADPH-ferrihemoprotein reductase [ <i>Helianthus tuberosus</i> ]	NE	3	
<i>OC2147</i>		putative glycerol-3-phosphate dehydrogenase [ <i>Arabidopsis thaliana</i> ]	NE	3	
<i>OC0024</i>	MADS	<b>MADS box protein (AGL24) [<i>Cichorium intybus</i>]</b>	NE	<b>4</b>	
<b>ESTE002B7</b>		<b>amino acid permease 6 [<i>Arabidopsis thaliana</i>]</b>		<b>4</b>	
<i>OC0390</i>		40S ribosomal protein S2 [ <i>Arabidopsis thaliana</i> ]	E	4	

## Gènes candidats

OC1412		unknown [ <i>Arabidopsis thaliana</i> ]	NE	4	
OC1796_b		vegetative storage protein, VSP [ <i>Cichorium intybus</i> ]	NE	4	
OC1833_b		ribosomal protein L3A [ <i>Nicotiana tabacum</i> ]	E	4	
<b>SK95UTR9L19R6</b>		<b>sérine protéine kinase SERK</b>		<b>5</b>	
OC0839		dormancy-associated protein [ <i>Codonopsis lanceolata</i> ]	NE	5	
OC1635_a		lipid transfer protein (LTP) family protein [ <i>Arabidopsis thaliana</i> ]	NE	5	
<b>GTPB L3R1</b>		<b>GTP binding protein (GTP1)</b>		<b>6</b>	
<b>3262L1R1</b>	Dehydrine	<b>dehydrine</b>		<b>6</b>	QTL6
CiSTM	STM	shootmeristemless		6	QTL6
OC0687	AGP	arabinogalactan protein [ <i>D. carota</i> ]	E	6	QTL6
OC0344		unknow protein	NE	7	
OC1145		delta-12 fatty acid desaturase [ <i>Borago officinalis</i> ]	NE	7	
OC1517		putative trypanothione-dependent peroxidase [ <i>Oryza sativa</i> ]	NE	7	
OC1696		P protein [ <i>Flaveria pringlei</i> ]	NE	7	
OC1821		40S ribosomal protein S11 [ <i>Euphorbia esula</i> ]	E	7	
OC1932		hybrid proline-rich protein [ <i>Catharanthus roseus</i> ]	NE	7	
<b>3142L2R1*</b>		<b>hydroxy-methylglutaryl-coenzyme A syntase</b>		<b>8</b>	
<b>3704L2R1</b>	LTP2	<b>lipid Transfer Protein 2</b>		<b>8</b>	QTL8
<b>3361L1R1</b>	LTP1	<b>lipid Transfer Protein 1</b>		<b>8</b>	QTL8
<b>Tubulin L2R2</b>	B-tubulin	<b>B-tubulin</b>		<b>8</b>	QTL8
OC1822_b	RPL10a	putative ribosomal protein L10a [ <i>Oryza sativa (japonica cultivar-group)</i> ]	E	8	QTL8
OC1976	FAD	delta 12 fatty acid desaturase [ <i>Crepis palaestina</i> ]	NE	8	QTL8
OC2060	LTP	lipid transfer protein [ <i>Helianthus annuus</i> ]	NE	8	QTL8
OC0023_a		gluthatione transferase [ <i>H. muticus</i> ]	NE	8	
OC0031		peroxidase 3 [ <i>Scutellaria baicalensis</i> ]	NE	8	
OC0347a		poly(ADP-ribose) polymerase [ <i>Zea mays</i> ]	E	8	
OC0565_a		acetoacetyl-CoA thiolase [ <i>Arabidopsis thaliana</i> ]	NE	8	
OC0748		ribosomal protein L7 [ <i>Solanum tuberosum</i> ]	E	8	
OC0812		RNA Binding Protein 47 [ <i>Nicotiana plumbaginifolia</i> ]	NE	8	
OC0818		glutamate synthase (ferredoxin) [ <i>Nicotiana tabacum</i> ]	NE	8	
<b>GTPB L2R2</b>		<b>GTP binding protein (GTP1)</b>		<b>9</b>	
<b>3474 L1R1</b>		<b>hemoglobine 1</b>		<b>9</b>	
OC0183		salt tolerance protein 5 [ <i>Beta vulgaris</i> ]	E	9	
OC0603_a		cobalamine-independent methionine synthase [ <i>S. scutellarioides</i> ]	NE	9	
OC1580		40S ribosomal protein S6 [ <i>Cicer arietinum</i> ]	E	9	
OC1929		G protein beta subunit [ <i>N. plumbaginifolia</i> ]	E	9	

## 1. Protéines impliquées dans la biogenèse des ribosomes

Sur les 113 gènes exprimés préférentiellement chez K59 (45 gènes) et C15 (68 gènes) au 4<sup>ième</sup> jour de culture en condition d'ES, 18 codent des protéines ribosomales. L'analyse transcriptionnelle a révélé que ces gènes montrent globalement une hausse d'expression chez le génotype embryogène (E) au cours de la cinétique alors qu'aucune modulation, ou une baisse d'expression, n'a été observée chez le génotype C15 (NE) (Legrand, 2006).

Les ribosomes synthétisés dans le cytoplasme sont des éléments indispensables à la synthèse des protéines, leur activité et leur nombre varient en fonction de l'intensité de la synthèse des protéines dans la cellule.

La forte proportion de gènes codant des protéines ribosomales, soit 40% des gènes préférentiellement exprimés chez K59, parmi les 113 gènes analysés, reflète la synthèse des ribosomes dans les cellules se préparant à entrer dans une voie de transdifférenciation et à réinitier les divisions cellulaires (Legrand, 2006).

La relation entre une augmentation de l'expression de gènes codant des protéines ribosomales et la division cellulaire a par ailleurs été montrée dans diverses études (Barakat *et al.*, 2001 ; Dresselhaus *et al.*, 1999). Chez le zinnia, (*Zinnia elegans*) plusieurs gènes codant des protéines ribosomales ont montré une expression différentielle au cours de la transdifférenciation de cellules de mésophylle en cellules du xylème (Demura *et al.*, 2002). Chez la luzerne (*Medicago trunculata*), une augmentation relative de l'expression des protéines ribosomales a également été observée et pourrait être liée à la division cellulaire et à la différenciation des embryons somatiques (Domoki *et al.*, 2006).

Parmi les 18 gènes codant les protéines ribosomales, 12 sont polymorphes et 4 co-localisent avec les QTL des groupes de liaison 1, 2 et 8. Cependant, la probabilité de co-localisation d'un QTL avec des gènes ribosomaux est élevée puisque de nombreux gènes codent pour ce type de protéine. Par exemple, chez *Arabidopsis*, il existe 249 gènes qui codent pour 80 protéines ribosomales (32 pour la petite sous-unité, 48 pour la grande sous-unité) et leurs locus sont dispersés sur les 5 chromosomes constituant le génome d'*Arabidopsis* (Barakat *et al.*, 2001).

De plus, ces protéines ribosomales semblent participer à de nombreux processus de développement comme la division cellulaire et la transdifférenciation qui ne sont pas exclusivement liées à l'embryogenèse somatique.

Ainsi, même si ces protéines ribosomales semblent participer à la formation des embryons somatiques, elles ne semblent pas jouer un rôle majeur et déterminant lors du processus lié à l'ES. Notre étude avait pour objectif d'identifier les gènes majeurs liés à l'ES et si nous hiérarchisons les gènes impliqués dans l'ES, les gènes codant des protéines ribosomales ne semblent pas faire partie de ces gènes majeurs.

### 2. Les gènes en lien avec les divisions cellulaires

Parmi l'ensemble des EST identifiés au jour 4 de culture, c'est-à-dire à l'instant où sont initiées les premières divisions cellulaires, il n'a pas été surprenant d'identifier des gènes liés à la division cellulaire. Au niveau du groupe de liaison 1, un gène code une protéine STP (Salt Tolerance Protein) ayant une forte homologie avec une protéine identifiée chez *Beta vulgaris* impliquée dans le déplacement du noyau. Le gène identifié présente une forte hausse d'expression significative à jour 4 de culture chez le génotype K59 et ne montre pas de modulation d'expression significative chez le génotype C15.

Un autre gène codant pour une arabinogalactane protéine (AGP) présente le même profil d'expression c'est à dire seulement une hausse d'expression chez le génotype embryogène K59 au 4<sup>ième</sup> jour d'induction de l'embryogenèse somatique. Ce gène présente des homologies avec le gène DcAGP1 de carotte dont l'ADNc a été isolé dans des suspensions cellulaires. Le gène DcAGP1 code une AGP dite non classique qui présente de forte similitude avec une famille de protéines riches en proline (Baldwin *et al.*, 2001). Les AGP participent à des interactions moléculaires et ont un rôle dans de nombreuses voies de signalisation (Schultz *et al.*, 1998 ; Serpe, 1999 ; Showalter, 2001). Les AGP sont connues pour être essentielles à la formation d'embryons somatiques (Kreuger and Van Holst, 1996; Toonen *et al.*, 1997 ; Egertsdotter and von Arnold, 1995). En effet, l'addition dans le milieu de culture de réactif  $\beta$ -D-glycosyl Yariv, réactif qui interagit avec les AGP, bloque la formation du développement des embryons somatiques mais pas le processus de dédifférenciation et de réactivation cellulaire (Chapman *et al.*, 2000). Après l'ajout du réactif Yariv des cellules dédifférenciées et réactivées ont pu être observées. Les AGP semblent intervenir plus tardivement dans le processus de développement des embryons somatiques.

Ensuite, des travaux plus récents au sein de notre laboratoire ont révélé une corrélation entre le nombre d'embryons somatiques et la quantité d'AGP libérée dans le milieu de culture (Windels, 2005). Cependant, ces travaux ont porté en particulier sur l'étude d'AGP classiques alors que le gène identifié et localisé sur le groupe de liaison 6 au niveau du QTL, code potentiellement pour une AGP non classique. Les AGP classiques et les AGP non classiques telles que l'AGP-PC2 du Pin (Mau *et al.*, 1995) et, les AGP-Na2 et PvPRP1 du tabac (Mau *et al.*, 1995 ; Chen *et al.*, 1993) présentent en N-terminal un peptide signal mais différent dans leur extrémité C-terminale hydrophobe. En effet, contrairement aux AGP non classiques, les AGP classiques présentent une ancre GPI permettant un ancrage membranaire (Gaspar *et al.*, 2001)

A l'heure actuelle, le rôle des AGP non classiques et des AGP classiques dans l'embryogenèse somatique n'est pas encore connu. Cependant, des expériences récentes ont montré que l'expression du gène codant l'AGP non classique est bloquée par l'addition de réactif Yariv dans le milieu de culture (Lucau *et al.*, soumis). Ce résultat suggère que cette AGP est peut-être indispensable au développement des embryons somatiques mais seulement au cours des phases tardives.

Un gène codant pour une  $\beta$ -tubuline est localisé sous le QTL du groupe de liaison 8. Les tubulines  $\alpha$  et  $\beta$  s'associent en dimères pour former les microtubules, qui sont les constituants principaux de la charpente cellulaire, le cytosquelette, impliqué dans divers processus cellulaires comme les divisions cellulaires et le transport intracellulaire des vésicules et des organites (Wasteneys, 2002).

Chez la carotte, il a été montré que le nombre de microtubules augmente considérablement dans les cellules initiant l'embryogenèse somatique (Cyr *et al.*, 1987). Parallèlement, l'augmentation de la quantité de tubuline est corrélée avec l'augmentation de la taille des cellules présentes dans les phases tardives de l'ES (Cyr *et al.*, 1987). Chez les protoplastes de carottes, l'induction de l'ES a également été corrélée avec un remodelage des microtubules du cytosquelette (Staxen *et al.*, 1994). Un traitement court avec amiprofos-methyl (APM), une drogue qui déstabilise les microtubules, bloque le développement des embryons somatiques chez la luzerne et la carotte (Binorova and Dolezel, 1993). De plus, les quantités de protéines  $\alpha$ -tubuline et  $\beta$ -tubuline diminuent fortement dans les cellules indifférenciées initiant par la suite un programme embryogène (Cyr *et al.*, 1987).

Ainsi, le remodelage du cytosquelette semble indispensable pour réinitier un programme de différenciation cellulaire au cours de l'ES. L'expression du gène codant la  $\beta$ -tubuline pendant l'ES pourrait donc être corrélée avec la redifférenciation cellulaire.

### 3. Les gènes impliqués dans le transport

Trois gènes codant des protéines de transport des lipides (LTP) sont localisés sous les QTL du groupe de liaison 8. Les gènes codant les LTPs ont été caractérisés dans de nombreuses espèces comme *Arabidopsis thaliana* (Thoma *et al.*, 1994), *Daucus carota* (Sterk *et al.*, 1991), *Zea mays* ( Sossountzov *et al.*, 1991), *Hordeum vulgare* (Pearce *et al.*, 1998 ; Chen and Foolad, 1999), *Oryza sativa* (Vignols *et al.*, 1997), *Sorghum vulgare* (Pelèse-Siebenbourg *et al.*, 1994) et *Triticum aestivum* (Jang *et al.*, 2002). Les LTP ont été dans un premier temps décrites pour leur implication dans le transport des phospholipides (Kader *et al.*, 1996, 1997). Cette fonction a été rapidement abandonnée lorsque la pré-protéine présentant une séquence signalée liée à la sécrétion extracellulaire a été identifiée (Yeast and Rose, 2008). Les LTPs sont impliquées dans la défense des végétaux ou la prévention à la déshydratation et quelques-unes sont connues pour être induites par l'ABA (Kader *et al.*, 1996, 1997).

Chez la carotte, l'expression d'un gène codant une nsLTP (EP2) est corrélée à la formation du protoderme, premier tissu différencié chez les embryons somatiques et zygotiques (Sterk *et al.*, 1991). Pendant l'embryogenèse somatique, le gène est exprimé uniformément dans les masses pro-embryogènes mais son expression diminue dans les cellules non embryogènes. Le gène *EP2* se révèle être un bon marqueur de l'embryogenèse (voir analyse bibliographique). *EP2* semble impliqué dans le transport des monomères de cutine, nécessaire à la formation du protoderme entourant les embryons somatiques (Sterk *et al.*, 1991; Hendriks *et al.*, 1994).

Chez *Camellia*, lors de l'induction des feuilles, un dépôt de cutine est nécessaire pour un développement normal des embryons somatiques, ce qui suggère également l'implication des LTPs (Pedroso *et al.*, 1995). Une analyse par microarray a été réalisée chez des embryons somatiques développés à partir de microspores chez le colza (*Brassica napus*). Ceci a permis de caractériser le

changement d'expression des gènes lors du développement gamétophytique, des embryons développés à partir de microspores mais aussi lors de la croissance des embryons (Joosen *et al.*, 2007). Cent trente cinq marqueurs robustes de l'ES haploïdes ont ainsi pu être identifiés, dont 23 correspondent à des LTP ou LTP-like. Les LTP identifiés au cours de la formation des embryons somatiques mais aussi des embryons développés à partir de microspores constituent ainsi de bons marqueurs pour l'ES. Cependant, chez la carotte toutes les cellules embryogènes exprimant les LTP n'aboutissent pas à la formation d'un embryon somatique indiquant que les LTP à elles seules ne permettent pas le développement d'embryons somatiques (Toonen *et al.*, 1994).

L'analyse du profil transcriptomique du gène OC2060 codant pour une LTP présente une hausse d'expression aussi bien chez le génotype embryogène que non embryogène (Legrand *et al.*, 2007). Ces résultats indiqueraient que ce gène n'est pas spécifique de l'embryogenèse somatique mais pourrait être lié au stress induit lors de l'induction des explants.

Deux gènes (OC1138 et OC0206) présents au niveau du QTL détecté sur le groupe de liaison 3, codent pour une HMG-CoA réductase. La présence d'une fonction similaire et la proximité des deux locus sur le même groupe de liaison semblent indiquer que ces locus constituent un seul gène. Deux autres gènes impliqués dans la même voie de biosynthèse des terpènes, HMG-CoA synthase et acetoacetyl-CoA thiolase, sont localisés sur le groupe de liaison 8, mais aucune co-localisation avec un QTL n'a été déterminée pour ces deux gènes. Les profils obtenus en PCR en temps réel (Legrand, 2006) montrent une augmentation d'expression aussi bien chez le génotype embryogène K59 que non embryogène C15 au cours de l'ES. Néanmoins, les gènes impliqués dans la voie de biosynthèse des terpénoïdes sont peut-être liés à la réactivation cellulaire puisque le génotype C15 est capable de démarrer un processus de réactivation cellulaire comme observé chez le génotype K59 (Legrand *et al.*, 2007).

4. Les protéines impliquées dans les phases tardives de l'embryogénèse (LEA, Late Embryogenesis Abundant).

Les déhydrines font partie d'un large groupe de protéines hautement hydrophiles et sont classées dans le groupe II des LEA (Late Embryogenesis Abundant). Les protéines LEA présentent un motif de réponse à l'ABA et s'accumulent particulièrement dans les embryons matures (Hundertmark and Hinch, 2008 ; voir analyse bibliographique).

Afin d'identifier les gènes impliqués lors de l'embryogénèse somatique chez la carotte et d'étudier le mécanisme de leur régulation, Ko et ses collaborateurs, ont développé un système « stress-induction » lors de l'embryogénèse somatique. Utilisant ce système, ils ont isolé et caractérisé un ADNc codant une phosphoprotéine (ECP44) qui apparaît être impliquée dans l'induction de l'embryogénèse somatique (Ko *et al.*, 2006). L'analyse phylogénétique de la séquence en acides aminés de l'ECP44 montre que l'ECP44 a une homologie élevée avec la famille des déhydrines chez la carotte. L'analyse par northern-blot révèle que l'ARNm de l'ECP44 est accumulé dans les embryons somatiques, les bourgeons apicaux non stressés et stressés, les cellules embryogènes mais pas dans les cellules non embryogènes. (Ko *et al.*, 2006).

Chez la chicorée, quatre protéines identifiées comme des déhydrine-like ont été détectées après électrophorèse bidimensionnelle dans les explants foliaires de l'hybride "474" (génotype embryogène) cultivés *in vitro* en condition d'embryogénèse somatique (Dupire, 1995). Ces protéines de différents poids moléculaires (26, 27, 29 et 30 kDa) sont présentes de manière constitutive dans les tissus foliaires et leur quantité diminue au cours de l'induction des embryons somatiques. Il a été supposé que leur disparition pourrait entraîner l'acquisition des potentialités de division en permettant la reprise des divisions abondantes ou en favorisant la dérégulation de la voie embryogène chez le génotype "474" (Dupire, 1995).

### III. STM et ARGONAUTE

Le développement de l'appareil aérien des végétaux supérieurs se caractérise par la production continue d'organes à partir des méristèmes apicaux caulinaires, eux-mêmes constitués de cellules souches mises en place au cours de

l'embryogenèse. Le méristème apical caulinaire maintient la présence d'une population de cellules en prolifération, tout en assurant la différenciation des organes latéraux : feuilles et bourgeons (Sablowski, 2004, 2007).

Des analyses génétiques, réalisées chez l'espèce modèle *Arabidopsis*, ont permis l'identification de nombreux mutants affectés dans les différentes fonctions du méristème. Le clonage et l'étude de l'expression des gènes identifiés ont démontré le haut niveau d'organisation des cellules méristématiques. Plusieurs gènes impliqués dans l'initiation, le développement ou encore le maintien du méristème ont été identifiés comme les gènes *AGO1* (*Argonaute1*) et *STM* (*Shootmeristemless*).

Les gènes *AGO1* et *CiSTM* ont été localisés sous les QTL des groupes de liaison 2 et 6, respectivement. Chez *Arabidopsis*, ces deux gènes sont impliqués dans l'initiation et/ou le développement du méristème caulinaire chez les embryons, et plus spécifiquement dans le maintien des cellules méristématiques dans un état dédifférencié (Kidner et Martienssen, 2005). Des analyses par PCR en temps réel ont révélé une augmentation de l'expression de ces deux gènes au 4<sup>ème</sup> jour de culture en condition d'embryogenèse somatique (S. Legrand, données personnelles). Pour leur expression différentielle et leur co-localisation avec les QTL liés à l'ES, ces deux gènes méritent une étude plus approfondie de leur implication dans l'ES chez la chicorée.

Chez *Arabidopsis*, les mutants *stm* sont caractérisés par une réduction ou une absence du méristème caulinaire dans les embryons matures et les plantules présentant une fusion partielle des cotylédons ; ceci indique que le gène *STM* est impliqué dans le développement et le maintien du méristème caulinaire (Endrizzi *et al.*, 1996). *STM* est un facteur de transcription nucléaire contenant un homéodomaine, elle appartient à la classe 1 des gènes KNOX (KNOTTED1-like homeobox) (Scofield *et al.*, 2006). La protéine *STM* participe activement avec la protéine *WUS* (*Wuschel*) à initier et maintenir les cellules du méristème dans un état dédifférencié (Scofield *et al.*, 2006).

Pour étudier le profil d'expression de *STM*, des hybridations *in situ* ont été réalisées chez les plantes sauvages d'*Arabidopsis* (Long *et al.*, 1996). Dans l'embryon zygotique, les transcrits apparaissent avant l'initiation du méristème, à partir du stade globulaire moyen dans une ou deux cellules. Dans l'embryon tardif, les transcrits *STM* sont détectés dans une bande traversant la moitié supérieure de

l'embryon. Aux stades cœur, torpille et embryon mature, l'expression est restreinte à la base des cotylédons.

Chez *Arabidopsis*, la sous-classe des gènes KNOX comprend également les gènes *KNAT* ou *BP* (Knotted in *Arabidopsis thaliana*, *BREVIPEDICELLUS*) avec *KNAT1*, *KNAT2* et *KNAT6* (Lincoln *et al.*, 1994 ; Dockx *et al.*, 1995 ; Pautot *et al.*, 2001, Scofield *et al.*, 2005). Ces gènes s'expriment dans le méristème apical caulinaire dans des domaines distincts de ceux de *STM* mais chevauchants.

*KNAT1* s'exprime selon un profil proche de celui de *STM* alors que *KNAT2* est présent dans l'assise L3 du méristème apical caulinaire principalement et s'étend dans les parties aériennes les plus internes de l'apex couvrant la zone médullaire. Le gène *KNAT6* est exprimé dans plusieurs organes et tissus, incluant l'apex racinaire et foliaire (Dean *et al.*, 2004).

Aucun mutant de type "perte de fonction" des gènes *KNAT1/KNAT2* n'a été décrit. Cependant, la surexpression du gène *KNAT1*, sous le contrôle d'un promoteur constitutif induit la formation de lobes et de folioles sur les feuilles, ainsi que des tiges ectopiques, suggérant que *KNAT1* est suffisant pour la mise en place de méristème (Chuck *et al.*, 1996).

Les gènes de la famille Argonaute ont été identifiés pour la première fois chez *Arabidopsis* (Vaucheret, 2008). Ils sont au nombre de 10 nommés de *AGO1* à *AGO10*. Ces gènes ont ensuite été isolés chez d'autres espèces (Tableau 2).

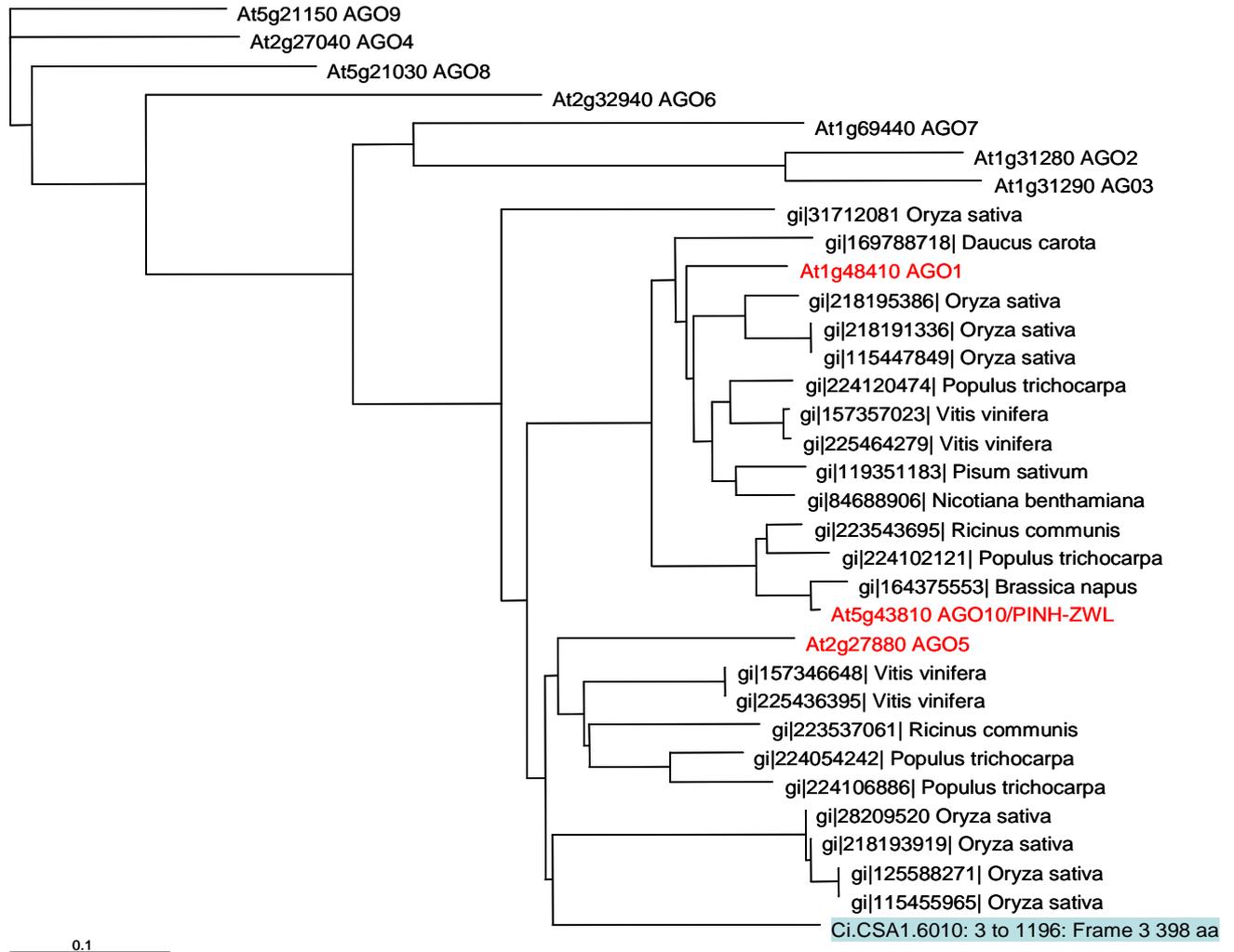
Pour confirmer la fonction du gène *CiAGO*, nous avons procédé à son alignement avec les séquences protéiques codées par les membres de la famille Argonaute d'*Arabidopsis* (logiciel BioEdit v. 7.0.5.3). Les alignements des séquences protéiques partielles (Annexe 18) de l'EST OC2006 (DT213997) de chicorée avec les séquences protéiques Argonaute d'*Arabidopsis* n'ont pas permis de déterminer de manière précise l'homologue de *CiAGO1* mais la séquence protéique semble plus proche de *AtAGO1*, *AtAGO5* ou *AtAGO10/PNH/ZLL*.

**Tableau 2 :** Nombre de gènes de la famille Argonaute identifiés chez différentes espèces (d'après Höck and Meister, 2008).

<b>Espèces</b>	<b>Nombre de gènes identifiés</b>	<b>Références</b>
<i>Homo sapiens</i>	8	Sasaki <i>et al.</i> , 2003
<i>Rattus norvegicus</i>	8	Carmel <i>et al.</i> , 2002
<i>Mus musculus</i>	8	Carmel <i>et al.</i> , 2002
<i>Drosophila melanogaster</i>	5	Williams and Rubis, 2002
<i>Caenorhabditis elegans</i>	27	Yigit <i>et al.</i> , 2006
<i>Arabidopsis thaliana</i>	10	Hunter <i>et al.</i> , 2003
<i>Daucus Carota</i>	1	Takahata, 2008
<i>Schizosaccharomyces pombe</i>	1	Carmel <i>et al.</i> , 2002
<i>Neurospora crassa</i>	2	Carmel <i>et al.</i> , 2002

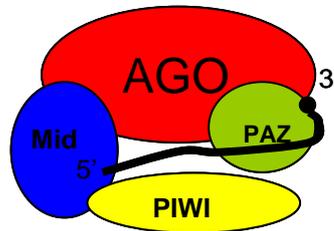
La construction d'un arbre phylogénétique à partir des séquences protéiques de la famille Argonaute de différentes espèces a été réalisée (Figure 1). Un arbre phylogénétique à partir des séquences Argonaute d'*Arabidopsis* a déjà montré qu'AGO1, AGO5 et AGO10/PNH/ZLL sont très proches et présents au niveau du même clade (Höck and Meister, 2008). Les séquences protéiques de *Cichorium intybus* EH677381, EH690450 et EH708067 de la base de données de Davis (CGPD) et la séquence chicorée DT213997 mise en évidence au 4<sup>ème</sup> jour d'induction de l'embryogenèse somatique par analyse microarray (Legrand, 2006) ont été alignées afin de construire un contig (Ci. CSA). Les résultats indiquent que la séquence chicorée CiAGO1 (DT213997) est proche (70% d'homologie) de la séquence d'*Arabidopsis* AGO5.

## Gènes candidats



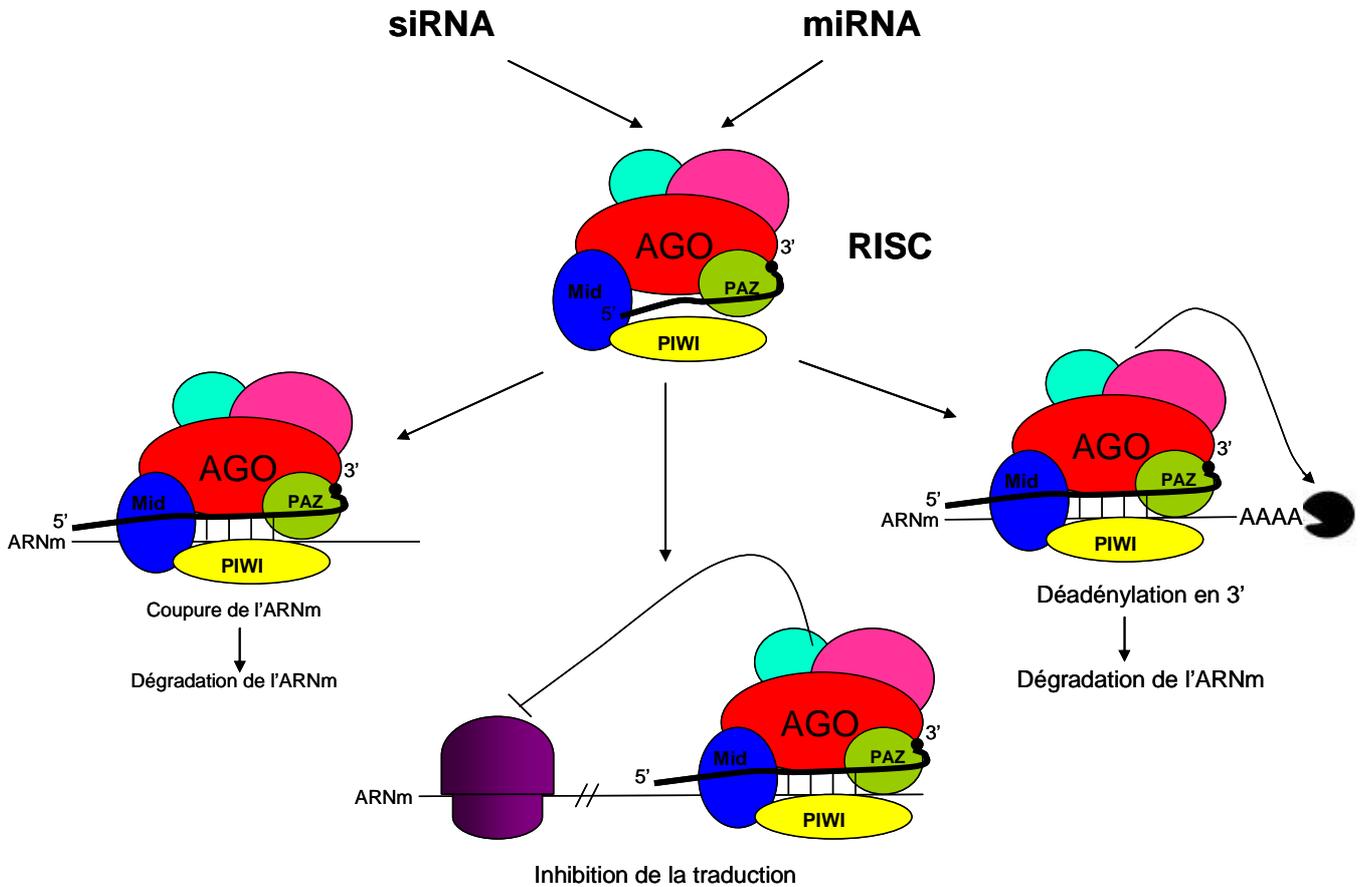
**Figure 1** : Arbre phylogénétique (M.C. Quillet, communication personnelle) issu de l'alignement des séquences protéiques des membres de la famille Argonaute chez différentes espèces. La séquence Ci. CSA (surlignée en bleu) a été obtenue après l'alignement des séquences protéiques EH677381, DT213997, EH690450 et EH708067 de *Cichorium intybus*. Les séquences EH sont issues des bases de données de Davis (CGPD, Compositae Genome Project Database, <http://cgpdb.ucdavis.edu/>). La séquence DT213997 a été mise en évidence au 4<sup>ème</sup> jour d'induction de l'embryogenèse somatique par analyse microarray (Legrand, 2006).

L'étude du mutant *ago1* chez *Arabidopsis* a révélé une forte accumulation des ARNm, connus pour être des cibles des miRNA (Vaucheret *et al.*, 2004), indiquant que ce gène pourrait être impliqué dans le processus de silencing. Un peu plus tard, la protéine AGO1 a été caractérisée (Figure 2) (Song *et al.*, 2004). C'est une protéine de 115 kDa avec un domaine PIWI de 300 acides aminés à activité RNase requis pour le clivage des ARNm, un domaine de 110 acides aminés nommés PAZ capable de se lier à la fin des petits ARN du côté 3' phosphate et un domaine Mid capable de se lier aux petits ARN du côté 5' phosphate (Song *et al.*, 2004).



**Figure 2:** La protéine Argonaute présentant un domaine PAZ capable de se lier aux petits ARN du côté 3' phosphate, un domaine PIWI à activité RNase et un domaine MID capable de se lier aux petits ARN du côté 5' phosphate. Le siARN est représenté en noir (D'après Song *et al.*, 2004).

La protéine Argonaute fait partie de complexes ribo-nucléoprotéiques appelés RISC (RNA-induced silencing complex) impliqués dans la dégradation d'ARNm (Figure 3). Ces complexes contiennent des petits ARNs (21 nucléotides) opérant comme des répresseurs de l'expression génique. Ces petits ARNs sont codés par des unités de transcription autonomes du génome et servent d'adaptateurs aux protéines Argonaute. Grâce à leur complémentarité avec des ARNm, ces petits ARNs placent les protéines Argonaute sur l'ARNm et guident le clivage de celui-ci (Vaucheret, 2008).



**Figure 3:** Action des siARN et miARN chez la plante modèle *Arabidopsis*.

On distingue deux catégories de petits ARNs de 21 nucléotides guidant l'activité des protéines Argonaute du complexe RISC : les petits ARN interférants (siARN) qui dérivent de molécules d'ARN double brin résultant de l'activité d'enzymes cellulaires appelées ARN polymérase ARN dépendante, essentiellement impliqués dans la résistance aux pathogènes et dans les phénomènes d'extinction des transgènes et les microARN (miARN) qui dérivent de molécules d'ARN simple brin, impliqués dans le contrôle du développement et dans la réponse aux stress abiotiques. Les siARN et les miARN guident la protéine Argonaute jusqu'à l'ARN cible conduisant à la dégradation de l'ARNm ou à l'inhibition de la traduction (d'après Song *et al.*, 2004).

Chez *Arabidopsis*, la protéine AGO10, aussi appelée PINHEAD (PHN) ou ZWILLE (ZLL), présente une forte similarité de séquence avec la protéine AGO1 mais aussi une redondance de fonction au cours du développement du méristème (Lynn *et al.*, 1999). De la même manière que AGO1, AGO10 est capable de dégrader les ARNm et est impliquée dans le processus de silencing des gènes. Chez le mutant *ago10/pnh/zll*, le niveau de miARN165/166 est relativement élevé entraînant une forte réduction de transcrits HD-ZIP III (PHAVOLUTA, PHABULOSA et REVOLUTA) impliqués dans le développement des organes latéraux (Liu *et al.*, 2008). AGO10 semble moduler l'accumulation des miARN165/166 afin de préserver le méristème apical caulinaire. La protéine AGO5 a fait l'objet de peu d'études contrairement à AGO1 et AGO10. Des études d'immunoprécipitations ont révélé que AGO5 est préférentiellement liée à des petits ARN avec une cytosine en 5' terminale comme le miARN169 (Takeda *et al.*, 2008). Cependant, il est difficile de comprendre le rôle de ce miRNA puisque le mutant *ago5* d'*Arabidopsis* ne présente pas de développement anormal (Vaucheret, 2008).

L'accumulation des transcrits d'AGO1 est observée dès le stade globulaire. AGO1 est uniformément distribué dans l'embryon propre et le suspenseur et ce profil d'expression reste inchangé au cours de l'embryogenèse. Lors du développement des plantules, les transcrits d'AGO1 sont présents dans le méristème apical caulinaire et les jeunes feuilles (Lynn *et al.*, 1999). Au contraire, les transcrits de AGO10/PNH/ZLL sont détectables à deux niveaux dans l'embryon : les transcrits sont fortement accumulés au niveau des cellules provasculaires et plus faiblement accumulés du côté adaxial des cotylédons et dans les méristèmes. (Lynn *et al.*, 1999).

L'étude des doubles mutants a permis d'établir un lien entre AGO1, AGO10/PNH/ZLL et STM dans le développement du méristème caulinaire (Lynn *et al.*, 1999). AGO1 et AGO10/PNH/ZLL semblent impliqués dans l'accumulation de STM dans le méristème caulinaire. En effet, dans le double mutant *pnh/zll/ago1*, l'accumulation de la protéine STM est diminuée alors que l'accumulation de l'ARNm ne semble pas affectée (Lynn *et al.*, 1999). AGO1 et AGO10/PNH/ZLL semblent jouer un rôle dans la régulation post-transcriptionnelle de STM.

Au contraire, chez le mutant *ago1*, le maintien du méristème et l'expression de *STM* sont seulement affectés de 10%. De manière similaire, chez le mutant *pnh*, l'expression de *STM* pendant les phases précoces n'est pas modifiée dans les embryons, mais lors des phases tardives, l'expression de *STM* est fortement modifiée. En effet, les transcrits *STM* sont confinés dans les cellules latérales du méristème. Ceci suggère que AGO10/PNH/ZLL et AGO1 sont requis pour l'accumulation de la protéine *STM* dans l'embryon mais pas lors de l'initiation du méristème caulinaire (Lynn *et al.*, 1999). Ainsi, le maintien du méristème caulinaire pourrait être sous l'action coordonnée des protéines *STM*, AGO1 et AGO10/PNH/ZLL (Kidner and Martienssen, 2005).

L'effet de AGO1 sur la fonction de *STM* pourrait être direct, cependant à ce jour aucun miARN correspondant à *STM* n'a encore été identifié. Alors, il est vraisemblable que AGO1 agit de manière indirecte sur *STM*. L'identification du miARN (miRNA164) (Reinhard *et al.*, 2002) chez *Arabidopsis* correspondant aux gènes *CUC1* et *CUC2* (*CUPSHAPED COTYLEDONS*), lesquels favorisent la fonction du méristème *via* *STM* (Aida *et al.*, 1999), fournit à ce jour le seul lien connu entre l'action des miARN et l'expression des gènes *KNOX*. Cependant, l'action de ces miARN164 vient inhiber l'action des gènes *CUC* et par la même l'action de *STM* dans le maintien des cellules différenciées. Ces miRNA ont un rôle négatif sur le développement du méristème alors que la protéine AGO1 favorise l'expression des gènes *KNOX*.

Alors AGO1 et AGO10/PNH/ZLL pourraient agir plutôt en inhibant certains répresseurs de *STM* impliqués dans le développement des organes latéraux. La répression des gènes *KNOX* requiert la présence de facteurs de la famille MYB, codés par les gènes *ROUGH SHEATH 2* (*RS2*, Timmermans *et al.*, 1999 ; Tsiantis *et al.*, 1999), *PHANTASTICA* (*PHAN*, Waites *et al.*, 1998) et *ASYMMETRIC LEAVES 1* (*AS1*, Byrne *et al.*, 2000). Une mutation associée à une perte de fonction dans un de ces gènes donne un phénotype proche du phénotype associé à une sur-expression des gènes *KNOX* (Timmermans *et al.*, 1999 ; Tsiantis *et al.*, 1999 ; Byrne *et al.*, 2000 ; Ori *et al.*, 2000 ; Semiarti *et al.*, 2001). Chez les mutants *as1* d'*Arabidopsis*, les gènes *KNAT* (gènes de la famille *KNOX*) sont exprimés ectopiquement dans les feuilles en développement (Byrne *et al.*, 2000 ; Ori *et al.*, 2000), suggérant que *AS1* pourrait réprimer leur expression dans les primordia.

Toutefois, les mutations *as1*, *rs2* et *phan* n'affectent pas l'expression des gènes KNOX dès le recrutement des cellules du primordium. Alors, l'extinction des gènes KNOX se fait normalement dans ces mutants, ce n'est qu'après qu'ils sont réactivés (Timmermans *et al.*, 1999 ; Tsiantis *et al.*, 1999 ; Ori *et al.*, 2000). Il semble que ces gènes soient importants pour maintenir l'identité organogène des cellules mais qu'ils ne soient pas responsables de la perte de l'identité des cellules méristématiques.

A ce jour, aucun gène réprimant l'expression de *STM* conduisant à une perte de l'identité des cellules méristématiques n'a été identifié.

#### IV. Conclusion

Au total, 16 gènes, sur les 63 gènes cartographiés sur la carte génétique F1'a et liés au processus de l'embryogenèse somatique, co-localisent avec les QTL détectés sur les groupes de liaison 1, 2, 3, 6 et 8.

Les gènes non liés significativement aux QTL peuvent participer directement ou indirectement à la formation des embryons somatiques mais ne semblent pas impliqués dans la variation du nombre d'embryons somatiques observés dans la descendance F1'a.

Plusieurs de ces gènes co-localisant avec les QTL semblent impliquer dans des processus généraux de développement comme les divisions cellulaires et le transport cellulaire. Cependant, deux de ces gènes, *AGO1* et *STM*, impliquer dans le développement du méristème caulinaire chez les embryons peuvent s'avérer de très bons gènes candidats pour accroître la capacité embryogène des individus.

Pour tenter d'élucider leur rôle, des expériences d'hybridation *in situ* pourraient être envisagées afin d'observer le profil d'expression de ces gènes au niveau tissulaire et cellulaire. Pour valider leur fonction dans la formation des embryons somatiques, l'expression de *AGO1* et *STM* pourrait être éteinte chez un génotype très embryogène afin de voir si ces deux gènes sont impliqués dans la variation quantitative du nombre d'embryons somatiques formés.

Une étude plus poussée de *CiAGO1* et *CiSTM* chez la chicorée pourrait être envisagée par l'utilisation des techniques de génétique reverse par mutagenèse insertionnelle ou TILLING (Targeted Induced Local Lesions In Genes). Contrairement aux bactéries, l'inactivation par recombinaison homologe est difficile à contrôler

chez les plantes (Kempin *et al.*, 1997). L'approche anti-sens s'est avérée plus efficace pour bloquer la production d'une protéine. Elle consiste à faire exprimer chez les plantes un ARN (ARN anti-sens) complémentaire à l'ensemble ou à une partie d'un ARNm visé auquel il peut s'hybrider et ainsi servir de répresseur. En effet, la formation d'un ARN peut interférer avec la maturation de l'ARNm, son transport à partir du noyau ou sa traduction dans le cytoplasme entraînant la dégradation de l'ARNm.

Il serait intéressant de réaliser des transformations de tissus racinaires mais aussi foliaires de chicorée de génotypes très embryogènes, de façon à surexprimer, grâce à un promoteur fort (35S) des ARN anti-sens correspondant à l'ADNc de *CiAGO1* et *CiSTM*. Les plantes transformées seraient sélectionnées après une étude moléculaire en observant le niveau d'expression des gènes et la quantité de protéines présentes pour *CiAGO1* et *CiSTM*. Pour confirmer l'implication de ces deux gènes dans l'ES, une analyse phénotypique des plantes transformées, par un comptage du nombre d'embryons somatiques formés, devra être réalisée. En effet, si la capacité embryogène des plantes transformées est fortement diminuée alors le rôle de *CiAGO1* et *CiSTM* dans l'ES sera peut-être validé. Cependant, à ce jour, seules quelques plantules transformées pour *STM* et cultivées *in vitro* (ne présentant pas de méristème apical caulinaire), ont pu être régénérées chez *Arabidopsis* (Barton *et al.*, 1993). Nous pourrions alors envisager d'observer la capacité embryogène des individus à partir de cals transformés. L'avantage de cette approche, est qu'elle ne nécessite pas d'aboutir au développement complet et au maintien des plantules transformées.

Toutefois, chez *Arabidopsis*, la redondance de fonction et de localisation des protéines AGO10/PNH/ZLL et KNAT1 avec AGO1 et STM, respectivement, nous amène à considérer les gènes codants pour AGO10/PNH/ZLL et KNAT1 dans la suite de l'étude. En effet, la répression du gène *CiAGO1* et ou *CiSTM* seul ne permettrait peut-être pas d'observer un phénotype et donc de confirmer l'implication du gène dans l'ES.

Cependant, contrairement à *Arabidopsis* où l'obtention et le maintien de mutants sont relativement faciles, la chicorée est une espèce plus difficile à transformer. Toutefois, des transformations par RNAi de racines chez *Cichorium intybus* ont déjà été réalisées. En effet, dans le but d'inhiber l'action d'une enzyme

## Gènes candidats

(fructan-1-exohydrolase) impliquée dans la dégradation de l'inuline (Asad, 2006) les auteurs ont utilisé une approche anti-sens.

## Chapitre V

### Discussion générale et Perspectives

Ce travail avait pour but, l'étude du contrôle génétique des phases de l'embryogenèse somatique chez la chicorée. En effet, nous avons pour objectif de détecter les régions impliquées dans l'ES. Pour cela, nous avons emprunté une approche classique : obtention d'une descendance en ségrégation, phénotypage du caractère d'intérêt, construction d'une carte génétique, analyse QTL et identification des gènes candidats qui co-localisent avec les QTL détectés.

Notre analyse a débuté par la sélection de parents à phénotype extrême pour la capacité à former des embryons somatiques : K59, génotype embryogène, et K28, génotype peu embryogène. Le génotype peu embryogène a été utilisé comme parent femelle et le génotype embryogène K59 comme parent mâle permettant d'obtenir une descendance en ségrégation pour le caractère d'embryogenèse somatique. Ainsi, une première sous-descendance F1'a de 123 génotypes a été obtenue en 2001. Puis, afin d'augmenter l'effectif de notre sous-descendance, de nouveaux croisements entre les parents K28 et K59 ont été réalisés en 2004, générant ainsi 104 nouveaux individus correspondant à la sous-descendance F1'b. Ces deux descendances ont été caractérisées quant à leur capacité à former des embryons somatiques. L'évaluation du test d'embryogenèse somatique a été réalisée à partir d'explants racinaires et ceci après 30 jours de culture en condition de l'embryogenèse somatique. Nous avons comptabilisé des plantules, issues du développement des embryons somatiques mais aussi des structures chlorophylliennes unipolaires que nous avons préféré, dans un premier temps, distinguer des plantules ne connaissant pas l'origine de ces structures. Après transformation des valeurs obtenues en  $\log_{10}$ , la fréquence de distribution s'est avérée normale et continue quelle que soit la sous-descendance étudiée. Ce résultat est venu conforter notre choix quant aux génotypes K28 et K59 comme parents pour l'obtention de descendances en ségrégation pour la capacité à former des embryons somatiques. Les caractères liés aux plantules et aux structures chlorophylliennes unipolaires formées présentent une héritabilité forte avec, en moyenne, 75 et 70% dans la sous-descendance F1'a et F1'b, respectivement. Dans la sous-descendance F1'b, parallèlement à l'évaluation du potentiel embryogène, nous avons également caractérisé la capacité organogène des individus à partir des pétioles. Nous avons, pour chaque sous-descendant, comptabilisé le nombre de cals formés mais aussi le

nombre de bourgeons régénérés. Après transformation des valeurs en  $\log_{10}$ , la distribution s'est avérée normale et continue. L'héritabilité calculée est de 65% et 35% pour la formation des bourgeons et des cals, respectivement. La faible héritabilité pour la régénération des cals par rapport aux bourgeons pourrait indiquer que ces 2 processus sont indépendants. En effet, certains génotypes ont développé des cals sans toutefois initier la formation de bourgeons.

Dans le chapitre II, nous avons dans un premier temps cartographié de nouveaux marqueurs SSR, EST et AFLP—sur la carte génétique K28\*K59 déjà existante au laboratoire. Ces marqueurs sont venus améliorer la densité de la carte génétique de la chicorée (*Chicorium intybus* L.). La carte comporte ainsi 518 marqueurs répartis sur les 9 groupes de liaison correspondant au nombre de chromosomes haploïdes chez la chicorée.

La recherche de QTL a nécessité pour chacun des individus des sous-descendances de connaître leur valeur phénotypique et leur génotype pour chaque marqueur utilisé dans l'analyse QTL. Il n'était pas possible dans le temps qui nous était imparti de génotyper l'ensemble des marqueurs, à partir des ADN de la sous-descendance F1'b, présents sur la carte génétique F1'a. Nous avons alors choisi certains marqueurs, si possible codominants, sur chaque groupe de liaison distants en moyenne de 10 cM. Ainsi, une carte composée de 124 marqueurs (52 SSR, 33 STS et 39 AFLP) a été construite pour permettre une analyse QTL.

A partir des données phénotypiques et génotypiques obtenues, nous avons réalisé une analyse QTL pour identifier les locus impliqués dans la variabilité embryogène et organogène des individus des sous-descendances F1'a et F1'b. Au total, 6 QTL ont été identifiés avec une valeur de LOD entre 3,85 et 8,8 et, expliquant plus de 23% et 44% de variation phénotypique totale pour les caractères PL et SH, respectivement, dont 2 sont spécifiques de la sous-descendance F1'a et 1 de la sous-descendance F1'b. Les QTL détectés pour les caractères PL et SH sont fortement corrélés indiquant une base génétique commune. La caractérisation du potentiel organogène des génotypes de la sous-descendance F1'b a permis de détecter 2 QTL pour chacun des caractères liés à la callogenèse (Cal2 et Cal8) et à l'organogénèse (Bg3 et Bg8) expliquant 32,3% et 42,1% de variation phénotypique totale, respectivement. L'analyse à partir des 104 individus de la sous-descendance F1'b, n'a peut-être pas permis d'identifier l'ensemble des locus impliqués. Le potentiel organogène n'a pas pu être caractérisé dans les descendants de F1'a

puisque les plantules cultivées *in vitro* n'existaient plus. Ces données auraient peut-être permis d'augmenter l'effectif total de la descendance et ainsi favorisé la détection d'autres QTL à effet mineur.

Le but de notre analyse QTL était d'identifier les gènes candidats, présents au niveau des QTL, responsables de la variation phénotypique liée à l'ES. Pour cela, nous avons entrepris de cartographier les gènes mis en évidence au 4<sup>ème</sup> jour de l'ES par microarray et/ou PCR en temps réel. Le fait d'associer la cartographie de gènes candidats et l'identification de locus liés à l'embryogenèse somatique, nous a permis de restreindre le nombre de gènes candidats. En effet, 47 EST ont pu être cartographiés sur la carte génétique F1'a et seulement 16 co-localisent avec les QTL détectés dans les descendance K28\*K59. Cependant, comme nous l'avons indiqué dans le chapitre II, le taux d'EST détectés polymorphes était faible ce qui demanderait d'utiliser une autre technique comme la HRM (High Resolution Melting) afin de mieux identifier les SNP au sein des séquences. Cette méthode nous permettrait peut-être de cartographier d'autres gènes mis en évidence par l'analyse transcriptomique et validés par des analyses de PCR en temps réel et ainsi de confirmer l'implication de ces gènes dans l'embryogenèse somatique. De plus, les analyses des descendance issues des croisements entre d'autres plantes avec une variation pour la formation des embryons somatiques pourraient conduire à la détection des QTL additionnels qui co-localisent avec d'autres gènes candidats.

Les résultats présentés ci-dessus ont soulevé trois points intéressants que nous allons discuter dans la suite de ce chapitre : la différence phénotypique entre les 2 sous-descendance, la présence d'une région en distorsion sur le groupe de liaison 7 dans la sous-descendance F1'b et l'origine des structures chlorophylliennes unipolaires.

Le potentiel embryogène des sous-descendance F1'a et F1'b a été évalué en 2004 et 2006, respectivement. La formation des plantules dans les deux descendance s'est avérée différente, avec un nombre plus élevé dans la sous-descendance F1'a comparé à la sous-descendance F1'b. L'inverse a été observé pour le nombre de structures chlorophylliennes. De plus, la corrélation entre le nombre d'embryons somatiques et le nombre de structures chlorophylliennes unipolaires formées semble plus élevée dans la sous-descendance F1'b que dans la

sous-descendance F1'a. Au contraire, le nombre d'événements totaux, c'est à dire l'ensemble des plantules et des structures chlorophylliennes unipolaires formées, est comparable avec en moyenne  $104 \pm 105$  et  $117 \pm 82$  structures formées dans les sous-descendances F1'a et F1'b, respectivement. La variabilité observée pour la capacité embryogène entre les deux sous-descendances nous a contraints dans un premier temps à réaliser les analyses QTL séparément. Une première analyse a été réalisée à partir des individus de la sous-descendance F1'a et une deuxième à partir des individus de la sous-descendance F1'b. La recherche de QTL à partir d'un effectif faible ne permet pas toujours d'identifier la totalité des locus liés aux caractères étudiés. Toutefois, les trois QTL majeurs présents sur le groupe de liaison 2, 6 et 8 ont été détectés dans chacune des sous-descendances. De plus, la robustesse de ces 3 QTL a été confirmée par leur détection dans la descendance F1'a+b.

Comment expliquer une telle variabilité phénotypique ? Certains marqueurs du groupe de liaison 7 de la sous-descendance F1'b présentent une forte distorsion de ségrégation qui n'a pas été observée dans la sous-descendance F1'a. La présence d'une distorsion de ségrégation dans cette région a également été observée dans la population Rubis 118, une descendance F2 issue d'un croisement entre un clone mâle stérile MS8, ayant la même origine que K59 et K28, et un cultivar de chicorée industrielle "Cassel" (Florimond-Desprez, Capelle-en-Pevele, France) utilisée pour la construction de la carte consensus chez la chicorée (Cadalen *et al.*, 2009). La distorsion dans la sous-descendance F1'b pourrait être responsable d'une diminution de la capacité embryogène et/ou d'une augmentation du nombre de structures chlorophylliennes formées dans la sous-descendance F1'b. L'analyse, pour déterminer l'origine de la distorsion des marqueurs sur le groupe de liaison 7, indique que les facteurs de sélection gamétique seraient responsables de cette distorsion. La sélection s'est peut-être faite au niveau du pollen (germination, compétition du tube pollinique). Les conditions environnementales différentes en 2001 et 2004 ont peut-être influencé cette sélection. Les 2 séries de croisements ont été réalisées en serre au cours des mois de juillet et août. N'ayant pas les relevés de température, d'ensoleillement et d'humidité en 2001 et 2004 de la serre, nous nous sommes intéressés aux relevés météorologiques (Figure 1, p196) puisque les conditions en serre sont contrôlées par rapport aux conditions externes. Si les moyennes des données de température relevées au cours des mois de juillet et août en 2001 et 2004 présentent peu de différences ; le nombre d'heures d'ensoleillement

différent en août ;236h en août 2001 contre 163h en août 2004. Les conditions climatiques semblent assez semblables.

Une température plus élevée de 4°C en 2004 par rapport à 2001 a été relevée lors de la récolte des graines en septembre. Toutefois, cette différence de température ne semble pas avoir eu d'effet sur le développement des embryons zygotiques. En effet, en moyenne,  $12,2 \pm 3,6$  et  $12,6 \pm 3,3$  graines ont été récoltées par capitule en 2001 et 2004, respectivement, et le taux de germination de ces graines était de 100% pour les deux sous-descendances F1'a et F1'b. Ainsi, aucune sélection n'a été réalisée lors de la germination des graines. Ces résultats indiqueraient que les facteurs de sélection zygotiques ne seraient pas responsables de la variation du potentiel embryogène entre les 2 sous-descendances ce qui confirme la présence d'une sélection gamétique qui a été révélée par l'analyse génétique.

Nous avons ensuite réfléchi à d'autres paramètres (condition de culture *in vitro* et expérimentateur) qui auraient pu aboutir à des variabilités phénotypiques.

Le potentiel embryogène des plantules a été évalué sur la totalité des plantules pour la sous-descendance F1'b alors que pour 54 plantules issues de la sous-descendance F1'a celui-ci n'a pu être déterminé. Ces plantules n'ont pu être maintenues *in vitro*, ceci à cause des contaminations survenues lors des tests d'induction en condition d'embryogenèse somatique mais aussi lors de la multiplication par organogenèse. L'élimination de ces individus a peut-être induit un biais dans l'échantillonnage de cette sous-descendance en favorisant ou défavorisant certaines classes génotypiques. Prenons l'exemple de 2 situations extrêmes pour les 54 plantules, dont le potentiel embryogène n'a pas pu être évalué. Dans le premier cas, les plantules pourraient avoir un caractère embryogène élevé donc la proportion d'individus fortement embryogènes serait plus élevée dans la descendance F1'a. Dans le second cas, les plantules présenteraient un potentiel embryogène faible, et la proportion d'individus faiblement embryogènes serait ainsi plus élevée dans la descendance F1'a. Dans les 2 cas décrits, certes la moyenne serait modifiée mais la tendance observée dans la sous-descendance F1'a par rapport à la sous-descendance F1'b resterait la même c'est-à-dire une forte embryogénicité des génotypes issus de la première série de croisements réalisée en 2001.

## Discussion générale et Perspectives

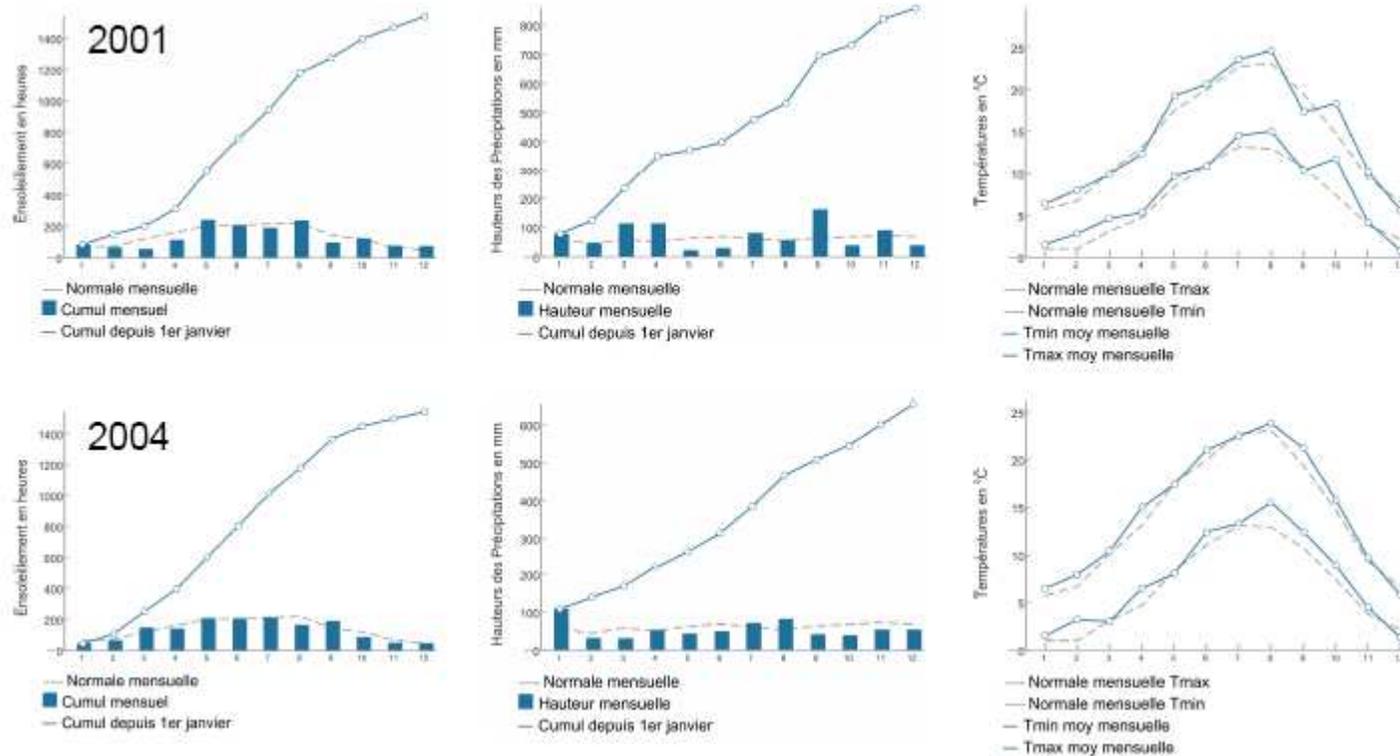


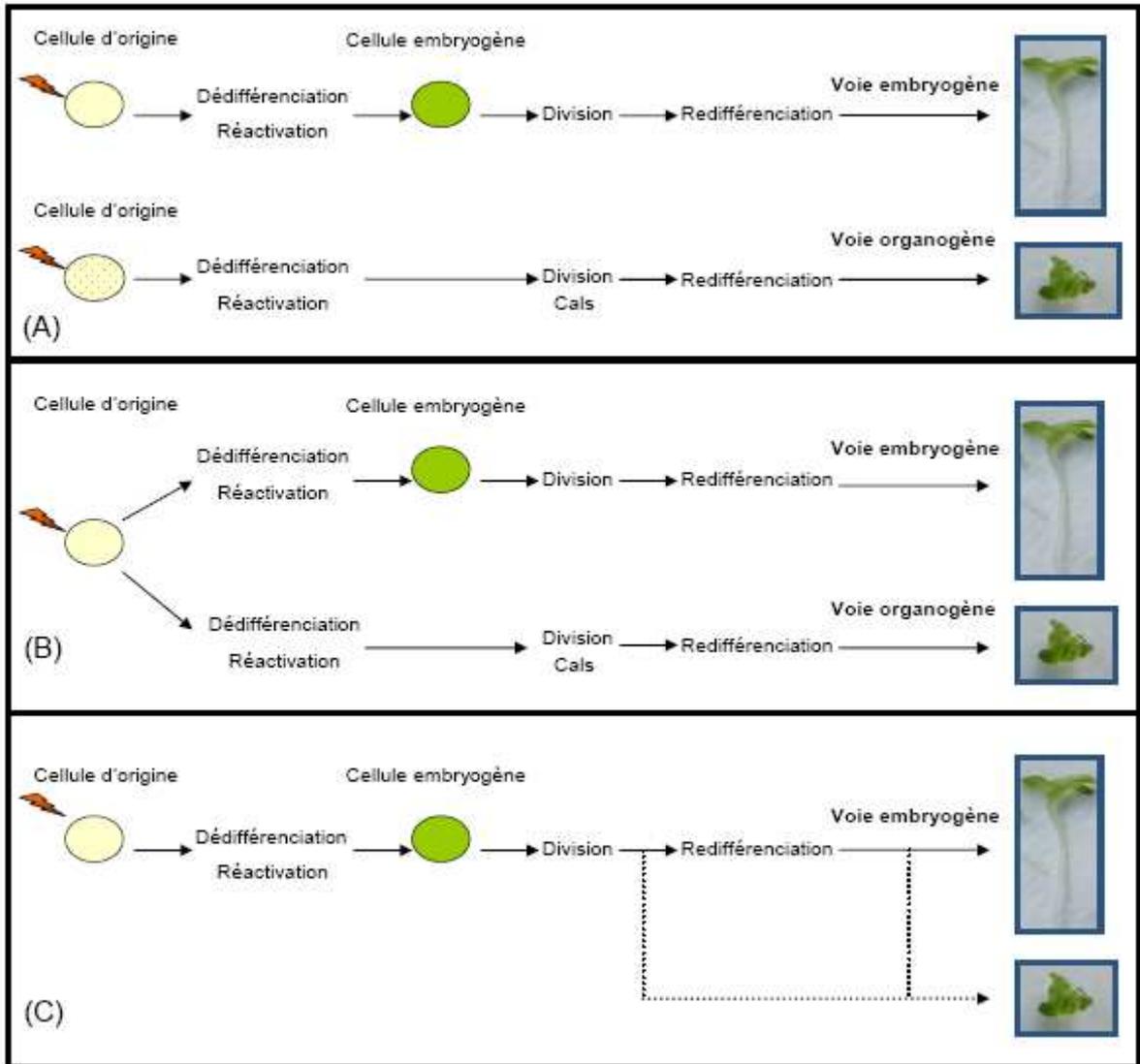
Figure 1 : Relevés météorologiques à la station de Lesquin des mois de janvier à décembre (1 à 12) en 2001 et 2004.

La première sous-descendance F1'a générée en 2001 a été évaluée par MC Quillet (1<sup>er</sup> expérimentateur). L'évaluation du potentiel embryogène des individus a été réalisée sur une période de 9 mois. Les critères de définition des plantules issues des embryons somatiques ont été définis par le 1<sup>er</sup> expérimentateur. La deuxième sous-descendance F1'b obtenue en 2004 a été évaluée sur une période de 1 an et demi par moi-même (2<sup>ème</sup> expérimentateur). Les critères de discriminations appliqués aux structures formées après 30 jours en condition d'embryogenèse somatique afin de définir la capacité embryogène des individus ont été les mêmes que lors de l'évaluation de la descendance F1'a. Des comptages de plantules et de structures chlorophylliennes unipolaires formées, ont été réalisés en parallèle, avec les deux expérimentateurs, à partir d'une même culture et ceci sur plusieurs génotypes, et n'ont révélé aucune différence.

La réalisation d'un test d'induction d'embryogenèse somatique nécessite de prélever les racines issues des plantules cultivées *in vitro*. Le prélèvement des racines des jeunes plantules a été réalisé de manière qualitative, c'est-à-dire lorsque les racines commencent à s'enrouler sur elles-mêmes au fond du tube à essai. Le stade de différenciation et de développement des racines pourrait avoir un impact sur la formation des embryons somatiques. Afin de réduire ce biais entre 10 et 26 répétitions ont été réalisées par génotype mais surtout un seul test d'induction a été réalisé pour un génotype donné à une date donnée. Le fait de réaliser un seul test pour un génotype à une date donnée permet également de réduire les effets du milieu d'induction et/ou de développement des embryons somatiques. En effet, durant la période pendant laquelle les tests d'induction d'embryons somatiques ont été réalisés, les différents milieux utilisés ont été préparés plusieurs fois. Selon la précision de la préparation, ces milieux peuvent très légèrement varier dans leur composition et ainsi influencer l'initiation ou le développement de la formation d'embryons somatiques ou de structures chlorophylliennes unipolaires dans les sous-descendances F1'a et F1'b. De plus, ce biais a aussi été écarté dans l'analyse avec l'élimination de 10% ou de manière plus stricte, 50% des valeurs comptabilisées correspondant aux valeurs les plus faibles.

Lors de l'évaluation du potentiel embryogène des génotypes à partir d'explants racinaires, nous avons pu observer des plantules issues de la formation des embryons somatiques mais aussi des structures chlorophylliennes unipolaires

d'origine inconnue. Ainsi, nous avons préféré, dans un premier temps, les considérer séparément. Les QTL identifiés à partir des données correspondant aux structures chlorophylliennes unipolaires co-localisent avec chacun des QTL liés à la formation des plantules dans chaque sous-descendance, indiquant que ces deux processus sont étroitement liés. Ce constat, nous a amené à émettre 3 hypothèses concernant l'origine des structures chlorophylliennes unipolaires (Figure 2). D'une part, les structures chlorophylliennes unipolaires pourraient être issues d'une voie de développement organogène (Figure 2A). Cependant, au regard des résultats obtenus lors des analyses QTL et du calcul de corrélation entre plantules et structures chlorophylliennes unipolaires formées, cette hypothèse semble peu probable. La deuxième hypothèse (Figure 2B) suggérait que les structures chlorophylliennes unipolaires auraient pour origine les mêmes cellules qui initient la formation des embryons somatiques. L'application d'un stimulus ou d'un stress par exemple aurait peut-être dirigé le développement vers une voie embryogène ou organogène. Enfin, les structures chlorophylliennes unipolaires pourraient suivre la même voie de développement que les embryons somatiques mais le développement n'aboutirait pas à la formation de plantules (Figure 2C). Ainsi, ces deux structures auraient emprunté les mêmes étapes de développement impliquant les mêmes gènes ce qui expliquerait la co-localisation des QTL liée aux structures embryogènes et aux structures chlorophylliennes unipolaires.



**Figure 2 :** Hypothèses quant à l'origine des structures chlorophylliennes unipolaires formées après induction d'explants racinaires en condition d'embryogenèse somatique.

Des analyses cytologiques sur des explants racinaires placés en condition d'induction de l'embryogenèse somatique ont révélé la présence de cellules en cours de réactivation mais également de cellules avec un nombre élevé de vacuoles, caractéristiques des cellules non différenciées ; les cals (A.S. Blervacq, données personnelles). Au même titre que le développement des embryons somatiques, la formation de cals passe par des étapes de dédifférenciation, division cellulaire et redifférenciation. Ainsi, certains gènes impliqués dans ces processus pourraient être communs expliquant la présence d'une co-localisation des QTL. Néanmoins, l'identification de 3 gènes *CiSTM*, *CiAGP* et *CiAGO1* associés aux QTL liés à la formation des plantules et à la formation des structures chlorophylliennes unipolaires pourrait nous amener à valider la troisième hypothèse énoncée (Figure 2C).

Le dénombrement des structures organogènes est plus aisé à partir des explants foliaires. En effet, après la mise en culture des fragments de pétioles, nous avons réalisé un comptage du nombre de cals puis de bourgeons formés. Nous pouvons remarquer qu'aucun des QTL détectés par cette analyse ne co-localise avec les QTL liés à la formation des plantules et des structures chlorophylliennes unipolaires. Même s'il est difficile de comparer les QTL obtenus à partir d'explants foliaires et racinaires, ce résultat pourrait peut-être indiquer que les structures chlorophylliennes unipolaires n'ont pas suivi la voie de développement organogène. Par conséquent, les structures chlorophylliennes unipolaires seraient des embryons somatiques n'ayant pas terminé leur développement et donc abouti à la formation de plantules. Des travaux récents (Joosen *et al.*, 2007) ont montré que le temps d'induction du processus d'embryogenèse somatique à partir de microspores chez le colza avait un impact sur le développement des embryons somatiques. En effet, un temps d'induction court de 8h permet l'obtention d'un grand nombre d'embryons avec un grand suspenseur (71,3%) et au contraire un temps d'induction plus long de 96h entraîne la formation d'une grande proportion d'embryons n'ayant pas développés de suspenseur (68,8%). Chez la chicorée, l'induction des embryons somatiques pourrait alors être diminuée afin de réduire la formation de structures chlorophylliennes unipolaires c'est-à-dire d'embryons malformés.

Dans la littérature, les gènes *STM*, *AGP* et *AGO1* seraient impliqués de manière directe ou indirecte dans l'ES. Dans notre étude, ces gènes sont liés à des

QTL robustes quant à l'évaluation du potentiel embryogène des génotypes de la descendance K28\*K59. Ceci nous permet d'envisager des études plus approfondies.

D'un point de vue fondamental, il serait intéressant d'identifier par hybridation *in situ*, dans quel type de cellules, au sein des explants K59, s'expriment *CiSTM*, *CiAGP* et *CiAGO1*. Ces observations, déjà réalisées pour *STM* et *AGO1* dans les embryons zygotiques –d'*Arabidopsis*, ont révélé que *STM* est confiné au niveau du méristème caulinaire alors qu'au contraire *AGO1* s'étend jusqu'aux organes latéraux. En effet, les transcrits de *STM* s'accumulent dans les cellules initiales du méristème, à partir du stade globulaire moyen dans une ou deux cellules (Long *et al.*, 1996). Chez la chicorée, comme chez *Arabidopsis*, *STM* est exprimé pendant l'embryogenèse zygotique au moment de la formation du méristème caulinaire (Da Silva, 2004). Au contraire, pendant l'ES, *STM* est exprimé bien avant la mise en place d'un méristème apical caulinaire. L'apparition précoce des transcrits dans des cellules non différenciées pourrait suggérer un autre rôle pour *STM* dans l'ES, par exemple, *STM* est peut-être impliqué dans le processus de dédifférenciation cellulaire nécessaire pour la génération d'embryons somatiques.

Des études de sous-expression d'*AGO1* et *STM* pourraient valider l'implication de ces gènes dans la formation des embryons somatiques. Il serait judicieux de sous-exprimer les deux gènes en même temps chez un génotype très embryogène pour voir un effet sur le développement des embryons somatiques. En effet, le processus d'embryogenèse somatique est un caractère quantitatif impliquant plusieurs gènes donc l'extinction d'un seul gène ne permettra peut-être pas d'observer une variation significative du nombre d'embryons somatiques formés. De plus, chez *Arabidopsis*, il existe une forte redondance des gènes participant au maintien et/ou au développement du méristème apical caulinaire notamment *AGO10/PHN/ZLL* et *KNAT1* avec *AGO1* et *STM*, respectivement. La sous-expression de *CiAGO1* et *CiSTM* seuls ne permettra peut-être pas de visualiser une modification du potentiel embryogène de l'individu transformé. Pour éviter de sous-exprimer plusieurs gènes simultanément, il serait peut-être plus facile de réaliser des transformations avec un génotype ne présentant qu'un seul QTL lié à *CiSTM* (QTL au niveau du groupe de liaison 6) ou *CiAGO1* (QTL au niveau du groupe de liaison 2).

Les résultats obtenus au cours de cette thèse pourraient être utilisés afin d'augmenter la capacité embryogène des chicorées commerciales généralement récalcitrantes à l'ES. En effet, des tests d'ES réalisés à partir de différents cultigrupes de chicorées commerciales n'ont pas permis de régénérer des embryons somatiques. Pour cela, l'utilisation de marqueurs pour l'introggression par backcross d'un ou plusieurs gènes ou QTL d'une lignée "donneuse" dans une lignée "receveuse" pourrait être envisagée. Cette stratégie, permettrait d'introggresser les QTL majeurs (QTL des groupes de liaison 2, 6 et 8) liés à la capacité embryogène dans les chicorées commerciales pour augmenter la capacité embryogène des plantes et ainsi, maintenir et multiplier rapidement les génotypes élites. Pour être économiquement rentable, il faudrait introgresser simultanément les QTL à effets majeurs. Dans le chapitre III, nous avons montré que le caractère lié à l'embryogenèse somatique est gouverné par 3 QTL majeurs. La faible proportion du génome qui semble responsable de la variabilité de la capacité embryogène pourra faciliter l'introggression du caractère dans le génotype élite. N'ayant pas identifié les gènes responsables des QTL, l'utilisation de marqueurs flanquants pour contrôler l'introggression des QTL sera nécessaire.

# Références bibliographiques

## A

Aida M., Ishida T., Tasaka M. (1999). Shoot apical meristem and cotyledon formation during *Arabidopsis* embryogenesis: interaction among the *CUP-SHAPED COTYLEDON* and *SHOOT MERISTEMLESS* genes. *Development* 126, 1563-1570.

Andersen PS, Jespersgaard C, Vuust J, Christiansen M, Larsen LA (2003). High-throughput single strand conformation polymorphism mutation detection by automated capillary array electrophoresis: validation of the method. *Human Mutat* 21:116-122.

Anil V.S., Harmon A.C., Rao K. S. (2000). Spatio-temporal accumulation and activity of calcium dependent protein kinases during embryogenesis, seed development, and germination in sandalwood. *Plant Physiology* 122:1035-1043.

Armstrong C.L., Romero-Severson J., Hodges T.K. (1992). Improved tissue culture response of an elite maize inbred through back-cross breeding, and identification of chromosomal regions important for regeneration by RFLP analysis. *Theor. Appl. Genet.* 84:755-762.

Avivi Y., Morad V., Ben-Meir H., Zhao J. et al., (2004). Reorganization of specific chromosomal domains and activation of silent genes in plant cell acquiring pluripotentiality. *Developmental Dynamics* 230:12-22

Asad M. (2006). RNA interference (RNAi) as a tool to engineer high nutritional value in chicory (*Chicorium intybus*). *Commun. Agric. Appl. Biol. Sci.* 71:75-78.

## B

Baldwin T.C., Domingo C., Schindler T., Seetharaman G., Stacey N., Roberts K. (2001). DcAGP1, a secreted arabinogalactan protein, is related to a family of basic praline-rich proteins. *Plant Mol. Biol.* 45:421-435.

## Références bibliographiques

Balestrazzi A., Toscano I., Bernacchia G., Luo M., Otte S., Carbonera D. (1996). Cloning of a cDNA encoding DNA topoisomerase I in *Daucus carota* and expression analysis in relation to cell proliferation. *Gene* 183:183-190.

Balestrazzi A., Bernacchia G., Pitto L., Luccarini G., Carbonera D. (2001). Spatial expression of DNA topoisomerase I genes during cell proliferation in *Daucus carota*. *Eur. J. Histochem.* 45:31-38.

Barakat A., Szick-Miranda K., Chang I.F., Guyot R., Blanc G., Cooke R., Delseny M., Bailey-Serres J. (2001). The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome. *Plant Physiol* 127:398-415.

Baudino S., Hansen S., Brettschneider R., Hecht V.R.G., Dresselhaus T., Lorz H., Dumas C., Rogowsky P.M. (2001). Molecular characterisation of two novel maize LRR receptor-like kinases, which belong to the SERK gene family. *Planta* 213:1-10.

Ben C. (2005). Analyse du transcriptome lors de l'embryogenèse précoce chez le tournesol. Thèse à l'université de Toulouse.

Ben Amer I.M., Korzun V., Worland A.J., Börner A. (1997). Genetic mapping of QTL controlling tissue-culture response on chromosome 2B of wheat (*Triticum aestivum* L.) in relation to major genes and RFLP markers. *Theor. Appl. Genet.* 94:1047-1052.

Bresford W.A. (1990). Direct transdifferentiation: can cells change their phenotype without dividing? *Cell Differ Dev.* 29:81-93.

Binarova P. and Dolezel J. (1993) Effect of anti-microtubular drug amiprophos-methyl on somatic embryogenesis and DNA ploidy levels in alfalfa and carrot cell suspension cultures. *Biology Plantarum* 25:329-339.

Blanckaert A., Belingheri L., Sautiere P.E., Vasseur J., Hilbert J.L. (2002). 9-kDa acidic and basic nsLTP-like proteins are secreted in the culture-medium conditioned

by somatic embryogenesis in *Cichorium*. *Plant Physiology and Biochemistry* 40:339-345.

Blervacq A.S., Dubois T., Dubois J., Vasseur J., (1995). First divisions of somatic embryogenic cells in *Cichorium* hybrid '474'. *Protoplasma* 186:163-168.

Bolibok H. and Rakoczy-Trojanowska M. (2006). Genetic mapping of QTLs for tissue-culture response in plants. *Euphytica* 149:73-83.

Bolibok H., Gruszczyńska A., Hromada-Judycka A., Rakoczy-Trojanowska M. (2007). The identification of QTLs associated with the *in vitro* response of rye (*Secale cereale* L.). *Cell. Mol. Biol.* 12:523-535.

Boyer L.A., Plath K. Zeitlinger J. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441:349-353.

Braybook S.A., Stone S.L., Park S., Bui A.Q., Le BH., Fisher R.L.(2006). Genes directly regulated by LEAFY COTYLEDON 2 provide insight into the control of embryo maturation and somatic embryogenesis. *Proc. Natl. Acad. Sci.* 103:3468-3473.

Brill L.M., Evans C.J., Hirsh A.M. (2001). Expression of MsLEC1-and MsLEC2-antisens genes in alfalfa plant lines causes severe embryogenesis developmental and reproductive abnormalities. *Plant J.* 25:453-461.

Briggs R. and King T.J. (1952). Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. *Proc. Natl. Acad. Sci. USA* 38:455-463.

Burglin T.R., Finney M., Coulson A., Ruoken G. (1989).Caenorhabditis elegans has scores of homeobox-containing genes. *Nature* 341:239-243.

Bush M., Mayer U., Jurgens G. (1996). Molecular analysis of the *Arabidopsis* pattern formation gene *GNOM*: gene structure and intragenic complementation. *Mol. Gen. Genet.* 250:681-691.

Byrne P.F., McMullen M.D., Snook M.E., Musket T.A., Theuri J.M., Widstrom N.W., Wiseman B.R., Coe E.H. (1996). Quantative trait loci metabolic pathways : genetic control of the concentration of maysin, a corn earworm resistance factor, in maize silks. *Proc. Natl. Acad. Sci. USA* 93:8820-8825.

Byrne M.E., Barley R., Curtis M., Arroyo J.M., Dunham M., Hudson A., Martienssen R.A. (2000). *Assymetric leaves1* mediates leaf patterning and stem cell function in *Arabidopsis*. *Nature* 408:967-971.

## C

Cadalen T., Morchen M., Blassiau C., Clabaut A., Sheer I., Hilbert J.L., Quillet M.C. (2009). A consensus genetic map for chicory (*Cichorium intybus* L.), soumis à la revue *Molecular Breeding*.

Chapman A., Blervacq A.S., Vasseur J., Hilbert J.L. (2000). Arabinogalactan-proteins in *Cichorium* somatic embryogenesis : effect of beta-glucosyl Yariv reagent and epitope localisation during embryo development. *Planta* 211:305-314.

Chen F. and Foolad M. R. (1999). Nucellar-cell-specific expression of a lipid transfer protein gene in barley (*Hordeum vulgare* L.). *Plant Cell Report* 18: 445-450.

Chen C.G., Mau S.L., Clarke A.E. (1993). Nucleotide sequence and style-specific expression of a novel praline –rich protein gene from *Nicotiana glauca*. *Plant Mol. Biol.* 21:391-395.

Cheng J.C., Seeley K.A., Goupil P., Sung Z.R. (1996). Expression of *DC8* is associated with, but not dependent on embryogenesis. *Plant Mol. Biol.* 1:127-41.

Chuck G., Lincoln C., Hake S. (1996). *KNAT1* induces lobed leaves with ectopic meristems when overexpressed in *Arabidopsis*. *Plant Cell.* 8:1277-1289.

Corre F., Henry Y., Rode A., Hartmann C. (1996). Em gene expression during somatic embryogenesis in the monocot *Triticum aestivum* L. *Plant Science* 117:139-149.

Costa S., Shaw P. (2006a). 'Open minded' cells: how cells can change fate. *Trends in Cell Biology* 17:101-106.

Costa S., Shaw P. (2006b). Chromatin organization and cell fate switch respond to positional information in *Arabidopsis*. *Nature* 439:493-496.

Cuming A.C., Lane B.G. (1979). Protein synthesis in imbibing wheat embryos. *Eur. J. Biochem.* 99:217-224.

Crea F., Belluci M., Damiani F., Arcioni S. (1995). Genetic control of somatic embryogenesis in alfalfa (*Medicago sativa* L.). *Euphytica* 81:151-155.

Cresti M., Blackmore S., Van Went J.L. (1992). Atlas of sexual reproduction in flowering plants. Berlin and New York: Springer-Verlag.

**A. Cyr R.J., Bustos M.M., Gultinan M.J., Fosket D.E. (1987). Developmental modulation of tubulin protein and mRNA levels during somatic embryogenesis in cultured carrot cells. *Planta* 171: 365-376.**

## D

Da Silva (2004). Identification des gènes *CiSTM* et *CICLVP1p* chez *Cichorium intybus*. Implication potentielle dans les modifications d'identité cellulaire au cours des phases précoces de l'embryogenèse somatique et du développement des embryons zygotiques et de la graine. Thèse à l'université de Lille1.

## Références bibliographiques

Darvasi A., Weinreb A., Minke V., Weller J.I., Soller M. (1993). Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetic* 134: 943-951.

Davletova S., Meszaros T., Miskolczi P., Oberschall A., Torok K., Magyar Z., Dudits D., Deak M. (2001). Auxin and heat shock activation of a novel member of the calmodulin like domain protein kinase gene family in cultured alfalfa cells. *J. Exp. Bot.* 52:215-221.

De Buyser J., Marcotte J.L., Henry Y. (1992). Genetic analysis of *in vitro* wheat somatic embryogenesis. *Euphytica* 63:265-270.

Dean G., Casson S., Lindsey K. (2004). *KNAT6* gene of *Arabidopsis* is expressed in roots and is required for correct lateral root formation. *Plant Mol. Biol.* 54:71-84.

Decout E. Dubois T. Guerida M., Dubois J., Audran J.C., Vasseur J. (1994). Role of temperature as a triggering signal for organogenesis or somatic embryogenesis in wounded leaves of chicory cultured *in vitro*. *Journal of Experimental Botany* 45:1859-1865.

De Jong A.J., Cordewener J., Lo S.F., Terzi M., Vandekerckhove J., van Kammen A., de Vries S.C. (1992). A carrot somatic embryo mutant is rescued by chitinase. *Plant Cell* 4:425-433.

De Jong A.J., Schmidt E.D.L., de Vries S.C. (1993). Early events in higher-plant embryogenesis. *Plant Mol. Biol.* 22:367-377.

De jong A.J., Hendriks T., Meijer E.A., Penning M., Loschiavo F., Terzi M., Van Kammen A., De vries S.C. (1995). Transient reduction in secreted 32 kD chitinase prevents somatic embryogenesis in the carrot (*Daucus carota L.*) variant *ts11*. *Dev. Genet.* 16:332-343.

Demura T., Tashiro G., Horiguchi G., Kishimoto N., Kubo M., Matsuoka N., Minami A., Nagata-Hiwatashi M., Nakamura K., Okamura Y., Sassa N., Suzuki S., Yazaki J.,

## Références bibliographiques

Kikuchi S., Fukuda H. (2002). Visualization by comprehensive microarray analysis gene expression programs during transdifferentiation of mesophyll cells into xylem cells. *Proc Natl Acad Sci USA* 99:15794-15799.

de Simone M., Morgante M., Lucchin M., Parrini P., Marocco A (1997). A first linkage map of *Cichorium intybus* L. Using a one-way pseudo-testcross and PCR-derived markers. *Mol Breed* 3:415-425.

de Vries S.C., Booij H., Meyernink P., Huisman G., Wilde H.D., Thomas T.L., van Kammen A. (1988). Acquisition of embryogenic potential in carrot cell-suspension cultures. *Planta*. 176:196-204.

Diatchenko L., Lau Y.F., Campbell A.P., Chenchik A., Moqadam F., Huang B., Lukyanov S., Lukyanov K., Gurskaya N., Sverdlov E.D., Siebert P.D. (1996). Suppression subtractive hybridisation: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* 93:6025-6030.

Dockx J., Quaedvlieg N., Keultjes G., Kock P., Weisbeek P., Smeekens S. (1995). The homeobox gene *ATK1* of *Arabidopsis thaliana* is expressed in the shoot apex of the seedling and in floxers and inflorescence stems of mature plants. *Plant. Mol. Biol.* 28:723-737.

Domoki M., Gyorgyey J., Biro J., Pasternak T.P., Zvara A., Bottka S., Puskas L.G., Dudits D., Feher A. (2006). The identification and characterization of genes associated with the induction of embryogenic competence in leaf-protoplast-derived alfalfa cells. *Biochim. Biophys. Acta*. 1759:543-551.

Donj J.Z. and Dunstan D.I. (1996b). Characterization of three heat-shock-protein genes and their development regulation during somatic embryogenesis in white spruce (*Picea Glauca*). *Planta* 200:85-91.

Donj J.Z. and Dunstan D.I. (1997a). Characterization of cDNAs representing five abscisic acid-responsive genes associated with somatic embryogenesis in *Picea*

*glauca*, and their responses their responses to abscisic acid stereostructure. *Planta* 203:448-453.

Donj J.Z. and Dunstan D.I. (1997b). Endochitinase and  $\beta$ -1,3-glucanase genes are developmentally regulated during somatic embryogenesis in *Picea glauca*. *Planta* 201:189-194.

Donj, J.Z. and Dunstan D.I. (1999). Cloning and characterization of six embryogenesis-associated cDNA from embryos of *Picea glauca* and their comparative expression during zygotic embryogenesis. *Plant. Mol. Biol.* 39:859-864.

D'Onofrio C.D., Morini S., Bellocchi G. (1998). Effect of light quality on somatic embryogenesis in quince leaves. *Plant Cell Tiss. Org. Cult.* 53:91-98.

Domoki I.M., Gyorgey J., Biro J., Pasternak T. P., Zvara A., Bottka S., Puska L. G., Dudits D., Feher A. (2006). Identification and characterization of genes associated with the induction of embryogenic competence in leaf-protoplast-derived alfalfa cells. *Biochimica et biophysica acta* 1759 :543-551

Dos Santos A.V.P., Cutter E.G., Davey M.R. (1983). Origin and development of somatic embryos in *Medicago sativa* L. *Protoplasma* 117:107-115.

Duban M. (2004). Clonage et caractérisation de deux genes codant des récepteurs transmembranaires à activité kinase chez *Cichorium intybus* L. Expression au cours des phases précoces de l'embryogenèse somatique et du développement de la graine. Thèse de l'Université de Lille 1.

Dubois T., Dubois J., Guedira M., Vasseur J. (1988). Direct somatic embryogenesis on styles of *Cichorium*: temperature effects and origin of embryoids. *Compte rendus de l'Académie des sciences de Paris, Série III*, 337:669-75.

Dubois T., Guedira M., Dubois J., Vasseur J. (1990). Direct somatic embryogenesis in roots of *Cichorium*: Is callose an early marker ?. *Ann Bot.* 65:539-545.

Dubois T., Guedira M., Dubois J., Vasseur J. (1991). Direct somatic embryogenesis in leaves of *Cichorium*: a histological and SEM study of early stages. *Protoplasma* 162:120-7.

Dudits D., Bogre L., Györgyey J. (1991). Molecular and cellular approaches to the analysis of plant embryo development from somatic cells in vitro. *J. Cell. Sci.* 99:475-484.

Dulieu H. (1991). Inheritance of regeneration capacity in the genus *Petunia*. *Euphytica* 53:173-181.

Dupire L. (1995). Contribution à l'étude des stades précoces de l'embryogenèse somatique chez le *Cichorium* hybride '474'. Corrélation entre l'évolution de 'dehydrins-like' protéines et la première division des cellules induites. Master recherche à l'université de Lille 1.

Dupire L., Decout E., Vasseur J., Delbreil B. (1999). Histological and 2-D protein patterns comparisons between a wild type and a somatic embryogenic mutant of *Asparagus officinalis* L. *Plant Sci.* 147:9-17

Dresselhaus T., Cordts S., Heuer S., Sauter M., Lorz H., Kranz E. (1999). Novel ribosomal genes from maize are differentially expressed in the zygotic and somatic cell cycles. *Mol Gen Genet* 261:416-427.

## E

Egertsdotter U. and von Arnold S. (1995). Importance of arabinogalactan proteins for the development of somatic embryos of Norway spruce (*Picea abies*). *Physiol. Plant* 93:334-345.

Endrizzi K., Moussian B., Haecker A., Levin J.Z., Laux T. (1996). The *SHOOT MERISTEMLESS* gene is required for maintenance of undifferentiated cells in

Arabidopsis shoots and floral meristems and acts at a different regulatory level than the meristem genes *WUSCHEL* and *ZWILLE*. *Plant Journal* 10:967-979.

## F

Fehér A. (2006). Why somatic plant cells start to form ES? In mujid A., Samaj J. eds, Somatic Embryogenesis. *Plant Cell Monographs*, Robinson D.G., series ed., Springer-Verlag, Berlin Heidelberg, Germany, 2:85-101.

Fehér A. (2008). The initiation phase of somatic embryogenesis: what we know and what we don't. *Acta Biologica* 52:53-56.

Fellers J.P., Geunzi A.C., Porter D.R. (1997). Marker proteins associated with somatic embryogenesis of wheat callus cultures. *J Plant Physiol.* 151:201-208.

Filipeki M.K., Sommer H., Malepszy S. (1997). The MADS-box gene *CUS1* is expressed during cucumber somatic embryogenesis. *Plant Sci.* 125:63-74.

Fisherova L., Fisher L., Vondrakova Z., Vagner M. (2008). Expression of the gene transcription factor PaVP1 differs in *Picea abies* embryogenic lines depending on their ability to develop somatic embryos. *Plant Cell Rep.* 27:435-441.

Florres Berrios E., Gentzbittel L., Mokrani G., Alibert G., Sarrafi A. (2000a). Genetic control of early events in protoplast division and regeneration pathway in sunflower. *Theor. Appl. Genet.* 101: 606-612

Flores Berrios E., Sarrafi A., Fabre F., Alibert G., Gentzbittel L. (2000b). Genotypic variation and chromosomal location of QTL for somatic embryogenesis revealed by epidermal layers culture of recombinant inbred lines in the sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 101:1307-1312

Footit S., Ingouff M., Clapham D., von Arnold S. (2003). Expression of the viviparous (Pavp1) and p34cdc2 protein kinase (cdc2Pa) genes during somatic embryogenesis in Norway spruce (*Picea abies* L. Karst). *J. Exp. Bot.* 54:1711-1719.

Franz G., Hatzopoulos P., Jones T.J., Krauss M., Sung Z.R. (1989). Molecular and genetic analysis of an embryogenic gene, DC8, from *Daucus carota* L. *Mol Gen Genet.* 218:143-151.

## G

Gaj M.D. (2002). Stimulation of somatic embryo formation by mutagens and darkness in culture of immature zygotic embryos of *Arabidopsis thaliana* (L.) Heynh. *Plant Growth Regul.* 37:93-98.

Galland R., Randoux B., Vasseur J., Hilbert J.L. (2001). A glutathione S-transferase cDNA identified by mRNA differential display is upregulated during somatic embryogenesis in *Cichorium*. *Biochim. Biophys. Acta* 1522:212-216.

Gaspar Y., Jonhson K.L., McKenna J.A., Bacic A., Schultz C. (2001). The complex structures of arabinogalactan-proteins and the journey towards understanding functions. *Plant mol. Biol.* 47 :161-176.

Guedira M., Dubois-Tylski T., Vasseur J., Dubois J. (1989). Embryogenèse somatique directe à partir de culture d'anthères de *Cichorium* (Asteraceae). *Can. J. Bot.* 67 :970-976.

Giroux R.W. and Pauls K.P. (1997). Characterization of somatic embryogenesis-related cDNAs from alfalfa (*Medicago sativa* L.). *Plant Mol. Biol.* 33:393-404.

Grabowska A., Wisniewska A., Tagashira N., Malepsy S., Filipecki M. (2009). Characterization of CsSEF1 gene encoding putative CCCH-type zinc finger protein expressed during cucumber somatic embryogenesis. *J. Plant Physiol.* 166:310-323.

Grafi G. (2004). How cells dedifferentiate: a lesson from plants. *Developmental biology* 268:1-6.

Grattapaglia D. and Sederoff R. (1994). Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137: 1121-1137.

Grimault V., Helleboid S., Vasseur J., Hilbert J.L. (2007). Co-localization of  $\beta$ -1,3 glucanases and callose during somatic embryogenesis in *Cichorium*. *Plant Signalling and Behavior* 2:6, 455-461.

Györgyey J., Gartner A., Németh K., Magyar Z., Hirt H., Heberle-Bors E., Dudits D.(1991). Alfalfa heat shock genes are differentially expressed during somatic embryogenesis. *Plant Mol. Biol.* 6:999-1007.

## H

Haiyan J., Dalong Y., Jie Y., Shulin X., Yang X., Caiquin Z et al., (2007). Mapping QTL for tissue culture response of mature wheat embryos. *Mol Cells* 23:323-330.

Hardke C.S. and Berleth T. (1998). The Arabidopsis gene *MONOPTEROS* encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO Journal* 17:1405-1411.

He P., Sheng L., Lu C., Chen Y., Zhu L. (1998). Analysis of quantitative trait loci, which contribute to anther culturability in rice (*Oryza sativa* L.). *Molecular Breeding* 4:165-172.

Hecht V., Vieille-Calzada J.P., Hartog M.V., Schmidt E.D., Boutilier K., Grossniklaus U., de Vries S.C. (2001). The Arabidopsis *SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1* gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. *Plant Physiol.* 127:803-816.

## Références bibliographiques

Helleboid S., Bauw G., Belingheri L., Vasseur J., Hilbert J.L. (1998). Extracellular beta-1,3-glucanases are induced during early somatic embryogenesis in *Cichorium*. *Planta* 205:56-63.

Helleboid S., Chapman A., Hendriks T., Inze D., Vasseur J., Hilbert J.L.(2000). Cloning of beta-1,3-glucanases expressed during *Cichorium* somatic embryogenesis.

Heller R. (1953). Researches on the mineral nutrition of plant tissues. *Am. Soc. Nat. Bot. Veg.* 14:223-246.

Hendriks T., Meijer E.A., Thoma S., Kader J.C., De Vries S.C. (1994) The carrot extracellular lipid transfer protein EP2: Quantitative aspects with respect to its putative role in cutin synthesis. In: NATO ASI Series, Vol H81, *Plant molecular biology*, Coruzzi G, Puigdomenech P, eds. Springer-Verlag, Berlin, Heidelberg, pp 85-94.

Hendriks T., Scheer I., Quillet M.C., Randoux B., Delbreil B., Vasseur J., Hilbert J.L. (1998). A nonsymbiotic hemoglobin gene is expressed during somatic embryogenesis in *Cichorium*. *Biochim. Biophys. Acta* 1443:193-197

Hernandez-Fernandez M.M and Christie B.R. (1989). Inheritance of somatic embryogenesis in alfalfa (*Medicago sativa* L.). *Genome* 32:318-321.

Heuer S., Hansen S., Bantin J., Brettschneider R., Kranz E., Lorz H., Dresselhaus T. (2001). The maize MADS box gene ZmMADS3 affect nod number and spikelet development and is co-expressed with ZmMADS1 during flower development, in egg cells, and early embryogenesis. *Plant Physiology* 127:33-45.

Hilbert J.L., Dubois T., Vasseur J. (1992). Characterization of embryogenesis related proteins and their use as markers during somatic embryo formation in *Cichorium*. *Plant Physiology Biochemistry* 30:733-741.

Hirt H., Pay A., Gyorgyey J., Bako L., Nemeth K., Bogre L., Schweyen R.J., Heberle-Bors E., Dudits D. (1991). Complementation of a yeast cell cycle mutant by an alfalfa

cDNA encoding a protein kinase homologous to p34cdc2. *Proc. Natl. Acad. Sci. U.S.A.* 88:1636-1640.

Higashi K., Shiota H., Kamada H. (1998). Patterns of expression of the genes for glutamine synthetase isoforms during somatic and zygotic embryogenesis in carrot. *Plant and Cell Physiology* 39:418-424.

Holk A., Kaldenhoff R., Ritcher G. (1996). Regulation of an embryonic carrot gene (DC 2.15) and identification of its active promoter sites. *Plant Mol. Biol.* 31:1153-1161.

Holme I.B. (2004). Quantative trait loci affecting plant regeneration from protoplasts of *Brassica oleracea*. *Theor. and Appl. Genet.* 108:1513-1520.

Huang X., Nabipour A., Gentzbittel L., Sarrafi A. (2007). Somatic embryogenesis from thin epidermal layers in sunflower and chromosol regions controlling the response. *Plant Science* 173:247-252.

Hundertmark M., Hinch D. (2008). LEA (Late Embryogenesis Abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC genomics* 9:118.

## J

Jang C. S., Kim D. S., Bu S. Y., Kim J. B., Lee S. S., Kim J. Y., Johnson J. W., Seo Y. W. (2002). Isolation and characterization of lipid transfer protein (LTP) genes from a wheat-rye translocationline. *Plant Cell Report* 20:961-966.

Jansen R.C. (1993). Interval mapping of multiple quantative trait loci. *Genetics* 135:205-211.

Jansen R.C. and Stam P. (1994). High resolution of quantative traits into multiple loci via interval mappig. *Genetics* 158:925-926.

Jimenez V. and Bangerth F. (2001). *In vitro* culture and endogenous hormone levels in immature zygotic embryos, endosperm and callus cultures of normal and high-lysine barley genotypes. *J. Appl. Bot.* 75:1-7.

Joosen R., Cordewer J., Supena E.D.J., Vorst O., Lammers M., Maliepaard C., Zeilmaker T., Miki B., America T., Custers J., Boutillier K. (2007). Combined transcriptome and proteome analysis identifies pathways and markers associated with the establishment of rapessed microspore-derived embryo development. *Plant Phys.* 144:155-172.

## K

Kader J.C. (1996). Lipid Transfer Proteins in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47:627-654.

Kader, J.-P. (1997). Lipid-transfer proteins: a puzzling family of plant proteins. *Trends Plant Sci.* 2:66–70.

Kao C.H., Zeng Z.B., Teasdale R.D. (1999). Multiple interval mapping for quantitative trait loci. *Genetics* 152:1203-1216.

Kapros T., Bogre L., Nemeth K., Bak L., Gyorgyey J., Wu S. C., Dudits D. (1992). Differential expression of histone H3 gene variants during cell cycle and somatic embryogenesis in alfalfa. *Plant Physiology* 98:621-625.

Kawahara R., Sunabori S, Fukuda H, Komamine A. (1992). A gene expressed preferentially in the globular stage of somatic embryogenesis encodes elongation-factor 1 alpha in carrot. *Eur. J. Biochem.* 209:157-162.

Kawahara R., Komamine A., Fukuda H. (1995). Isolation and characterization of homeobox-containing genes of carrot. *Plant Mol. Biol.* 27:155-164.

Kempin S.A., Liljegren S.J., Block L.M., Rounsley S.D., Yanosky M.F., Lam E. (1997). Targeted disruption in Arabidopsis. *Nature* 389: 802-803.

Kennedy D. and Norman C. (2005). What don't we know? *Science* 309:75

Kidner C.A. and Martienssen R.A. (2005). The role of argonaute 1 (AGO1) in meristem formation and identity. *Dev. Biol.* 280:504-517.

Kielly G.A. and Bowley S.R. (1992). Genetic control of somatic embryogenesis in alfalfa. *Genome* 35:474-477.

**Kim J.B., Sebastiano V., Wu G., Araúzo-Bravo M.J., Sasse P., Gentile L, Ko K., Ruau D., Ehrich M., van den Boom D., Meyer J., Hübner K., Bernemann C., Ortmeier C., Zenke M., Fleischmann B.K., Zaehres H., Schöler H.R. (2009).** Oct4-induced pluripotency in adult neural stem cells. *Cell* 136:411-419.

Kitamiya E., Susuki S., Sano T., Nagata T. (2000). Isolation of two genes that were induced upon the initiation of somatic embryogenesis on carrot hypocotyls by high concentrations of 2,4-D. *Plant Cell Rep.* 19:551-557.

Kiyosue T., Yamaguchi-Shinozaki K., Shinozaki K., Kamada H., Harada H. (1998). A chromo box gene from carrot (*Daucus carota* L.): its cDNA structure and expression during somatic and zygotic embryogenesis. *Biochim. Biophys. Acta* 1398:42-46.

Ko S., Tan S.K., Kamada H. (2006). Characterization of a dehydrin-like phosphoprotein (ECPP44) relating to somatic embryogenesis in carrot. *Plant physiol. Bioch.*

Koorneef M., Bade J., Hanhart C., Horsman K., Schel J., Soppe W., Verkerk R., Zabel P. (1993). Characterization and mapping of a gene controlling shoot regeneration in tomato. *Plant Journal* 3:131-141.

Kragh K.M., Hendriks T., Dejong A.J., Loschiavo F., Bucherna N., Hojrup P., Mikkelsen J.D., de Vries S.C. (1996). Characterization of chitinases able to rescue somatic embryos of the temperature-sensitive carrot variant *ts11*. *Plant Mol. Biol.* 31:631-645.

Krakowsky M.D., Lee M., Garay L., Woodman-Clikeman W., Long M.J., Sharopova N., Frame B., Wang K. (2006). Quantitative trait loci for callus initiation and totipotency in maize (*Zea mays* L.). *Theoretical and Applied Genetic* 113:821-830.

Kreuger and Van Holst (1996). Arabinogalactan proteins and plant differentiation. *Plant Mol. Biol.* 30:1077-1086.

***B. Kumar Sharma S., Steve Millam S., Ingo Hein I., Bryan G.J. (2008). Cloning and molecular characterisation of a potato SERK gene transcriptionally induced during initiation of somatic embryogenesis. Planta 228:319-330.***

## L

Lander E.S. and Botstein D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185-199.

Lee T.I., Jenner R.G., Boyer L.A. (2006). Control of developmental regulators by polycomb in human embryonic stem cells. *Cell* 125:301-313.

Legrand (2006). Etude transcriptomique au cours des phases précoces de l'embryogenèse somatique chez la chicorée. Thèse à l'université de Lille1.

Legrand S., Hendriks T., Hilbert J.L., Quillet M.C. (2007). Characterization of expressed sequence tags obtained by SSH during somatic embryogenesis in *Cichorium intybus* L. *BMC Plant Biol.* 7:27.

Lehmann E.L. (1975). Nonparametrics. McGraw-Hill, New York.

Li J. and Chory J. (1997). A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* 90: 929–938

## Références bibliographiques

- Li J. (2003). Brassinosteroids signal through two receptor-like kinases. *Curr. Opin. Plant Biol.* 6:494–499
- Li G., Hall T., Holmes-Davis R. (2002). Plant chromatin: development and genetic control. *Bioessays* 24:234-243.
- Liang P. (2002). A decade of differential display. *BioTechniques* 33:338-346.
- Liang P. and Pardee A.B. (1992). Differential display of eukaryotic mRNA by means of the polymerase chain reaction. *Science* 257:967-971.
- Liew M., Pryor R., Palais R., Meadows C., Erali M., Lyon E., Wittwer C. (2004). Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. *Clinical Chemistry* 50:7,1156-1164.
- Lincoln C., Long J., Yamagushi J., Serikawa K., Hake S. (1994). A *knotted1*-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *Plant Cell* 6:1859-1876.
- Lindzen E. and Choi J.H. (1995). A carrot cDNA encoding an atypical protein kinase homologous to plant calcium-dependent protein kinases. *Plant Mol. Biol.* 28:785-797.
- Liu Q., Yao X., Pi L., Wang H., Xiaofeng C., Huang H. (2006). The argonaute gene modulates shoot apical meristem maintenance and establishment of leaf polarity by repressing miR165/166 in *Arabidopsis*. *The Plant Journal* 58:27-40.
- Long J.A., Moan E.I., Medford J.I., Barton M.K. (1996). A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature* 379:66-69.

Lorieux M., Perrier X., Goffinet B., Lanaud C., Gonzales de Leon D (1995). Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F<sub>2</sub> populations. *Theor. Appl. Genet.* 90:81-89.

Lucau-Danila A., Laborde L., Legrand S., Huot L., Hot D., Lemoine Y., Hilbert J., Hawkins S., Quillet M.C., Hendriks T., Blervacq A.S. The  $\beta$ -D-1,3 Glucosyl Yariv reagent affects the differential expression of certain genes during somatic embryogenesis in leaf explants of *Cichorium intybus* L., soumis à la revue *Planta*.

Lynn K., Fernandez A., Aida M., Sedbrook J., Tasaka M., Masson P., Barton K. (1999). The *PINHEAD/ZWILLE* gene acts pleiotropically in Arabidopsis development and has overlapping functions with the *ARGONAUTE1* gene. *Development* 126:469-481.

## M

Ma H., McMuller M.D., Finer J.J. (1994). Identification of a homeobox containing gene with enhanced expression during soybean (*Glycine max* L.) somatic embryos development. *Plant Mol. Biol.* 24:465-473.

Maheswaran G. and Williams E.G. (1985). Origin and development of somatic embryos formed directly on immature embryos of *Trifolium repens in vitro*. *Ann Bot* 56:619-630.

Mano Y., Takahashi H., Sato K., Takeda K. (1996). Mapping genes for callus growth and shoot regeneration in barley. *Breeding science* 46:137-142.

Mano Y., Komatsuda T. (2002). Identification of QTLs controlling tissue-culture traits in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetic* 105:708-715.

Mantiri F.R., Kurdyukov S., Lohar D.P., Sharapova N., Saeed N.A., Wang X.D., Vandenbosh K.A., Rose R.J. (2008). The transcription factor MtSERF1 of the ERF

subfamily identified by transcriptional profiling is required for somatic embryogenesis induced by auxin plus cytokinin in *Medicago trunculata*.

Mau S.L., Chen C.G., Pu Z.Y., Moritz R.L., Simpson R.J., Bacic A., Clarke A.E. (1995). Molecular cloning of cDNAs encoding the protein backbones of arabinogalactan-protein from the filtrate of suspension-cultures cells of *Pyrus communis* and *Nicotiana glauca*. *Plant J.* 8:269-281.

Mayer U., Buttner G., Jurgens G. (1993). Apical-basal pattern formation in the *Arabidopsis* embryo – Studies on the role of the *Gnom* gene. *Development* 117:149-162.

McCabe P.F., Valentine T.A., Forsberg L.S., Pennell R.I. (1997). Soluble signals from cells identified at the cell wall establish a developmental pathway in carrot. *Plant Cell* 9:2225-2241.

McLean N.L. and Nowak J. (1998). Inheritance of somatic embryogenesis in red clover (*Trifolium pratense* L.). *Theoretical and Applied Genetics* 97:557-562.

Meijer A.H., Scarpella G., van Dijk E.L., Qin L., Taal A.J.C., Ruebs S., Harrington S.E., McCough S.R., Schilperoort R.A., Hoge J.H.C (1997). Transcriptional repression by oshox1 a novel homeodomain leucine zipper protein from rice. *Plant Journal* 11:263-276.

Meinke D.W. (1992). A homoerotic mutant of *Arabidopsis thaliana* with leafy cotyledons. *Science* 258:1647-1650.

Mordhost A.P., Toonen M.A.J., de Vries S.C. (1997). Plant embryogenesis. *Crit. Rev. Plant Sci.* 16:535-576.

Mordhost A.P., Voerman K.J., Hartog M.W., Meijer E.A., van Went J., Koorneef M., de Vries S.C. (1998). Somatic embryogenesis in *Arabidopsis thaliana* is facilitated by mutations in genes repressing meristematic cell divisions. *Genetics* 149:549-563.

## Références bibliographiques

Mundy J., Yamaguchi-Shinozaki K., Chua N-H. (1990). Nuclear proteins bind conserved elements in the abscisic acid responsive promoter of a rice *rab* gene. *Acad. Sci. USA* 87:1406-1410.

Murashige T., Skoog F., (1962). A revised medium for rapid growth and bioassays with tobacco tissue culture. *Physiol. Plant* 80: 91-98.

Murigneux A., Bentollila S., Hardy T., Baud S., Guitton C., Julien H., Ben Tahar S., Freyssinet G., Beckert M. (1994). Genomic variation of quantitative trait loci controlling in vitro androgenesis in maize. *Genome* 37:970-976.

## N

Nadolska-Orczyk A. and Malepszy S. (1989). In vitro culture of *Cucumis sativus* L. 7. Genes controlling plant regeneration. *Theor. Appl. Genet.* 78:836-840.

Neutelings G., Domon J.M., Membré N., Bernier F., Meyer Y., David A., David H. (1998). Characterization of a germin-like protein gene expressed in somatic and zygotic embryos of pine (*Pinus caribaea* Morelet). *Plant Mol. Biol.* 38:1179-1190.

Nishimura A., Ashikari M., Lin S., Takashi T., Angeles E.R., Yamamoto T., Matsuoka M. (2005). Isolation of a rice regeneration quantitative trait loci gene and its application to transformation systems. *Proc. Natl. Acad. Sci. USA* 102: 11940-11944.

Nishiwaki M., Fujino K., Koda Y., Masuda K., Kikuta Y. (2000). Somatic embryogenesis induced by the simple application of abscisic acid carrot (*Daucus carota* L.) seedlings in culture. *Planta* 211:756-759.

## O

Oetting WS., Lee H.K., Flanders D.J. (1995). Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics* 30:450-458.

## Références bibliographiques

Ogas J., Cheng J.C., Sung Z.R., Somerville C. (1997). Cellular Differentiation Regulated by Gibberellin in the *Arabidopsis thaliana* pickle Mutant. *Science* 277, 91–94.

Ogas J., Kaufmann S., Henderson J., Somerville C. (1999). PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. *Proc Natl Acad Sci USA* 96:13839-13844.

Ori N., Eshed Y., Chuck G., Bowman J., Hake S. (2000). Mechanisms that control KNOX gene expression in *Arabidopsis* shoot. *Development* 127:5523-5532.

Orita M., Iwahana H., Kanazawa H., Hayashi K., Sekiya T. (1989). Detection of polymorphism of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* 86:2766-2770.

Nadolska-Orczyk A., Malepszy S. (1989). *In vitro* culture of *cucumis sativus* L. *Theoretical and Applied Genetic* 78:836-840.

## P

Pautot V., Dockx J., Hamant O., Kronenberg J., Grand-Jean O, Jublot D., Traas J. (2001). *KNAT2*: evidence for a link between *knotted*-like genes and carpel development. *Plant Cell* 13:1719-1734.

Pearce R.S., Houlston C.E., Atherton K.M., Rixon J.E., Harrison P., Hughes M.A., Dun M.A. (1998). Localization of expression of three cold-induced genes, *blt101*, *blt4.9*, and *blt14*, in different tissues of the crown and developing leaves of cold-acclimated cultivated barley. *Plant Physiology* 117: 787-795.

Pedroso M. C. and Pais M. S. (1995). Factors controlling somatic embryogenesis. *Plant Cell Tissue Org. Cult.* 43:147–154.

Pelese-Siebenbourg F., Caelles C., Kader J.C., Delseny M., Puigdomenech P. (1994). A pair of genes coding for lipid transfer proteins in *Sorghum vulgare*. *Gene* 148:305-308.

Perry S.E., Lehti M.D., Fernandez D.E. (1999). The MADS-domain protein AGAMOUS-like 15 accumulates in embryonic tissues with diverse origins. *Plant Physiology*. 120:121-130.

Puupponen-Pimia R., Saloheim M., Vasara T., Ra R., Gaugecz J., Kurtén J, Knowles J., Keränen S., Kauppinen V. (1993). Characterization of birch (*Betula pendula* Roth) embryogenic gene, BP8. *Plant Mol. Biol.* 23:423-428.

## R

Rakoczy-Trojanowska M. and Malepszy S. (1993). Genetic factors influencing regeneration ability in rye (*Secale cereale* L. ). I. Immature inflorescences. *Theor. Appl. Genet.* 86: 406–410.

Randoux B., Quillet M.C., Rambaud C., Vasseur J., Hilbert J.L. (2002). Identification of cDNA encoding Rab-related GTP-binding proteins expressed during somatic embryogenesis in *Cichorium*. *Plant Science* 162:413-422.

Reed GH. and Wittwer CT. (2004). Sensitivity and specificity of single-nucleotide polymorphism scanning by high-resolution melting analysis. *Clin Chem* 50:1748-1754.

Reghavan V. (1986). Embryogenesis in Angiosperms. Cambridge University Press, Cambridge, UK.

Reich B. and Bingham E.T. (1980). Genetic control of bud formation from callus culture of diploid *alfalfa*. *Plant Sci. Lett.* 20:71-77.

Reinhard B.J. and Bartel D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. *Science* 297, 1831.

Rothschild M.F. and Soller M. (1997). Candidate gene analysis to detect gene controlling trait of economic important in domestic livestock. *Probe* 8: 13-20.

## S

Sablowski R. (2004). Plant and animal stem cells: conceptually similar, molecularly distinct? *Trends in cell biol.*14:

Sablowski R. (2007). The dynamic plant stem cell niches. *Curr. Opin. Plant Biol.* 10:639-644.

Sagare A.P., Lee Y.L., Lin T.C., Chen C.C., Tsay H.S. (2000). Cytokinin-induced somatic embryogenesis and plant regeneration in *Corydalis yanhusuo* (Fumariaceae), a medicinal plant. *Plant Sci.* 160:139-147.

Salaj J., von Recklinghausen I., Hecht V., de Vries S., Schel J., van Lammeren A. (2008). *AtSERK1* expression precedes and coincides with early somatic embryogenesis in *Arabidopsis thaliana*. *Plant Physiology and Biochemistry* 46:709-714.

Sato S., Toya T., Kawahara R., Whittier R. F., Fukuda H, Komamine A. (1995). Isolation of a carrot gene expressed specifically during early-stage somatic embryogenesis. *Plant Mol. Biol.* 28:39-46.

Sax K. (1923). The association of the size difference with seed coat pattern and pigmentation *Phaseolus vulgaris*. *Genetics* 8:552-560.

Schiantarelli E., De la Pena., Candela M. (2001). Use of recombinant inbred lines (RILs) to identify, locate and map major genes and quantitative trait loci involved with in vitro regeneration ability in *Arabidopsis thaliana*. *Theor. Appl. Genet.* 102:335-341.

## Références bibliographiques

Schrack K., Mayer U., Horrichs A., Kuhnt C., Bellini C., Dangl J., Schmidt J., Jurgens G. (2000). FACKEL is a sterol C-14 reductase required for organized cell division and expansion in *Arabidopsis* embryogenesis. *Genes Dev* 14: 1471-84.

Schmidt E.D., Guzzo F., Toonen M.A., de Vries S.C. (1997). A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. *Development* 124:2049-2062.

Scofield S. and Murray J.A.H. (2006). *KNOX* gene function in plant stem cells niches. *Plant mol. Biol.* 60:929-946.

Schultz C., Gilson P., Oxley D., Youl J., Bacic A. (1998). GPI-anchors on arabinogalactan-proteins: implications for signalling in plants. *Trends Plant Sci.* 3:426-431.

Semiarti E., Ueno Y., Tsukaya H., Iwakawa H., Machida C., Machida Y. (2001). The *ASYMMETRIC LEAVES2* gene of *Arabidopsis thaliana* regulates formation of a symmetric lamina, establishment of venation and repression of meristem-related homeobox genes in leaves. *Development* 128:1778-1783.

Serpe M.D. and Nothnagel E.A. (1999). Arabinogalactan-proteins in the multiple domains of the plant cell surface. *Adv. Bot. Res.* 30:207-289.

Shah K., Gadella T.W., Jr., van Erp H., Hecht V., de Vries S.C. (2001a). Subcellular localization and oligomerization of the *Arabidopsis thaliana* SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 PROTEIN. *J. Mol. Biol.* 309:641-655.

Sharma S.K., Millam S., Hein I., Bryan G.J. (2008). Cloning and molecular characterization of a potato SERK gene transcriptionally induced during initiation of somatic embryogenesis. *Planta.* 228:319-330.

Shevell D.E., Leu W.M., Gillmor C.S., Xia G.X., Feldmann K.A., Chua N.H. (1994). EMB30 is essential for normal division, cell expansion, and cell adhesion in *Arabidopsis* and encodes a protein that has similarity to sec7. *Cell* 77:1051-1062.

## Références bibliographiques

- Shiota H., Satoh R., Watabe K., Harada H., Kamada H. (1998). C-ABI3, the carrot homologue of the Arabidopsis ABI3, is expressed during both zygotic and somatic embryogenesis and functions in the regulation of embryo-specific ABA-inducible genes. *Plant and Cell Physiology* 39:1184-1193.
- Song J.J., Hannon G.J., Joshua-Tor L. (2004). Crystal structure of AGO. *Science* 305:1434-1437.
- Sossountzov L., Ruiz-Avial L., Vignols F. (1991). Spatial and temporal expression of a maize lipid transfer protein gene. *Plant Cell* 3: 923–933.
- Showalter A.M. (2001). Arabinogalactan-proteins: structure, expression and function. *Cellular and molecular life science* 58:1399-1417.
- Stam P (1993). Constuction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant Journal* 3:739-744.
- Staxén I , Klimaszewska K., Bornman C.H. (1994). Microtubular organization in protoplasts and cells of somatic embryo-regenerating and non-regenerating cultures of *Larix*. *Physiologia Plantarum* 91:680-686.
- Sterk P., Booij H., Schellekens G. A., van Kammen A., de Vries S.C. (1991). Cell-specific expression of the carrot EP2 lipid transfert protein gene. *Plant Cell* 3:907-921.
- Steinmann T., Geldner N., Grebe M., Mangold S., Jackson C. L., Paris S., Galweiler L., Palme K., Jurgens G. (1999). Coordinated polar localization of auxin efflux carrier PIN1 by GNOM ARF GEF. *Science* 286:316-8.
- Stone S.L., Kwong L.W., Yee K.M., Pelletier J., Lepiniec L., Fisher R.L. (2001). *LEAFY COTYLEDON 2* encodes a B3 domain transcription factor that induces embryo development. *Proc. Natl. Acad. Sci.* 98:11806-11811.

## Références bibliographiques

Stone S.L., Braybrook S.A., Paula S.A., Kwong L.W., Meuser J., Pelletier J. (2008). *Arabidopsis* *LEAFY COTYLEDON 2* induces maturation traits and auxin activity: Implications for somatic embryogenesis. *Proc. Natl. Acad. Sci.* 105:3151-3156

Sung Z.R. and Okimoto R. (1981). Embryonic proteins in somatic embryos of carrot. *Proc Natl Acad Sci USA.* 78:3683-3687.

Sung Z.R. and Okimoto R. (1983). Coordinate gene expression during somatic embryogenesis in carrots. *Proc Natl Acad Sci USA.* 80:2661-2665.

## T

Taguchi-Shiobara F., Lin S.Y., Tanno K., Komatsuda T., Yano M., Sasaki T., Oka S. (1997). Mapping quantitative trait loci associated with regeneration ability of seed callus in rice, *Oryza sativa* L. *Theor. Appl. Genet.* 96:828-833.

Takahashi T. and Yamanaka S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126:652-655.

Takeda A., Iwasaki S., Watanabe T., Utsumi M., Watanabe Y. (2008). The mechanism selecting the guide strand from small RNA duplexes is different among Argonaute Proteins. *Plant and cell Physiology* 49:493-500.

Takhare D., Tang W., Hill K., Perry S.E. (2008). The MADS-domain transcriptional regulator AGAMOUS-LIKE 15 promotes somatic embryo development in *Arabidopsis* and soybean. *Plant Physiology.* 146:1663-1672.

Takhur J., Tyagi A.K., Khurana J.P. (2001). OslAA1, an Aux/IAA cDNA from rice, and changes in its expression as influenced by auxin and light. *DNA Res.* 8:193-203.

Taylor R. L. (1967). The foliar embryos of *Malaxis paludosa*. *Can. J. Bot.* 45:1553-1556.

Terranova R., **Pereira C. P., Du Roure C., Merkschlager M., Fisher A.G. (2006)**. Acquisition and extinction of gene expression program are separable events in heterokaryon reprogramming. *J. Cell Sci.* 119:2065-2072.

Timmermans M.C., Hudson A., Becraft P.W., Nelson T. (1999). *ROUGH SHEATH 2* :a Myb protein that represses *knox* genes in maize lateral organ primordia. *Science* 284:151-153.

Thelen M.P. and Northcote D.H. (1989). Identification and purification of a nuclease from *Zinnia elegans* L: a potential marker for xylogenesis. *Planta* 179:181-195.

Thibaud-Nissen F., Shealy R.T., Khanna A., Vodkin L.O. (2003). Clustering of microarray data reveals transcript patterns associated with somatic embryogenesis in soybean. *Plant Physiology* 132:118-136.

Thoma S, Hecht U, Kippers A, Botella J, De Vries S, Somerville C.(1994). Tissue-specific expression of a gene encoding a cell wall localized lipid transfer protein from *Arabidopsis*. *Plant Physiology* 105 :35–45

Toonen M.A., Hendriks T., Schmidt E.D.L., Verhoeven H.A., van Kammen A., de Vries S.C. (1994). Description of somatic-embryo-forming single cells in carrot suspension cultures video cell tracking. *Planta.* 194:565-572.

Toonen M.A., Hendriks T., Schmidt E.D.L., Verhoeven H.A., van Kammen A., de Vries S.C. (1996). Expression of the JIM8 cell wall epitope in carrot somatic embryogenesis. *Planta* 200:167-173.

Toonen M.A., Verhees J.A., Schmidt E.D., van Kammen A., de Vries S.C. (1997). *A*tLTP1 luciferase expression during carrot somatic embryogenesis. *Plant J.* 12:1213-1221.

Touraev A., Ilham A., Vicente O., Heberle-Bors E. (1996). Stress-induced microspore embryogenesis in tobacco: an optimized system for molecular studies. *Plant Cell Rep* 15:561-565.

Trigiano R.N., Gray D.J., Conger B.V., McDaniel J.K. (1989). Origin of direct somatic embryos from cultured leaf segments of *Dactylis glomerata*. *Bot. Gaz* 150:72-77.

Tsiantis M., Schneeberger R., Golz J.F., Freeling M., Langdale J.A. (1999). The maize *rough sheath2* gene and leaf development programs in monocot and dicot plants. *Science* 284:154-156.

## V

Van Engelen F.A., Sterk P., Booij H., Cordewener J.H.G., Rook W., van Kammen A., de Vries S.C. (1991). Heterogeneity and cell-type specific localization of a cell wall glycoprotein from carrot suspension cells. *Plant Physiology* 96:705-712.

Van Engelen F.A., de Jong A.J., Meijer E.A., Kuil C.W., Meyboom J.K., Dirkse W.G., Booij H., Hartog M.V., Vandekerckhove J., de Vries S.C. (1995). Purification, immunological characterization and cDNA cloning of a 47kDa glycoprotein secreted by carrot suspension cells. *Plant Mol. Biol.* 27:901-910.

Van Hengel A.J., Guzzo F., Van Kammen A., De Vries S.C. (1998). Expression pattern of the carrot EP3 endochitinase genes in suspension cultures and developing seeds. *Plant Physiol.* 117:43-53.

Van Ooijen J.W., Boer M.P., Jansen R.C., Maliepaard C. (2002). *MapQTL® 4.0, Software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen, the Netherlands.*

Van Stallen N., Vandenbussche B., Verdoodt V., de Proft M. (2003). Construction of a genetic linkage map for witloof (*Cichorium intybus* L. var. *foliosum* Hegi). *Plant Breed* 122:521-525.

## Références bibliographiques

Van Stallen N., Vandebussche B., Londers E., Noten V., de Proft M. **(2005a)**. QTL analysis of important pith characteristics in a cross between two inbred lines of chicory (*Cichorium intybus* var. *foliosum*): a preliminary study. *Plant Breed* 124:54-58.

Van Stallen N., Vandebussche B., Londers E., Noten V., de Proft M. **(2005b)**. QTL analysis of production and taste characteristics of chicory (*Cichorium intybus* var. *foliosum*). *Plant Breed* 124:59-62.

Vasseur J., Dubois J., Hilbert J.L., Couillerot J.P. **(1995)**. Somatic embryogenesis in chicory (*Cichorium* species). *Biotechnology for Agriculture and Forestry*. Vol 31. Somatic Embryogenesis and Synthetic seed II, Springer, Berlin Heidelberg, New York, pp. 125-137.

Vaucheret H., Vasquez F., Crété P., Bartel D.P. **(2004)**. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes development* 18:1187-1197.

Vaucheret H. **(2008)**. Plant ARGONAUTES. *Plant in Trends Science* 13 :1360-1385

Vignols f., Wigger M., Garcia-Garrido J.M., Grellet F., Kader J.C., Delseny M. **(1997)**. Rice lipid transfer protein (LTP) genes belong to a complex multigene family and are differentially regulated. *Gene* 195:177-186.

Vos P., Hogers R., Bleeker M., Reijans M., Van de Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M., Zabeau M. **(1995)**. AFLP: a new technique for DNA fingerprints. *Nucleic Acids Research* 23:4407-4414.

## W

## Références bibliographiques

Waites R., Selvadurai H.R., Oliver I.R., Hudson A. (1998). The *PHANTASTICA* genes encode a MYB transcription factor involved in growth and dorsoventrality of lateral organs in *Antirrhinum*. *Cell* 93:779-789.

Wasteneys G.O. (2002). Microtubule organization in the green kingdom: chaos or self-order? *J. Cell. Sci.* 115:1345-1354.

Wilde H.D., Nelson W.S., Booij H., de Vries S.C., Thomas T.L. (1988). Gene expression programs in embryogenic and non-embryogenic carrot cultures. *Planta* 176:205-211.

Windels D. (2005). Etude de l'implication des arabinogalactanes protéines (AGPs) au cours de l'embryogenèse somatique de différents génotypes de chicorée. Thèse à l'université de Lille 1.

Wiwerge M., Farbos I., Ingouff M., Lagercrantz U., von Arnold S. (2003). Expression of chia4-Pa chitinase genes during somatic and zygotic embryo development in Norway spruce (*Picea abies*): similarities and differences between gymnosperm and angiosperm class IV chitinases. *J. Exp. Bot.* 54:2691-2699.

Wurtele E.S., Wang H., Durgerian S., Nikolau B.J., Ulrich T.H. (1993). Characterization of a gene that is expressed early in somatic embryogenesis of *Daucus carota*. *Plant Physiology* 102:303-312.

## Y

Yarbrough J.A. (1932). Anatomical and developmental studies of the foliar embryos of *Bryophyllum calycinum*. *Am. J. Bot.* 19:443-453.

Yasuda H., Nakajima M., Ito T., Ohwada T., Masuda H. (2001). Partial characterization of genes whose transcripts accumulate preferentially in cell clusters at the earliest stage of carrot somatic embryogenesis. *Plant Mol. Biol.* 45:705-712.

## Références bibliographiques

Yazawa K., Takahata K., Kamada H. (2004). Isolation of the gene encoding Carrot Leafy Cotyledon 1 and expression analysis during somatic and zygotic embryogenesis. *Plant Physiol. Biochem.* 42:215-223.

Yeats T.H. and Rose J.K.C.(2008). The biochemistry and biology of extracellular plant lipid-transfer protein (LTPs). *Protein Sci* 17:191-198.

Yu k. and Pauls K.P. (1993). Identification of a RAPD marker associated with somatic embryogenesis in alfalfa. *Plant Molecular Biology* 22:269-277.

## Z

Zeng Z.B.(1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci. USA* 90:10972-10976.

Zimmerman J.L. (1993). Somatic embryogenesis: a model for early development in higher plants. *Plant Cell* 5:1411-1423.

Zuo J., Niu Q.W., Frugis G., Chua N.H. (2002). The *WUSCHEL* gene promotes vegetative-to-embryonic transition in Arabidopsis. *Plant J.* 30:349-359.

***Annexe 1: Article soumis à Molecular Breeding: A consensus genetic map for chicory (Cichorium intybus L.)***

***A consensus genetic map for chicory (Cichorium intybus L.)***

***Cadalen, T.<sup>1</sup>, Morchen, M.<sup>1</sup>, Blassiau, C.<sup>1,2</sup>, Clabaut, A.<sup>2</sup>, Scheer, I.<sup>1</sup>, Hilbert, J-L.<sup>1,2</sup>, Hendriks, T.<sup>1,2</sup>. Quillet, M-C.<sup>1,2</sup>***

***Addresses:***

1: GIS CARTOCHIC, USTL, Bât SN2, 3<sup>ième</sup> étage, 59155 Villeneuve d'Ascq, France

2: UMR USTL / INRA 1281 « Stress Abiotiques et Différenciation des Végétaux Cultivés », ERT 1067, Bât SN2, 3<sup>ième</sup> étage, 59155 Villeneuve d'Ascq, France

***Correspondence should be addressed to:***

***Note : Thierry Cadalen and Monika Morchen should both be considered as first authors***

**Abstract**

A consensus genetic map for chicory ( $2n=2x=18$ ) was built from the integration of molecular marker data of two segregating industrial chicory progenies (K28K59, Rubis118) and one segregating witloof chicory progeny (BR). As a limited number of co-dominant markers was available, three SSR-enriched libraries (AGR, SW, EU) were produced from genomic DNA resulting in 420 AGR, 719 SW and 1251 EU sequences. The level of informative SSR sequences from the three libraries ranged from 28% to 40%, thus defining a set of 730 SSR markers available for polymorphism screening. A subset of 81 STS (EST, genes) was screened through Single Strand Conformational Polymorphism (SSCP) of intronic and exonic regions leading to 47 polymorphic loci integrated in the genetic maps. Markers were grouped and ordered on 9 homologous linkage groups: 274 markers in K28K59, 282 markers in Rubis118, 178 markers in BR. Co-linear regions between maps were identified through 194 'bridge' markers that allowed the integration of 9 homologous linkage groups. The consensus map contained 474 markers and covered 881.7 cM. Comparison across maps revealed the presence of 4 conserved regions with significant distorted markers, also defined as Segregation Distortion Regions (SDR), affected by gametic or zygotic selection factors. Marker distribution was not uniform; 6 linkage groups possessed homologous clustered regions in all maps. The consensus maps could be the starting point for the identification and the cloning of major genes and QTL in fundamental and applied genetic areas in chicory.

***Key words : SSR-enriched libraries; STS markers; SSR markers; genetic map; consensus map; Cichorium intybus***

## **Introduction**

Chicory (*Cichorium intybus*, L.) is a diploid plant species ( $2n=2x=18$ ), belonging to the *Asteraceae* family, subfamily *Cichoriodeae*, tribe *Lactuceae* or *Cichorieae* (Bremer 1994, Panero & Funk 2002). This species is naturally allogamous, due to a sporophytic self-incompatibility system (Eenink 1981, Varotto *et al.* 1995). In addition, outcrossing is favoured by a floral morphology unfavourable to selfing in the absence of pollinators (Pécaut 1962, Desprez *et al.* 1994) and by pollen tube competition (Desprez & Bannerot 1980, Eenink 1982). Long appreciated as a medical plant (Wittop Koning & Leroux 1972), *Cichorium intybus* L. is now cultivated for numerous different applications, and can be divided into varieties or cultigroups according to their use:

- 'industrial' or 'root' chicory, mainly cultivated in northwestern Europe (northern France, Belgium, The Netherlands) and in India, South Africa, Chile. Its taproot is used for the production of a coffee substitute or for the extraction of inulin and fructans which are known as prebiotics with promising features in the food industry (Kaur & Gupta 2002) or in preventive treatment of human diseases (Pool-Zobel 2005). Breeder's practices have evolved from mass selection to the use of a small set of selected families to obtain high-yielding synthetic varieties, and to hybrid production by the use of a cytoplasmic male sterility system (Desprez *et al.* 1994, Rambaud *et al.* 1993).
- 'Brussels' or 'witloof' chicory, commonly cultivated in the same european area than industrial chicory for the production of heads or chicons under forcing conditions. Witloof chicory is considered a derivative from the Magdeburg type of root chicory. In contrast to other chicory cultigroups, breeding efforts have lead to release inbred lines with reduced inbreeding depression for producing F<sub>1</sub> hybrids adapted to hydroponic forcing (Bannerot & de Coninck 1970, Eenink 1980).
- 'leaf' chicory, used as fresh or cooked vegetables, like the green-leaved varieties of the 'Pain de Sucre' cultigroup and the 'blond', 'red' or 'variegated' varieties of the 'Radicchio' cultigroup from northern Italy.
- 'forage' chicory, initially derived from wild chicory commonly found along roadsides and waste areas, and used since the mid-1970s to increase herbage availability in perennial pastures for grazing livestock. Thanks to the selection of varieties that meet

nutritional needs and stand performance (Li & Kemp 2005), the application of forage chicory is increasing worldwide (New Zealand, Australia, America, Europe).

During the last decade, the development of dominant molecular marker systems, essentially RAPD and AFLP, have largely contributed to classify germplasms and discriminate the genetic diversity within different species and cultigroups of *Cichorium* (Bellamy *et al.* 1996, Koch & Jung 1997, Kiers *et al.* 2000, van Stallen *et al.* 2000, Barcaccia *et al.* 2003, van Cutsem *et al.* 2003).

The same kind of markers was used to construct two genetic linkage maps for chicory. The first one was based on an interspecific F<sub>1</sub> outbred population between a 'variegated' chicory and an inbred line of *C. endivia* L., a self-compatible species with chromosome structure (n=9) closely related to chicory (Rick 1953). Following a 'pseudo-testcross' scheme, firstly described by Grattapaglia & Sederoff (1994) and used from that time on in many outbred species (Cervera *et al.* 2001, Kenis & Keulemans 2005, Yan *et al.* 2005), de Simone *et al.* (1997) used dominant PCR-based markers (RAPD, AFLP, SAMPL) specific for *C. intybus* L. to construct a map that covered 1204 cM with 351 markers dispersed over 13 linkage groups. The second map was based on an F<sub>2</sub> population derived from a cross between two inbred lines of witloof chicory varieties. This map covered 609 cM with 129 RAPD markers on 9 linkage groups (van Stallen *et al.* 2003), later completed by AFLP markers to dissect the quantitative variation of complex characters by detecting QTL (van Stallen *et al.* 2005a, van Stallen *et al.* 2005b).

Unfortunately, both maps suffer from a lack of codominant and readily transferable markers, rendering their integration impossible, and thus limiting their use in the construction of a consensus genetic map for *C. intybus* L. In addition, as pointed out by Maliepaard *et al.* (1997), the use of dominant markers may cause problems with respect to the probability to detect linkage and the ability to distinguish linkage phase combinations. The development of large sets of co-dominant markers will facilitate the process of constructing genetic maps and identifying major genes or QTL of interest in outbred species, and, ultimately, help breeders in marker-assisted-selection schemes involving diversified genetic material.

Microsatellite or Simple Sequence Repeat (SSR) markers are considered as ideal co-dominant markers for genetic linkage mapping in plants because of their neutral status, their abundance and polymorphism, their high information content and their mostly monogenic inheritance (Morgante & Olivieri 1993). All these features

have greatly facilitated SSR utilization in connecting maps from various mapping populations and constructing consensus genetic maps (Cervera *et al.* 2001, Yu *et al.* 2003, Gonzalo *et al.* 2005). SSR markers were initially obtained after screening genomic libraries enriched with SSR motifs and selecting unique SSR sequences suitable for primer design (Zane *et al.* 2002, Squirrell *et al.* 2003). More recently, the development of large sets of Expressed Sequence Tags (EST) resulted in genic sequence data available for SSR search and mapping (EST-SSR or genic SSR, Varshney *et al.* 2005). In chicory, the number of available EST to derive EST-SSR was limited until recently. Consequently, the isolation of a large number of SSR markers involved either the production and sequencing of large sets of EST data or the construction of SSR-enriched libraries. The second alternative was chosen when we started our mapping project (Cartochic project).

In this article, we report the construction of three molecular genetic maps in chicory, two derived from industrial chicory mapping populations, and one derived from a witloof mapping population. After screening 3 SSR-enriched libraries, a total of 730 SSR sequences were surveyed for identify polymorphic markers used for the construction of three genetic maps. In addition, a set of 81 sequence tagged sites (STS) was screened by single strand conformational polymorphism (SSCP), and polymorphic loci added to the maps. Homologous loci between the three maps were used as bridge markers to construct a consensus genetic map of 474 markers covering 881,7 cM on 9 linkage groups.

## ***Material and Methods***

### ***Plant Material***

Three mapping populations were used. The first one, called K28K59, was an outbred population of 177 F<sub>1</sub> individuals obtained after crossing the heterozygous parental genotypes K28 (male sterile) and K59 (hermaphrodite). Both parents are industrial chicories obtained from an improved Hungarian landrace population originally called 'Koospol' (Florimond-Desprez, Cappelle-en-Pévèle, France). The second population, called Rubis 118, was an F<sub>2</sub> population of 96 progenies originated

from the cross between the male sterile clone MS8 and one individual of the industrial chicory cultivar 'Cassel' (Florimond-Desprez, Cappelle-en-Pévèle, France). The third mapping population, called BR, was an F<sub>2</sub> population of 145 progenies originated from the cross between two inbred lines of parental witloof chicories: a white leaf-coloured genotype B and a red leaf-coloured genotype R (Hoquet Endives, Raillencourt, France). One F<sub>1</sub> hybrid plant was self-pollinated to generate the mapping population.

Parents K28, K59, MS8 and the F<sub>1</sub> Rubis 118 were all maintained through *in vitro* cloning procedures.

### **DNA extraction**

Genomic DNA was extracted with the GenElute™ Plant Genomic DNA Miniprep Kit (Sigma-Adrich Chemie GmbH, Germany) from 10 mg of dried leaves or 100 mg of fresh leaves. Yield and quality of extracted DNA was checked on a UV/vis spectrophotometer (Biophotometer, Eppendorf, Germany) and on 0.8% agarose gels.

### **Detection and scoring of markers**

#### **SSR markers**

Three SSR-enriched libraries were used. The first one was produced by the former company Agrogène, France. The library contained 420 enriched clones, referred to as 'AGR' clones, divided in 161 clones with a (CA)<sub>n</sub> motif, 193 clones with a (CT)<sub>n</sub> motif, 33 clones with a (CTT)<sub>n</sub> motif, 33 clones with a (ATT)<sub>n</sub> motif.

The second one was produced by the company Ecogenics GmbH (Zürich, Switzerland) from size selected genomic DNA of chicory (genotype K59) according to Armour *et al.* (1994), and enriched by magnetic bead selection with biotin-labelled (CA)<sub>13</sub> and (GA)<sub>13</sub> oligonucleotide repeats (Gautschi *et al.* 2000a, Gautschi *et al.* 2000b). Of 2185 colonies, 456 (111 with CA motif, 345 with GA motif) gave a positive signal after hybridisation. Positive recombinant clones, referred to as 'SW' clones, were sequenced by the company Genoscreen (Lille, France).

The third one was later produced in our laboratory using the enrichment protocol described by Glenn & Schable (2005). Using genomic DNA of K59 genotype, 2228 recombinant clones were obtained after enrichment by magnetic bead selection with a mixture of 3' biotinylated oligonucleotide repeats including (AG)<sub>12</sub> and (TG)<sub>12</sub>. These clones, referred to as 'EU' clones, were sequenced by the company Genoscreen (Lille, France).

Both SW and EU sequences were trimmed for vector sequences in BioEdit and submitted to Sputnik (<http://espressosoftware.com/pages/sputnik.jsp>), a program designed for the detection of SSR motifs. A Blast local database of all SSR-enriched clones was performed with Bioedit v 4.8.10 in order to search for inter- and intra-library redundancy. All groups of redundant clones were aligned and analysed in Bioedit (ClustalW Multiple Alignment) in order to define consensus sequences. Primers were designed with Primer3 v 0.3.0.

For each primer pair designed, one primer was labelled with the fluorescent dyes (6-Fam or Hex). The PCR assay was performed in 15 µl containing 1x PCR buffer with 2 mM MgCl<sub>2</sub>, 100 µM of each dNTP, 0.2 mg/ml BSA, 133 nM of each primer, 0.3 U Taq polymerase (Applied Biosystems) and 6 ng of template DNA. A touchdown (TD) procedure was applied : 5 min denaturation at 94°C, followed by (a) 5 cycles of 30s at 94°C, TD 30s at T<sub>m</sub> (-1°C per cycle), 30s at 72°C and (b) 35 cycles of 30s at 94°C, 30s at T<sub>m</sub>-5°C, 30s at 72°C, before 10 min of extension at 72°C.

PCR products were analysed either on a 3100 Avant Genetic Analyzer capillary sequencer (Applied biosystems). Genotyping data were automatically collected and analysed, (3100/3100 Avant Data Collection, Genescan<sup>®</sup> Analysis Software v 3.7.1, Genotyper v 3.7 NT).

### **STS markers**

Sequence Tagged Sites (STS) involved a set of EST, cDNA, genes and non-coding sequences of chicory listed in public databases and isolated in our lab. Both sources represented 81 STS. Tab 1 summarizes the information related to the 47 STS mapped. SNP and InDel polymorphism of STS was revealed through Capillary Electrophoresis – Single Strand Conformational Polymorphism (CE-SSCP) (Andersen *et al.* 2003) on a ABI 3100-Avant Genetic Analyser sequencer (Applied Biosystems).

Prior to primer design, sequence alignments between chicory cDNA or EST and Arabidopsis genes were searched with Blast tools in the NCBI site, Fasta 3 program in the EMBL-EBI site and ClustalW Multiple Alignment in Bioedit v 4.8.10. Intron positions were determined and then compared to predictions by models for intron/exon boundaries determination with RNASPL tool, <http://www.softberry.com/berry.phtml> or GeneMark cDNA, [http://opal.biology.gatech.edu/GeneMark/genemark\\_cDNA\\_all.cgi](http://opal.biology.gatech.edu/GeneMark/genemark_cDNA_all.cgi). For genes containing introns, primers were designed in regions surrounding putative intronic sequences. Primers were designed with Primer3 v 0.3.0.

Each primer of the primer pair designed was labelled with a different fluorescent dye (6-Fam or Hex). The PCR assay was performed in 15 µl containing 1x PCR buffer with 2 mM MgCl<sub>2</sub>, 100 µM of each dNTP, 0.2 mg/ml BSA, 133 nM of each primer, 0.3 U Taq polymerase (Applied Biosystems) and 6 ng of template DNA. A TD procedure was applied : 5 min denaturation at 94°C, followed by (a) 5 cycles of 30s at 94°C, TD 40s at T<sub>m</sub> (-1°C per cycle), 1min at 72°C and (b) 35 cycles of 30s at 94°C, 30s at T<sub>m</sub>-5°C, 1min at 72°C, before 10 min of extension at 72°C.

Sample preparation, denaturation and CE-SSCP fractionation were performed according to User Bulletin 'ABI PRISM<sup>®</sup> 3100 and 3100-Avant Genetic Analyzers: High Throughput Fluorescent SSCP Analysis' (Applied Biosystems) with a minor modification in the concentration of Genescan<sup>™</sup> polymer (3% instead of 7%). Data were collected (3100/3100 Avant Data Collection program) and normalised before analysis (Genescan<sup>®</sup> Analysis Software v 3.7.1 and Genotyper v 3.7 NT program).

### ***Segregation distortion analysis***

The segregation of each marker was evaluated by a  $\chi^2$  test of goodness-of-fit to the expected Mendelian segregation ratio *i. e.* 1:2:1 for co-dominant markers, 3:1 for dominant markers in F<sub>2</sub> configuration type (all markers of the mapping populations Rubis118 and BR, one part of the markers segregating in the K28K59 mapping population) 1:1 for markers in backcross configuration type, 1:1:1:1 for tri-allelic or four-allelic markers (different segregating type markers in the K28K59 mapping population).

For significantly skewed markers ( $P < 0.05$ ), the selection type (zygotic vs gametic) was then determined for co-dominant markers in  $F_2$  configuration type using two successive  $\chi^2$  tests (Lorieux *et al.* 1995). Similarly, following the procedure described by Lorieux *et al.* (1995),  $\chi^2$  tests for co-dominant markers with three and four alleles were derived (see the Appendix for more details)

### ***Genetic linkage map construction***

Linkage analysis and map calculations were performed using the program JoinMap v 3.0 (van Ooijen & Voorrips 2001). For each mapping population, linkage grouping was determined using a LOD threshold = 3. For linkage groups building, the linear order of markers was determined in a stepwise procedure by adding the most informative markers one at a time and accepting the most likely order with the best goodness-of fit. JoinMap setting parameters were: maximum recombination frequency = 0.45, LOD = 1.0, jump = 5. Pair-wise recombination frequencies were converted in map distances (cM) using the Kosambi mapping function (Kosambi 1944). Linkage analysis was systematically divided in two groups of markers for each mapping population. The first group involved the complete set of markers, the second group only included homologous markers, also referred to as 'bridge markers'. Bridge markers were markers that share common alleles between mapping populations, thus identifying homologous loci. One homologous group was identified when the linear order of the same set of bridge markers was 1) identical in two or three mapping populations, and 2) consistent with the order including the complete set of markers. By generalizing this mapping procedure to all linkage groups of the 3 populations, genetic maps of Rubis118, K28K59 and BR including specific and bridge markers were generated.

Homologous pair-wise marker data across two or three genetic maps were integrated by JoinMap (Map Integration option, van Ooijen & Voorrips 2001), before merging homologous groups and, ultimately constructing the consensus or framework linkage map. The mapping parameters above-mentioned for generating maps were used for the consensus map. As emphasized by Gonzalo *et al.* (2005), the inspection of significant heterogeneity tests ( $P < 0.01$ ) calculated by JoinMap has allowed the exclusion of markers causing severe misordering in the consensus map.

Further, all homologous groups were drawn and aligned using the program MapChart v 2.1 (Voorrips 2002) and compared to the consensus homologous group. When significant conflict orders appeared between individual maps and the consensus map, multiple rounds were assayed by excluding one marker at a time in the critical region until the final order was as consistent as possible with the initial order on the individual maps.

## Genome coverage and marker clustering

As described by Yu *et al.* (2003) for sunflower, expected genome coverage  $G_e$  was

estimated by the expression proposed by Fishman *et al.* (2001) :  $G_e = G_o + (2tG_o)/n$ , where  $G_o$  is the observed length in cM of the genetic map,  $t$  is the number of linkage groups and  $n$  is the number of marker intervals. We have also used the expression 4 of Chakravarti *et al.* (1991) generalized to all linkage groups :  $G_e = \sum G_{oi} [(k_i + 1)/(k_i - 1)]$  where  $G_{oi}$  is the observed length of the  $i$ th linkage group and  $k_i$  is the number of mapped markers on the  $i$ th linkage group. Both expressions giving similar results, only the Fishman's estimator was presented in Tab 3. The proportion of genome coverage was then estimated by the ratio  $G_o/G_e$ . The probability that at least one marker is within  $d$  cM of a randomly marker locus was estimated by Lange & Boehnke (1982):  $P_{LB} = 1 - e^{-2dk/G_o}$  where  $k$  is the number of loci.

Marker distribution between and within linkage groups of each population was evaluated according to Remington *et al.* (1999).

Inter-linkage groups comparisons (ILGC analysis) were done under the hypothesis of an uniform marker distribution for all linkage groups: the number  $k_i$  in linkage group  $i$  followed a Poisson distribution with parameter  $\lambda_i = kG_i/\sum G_i$  where  $G_i$  is the estimated length of the  $i$ th linkage group such as:  $G_i = G_{oi} + 2s$  ( $s$  is the average marker spacing). The cumulative probability function (cpf) was expressed for each linkage group by:

$$\sum_{k_i=0}^{k_i} \frac{e^{-\lambda_i} \lambda_i^{k_i}}{k_i!}$$

Poisson probabilities [ $P(\lambda_i \geq k_i)$  or  $P(\lambda_i \leq k_i)$ ] were significant when the cdf value or one minus the cdf value were less than  $\alpha/2$  (two-tailed Poisson test under cumulative distribution,  $\alpha=0.05$ ).

Within-linkage groups comparisons (WLGC analysis) for marker clustering were done after dividing each  $i$ th linkage group in windows of width  $W_i$ :

- 1)  $W_i$  was fixed for all linkage groups ( $W_i = 10$  cM )
- 2)  $W_{ij}$  was defined after selecting a subset of markers also called 'framework markers'

The criteria of choice for 'framework markers' were : uniform distribution on the linkage group, co-dominance inheritance, minimum number of missing data and low impact of segregation distortion. Intervals between framework markers  $j$  in the  $i$ th linkage group ( $W_{ij}$ ) were then variable.

In both cases, if markers are randomly distributed, the number of mapped markers in the  $i$ th linkage group ( $k_i$ ) followed a Poisson distribution with parameter  $\lambda_i = k_i W_i / G_i$  ( $W_i$  fixed) or  $\lambda_{ij} = k_i W_{ij} / G_i$  ( $W_{ij}$  variable). Poisson probabilities, calculated for each interval of the linkage group  $i$ , was significant to a  $\alpha/2$  threshold. Intervals with significant Poisson probabilities  $P(\lambda_i \leq k_i)$  revealed significant clustering of markers.

### **Analysis of parental genotypes through bridge markers**

A selection of 130 bridge markers representing 478 alleles was chosen from two main criteria: unambiguous identification of alleles in parents of the three mapping population and uniform distribution on the 9 LG of the genetic maps. The proportion of observed heterozygous loci  $p_n$  was scored for the  $n=6$  parental genotypes. To test the hypotheses,  $H_0: p_1 = p_2$  vs  $H_1: p_1 \neq p_2$ , a Z-test for comparison of two proportions from independent samples (Snedecor & Cochran 1989) was computed. The complete set of 478 alleles was then converted into binary variables (0/1), where the presence of the allele was encoded by 1 whereas the absence of the allele was encoded by 0. The estimation of genetic resemblance for each pair of parental genotypes was based on the calculation of the Dice's similarity coefficient,  $S_D$ , given by the following formula:  $S_D = 2a / (2a + b + c)$  (Dice

1945) where a codes for matches (1,1), b and c code for mismatches (1,0) and (0,1). According to Sneath & Sokal (1973), the matrix of similarity coefficients was transformed under the form of a dendrogram using the agglomerative and hierarchic cluster procedure UPGMA (Unweighted Pair-Group Method using arithmetic Average). The complete procedure was carried out using the program DARwin 5.0.157 (Perrier & Jacquemoud-Collet 2006).

## **Results**

### **SSR and STS screening**

All inserts from clones of SSR-enriched libraries were sequenced; the derived sequences were then included in a database containing 2367 sequences: 420 sequences from the AGR library, 719 sequences from the SW library, 1251 sequences from the EU library. The level of enrichment ranged from 75% to 85%. Unique sequences containing a SSR repeat varied from 42% for the AGR library to 53 and 57% for EU and SW library, respectively. Inter-library redundancy represented 10 to 20% of the copy sequences. Consequently, the samplings derived from three enriched-SSR libraries can be regarded as three complementary sets of SSR sequences.

In Fig 1, the distribution of sequences into different fractions has been illustrated for the EU library, which is the most productive enriched library for the sequences containing a SSR motif subsequently converted in markers. The whole set of 1251 EU sequences were partitioned in 608 copy or redundant sequences and 643 unique sequences. Three hundred and thirty-two sequences (26%) were removed according to several quality criteria: lack of SSR (9%), SSR number or repeats < 10 repeats (8%), poor sequence quality (1%), insufficient number of bases flanking the SSR to design primers (4%), repetitive sequences (4%). These last sequences were a set of repeated sequences, with similarities to minisatellites identified in *Beta* species by Jung *et al.* (1993). Alignments of redundant sequences

containing SSR repeats allowed the constitution of contigs from which 188 (15%) consensus SSR sequences were derived (Fig 1). By adding 311 (25%) unique SSR sequences available for primer design, the complete set of informative SSR was 499 SSR or 40% of the sequences in the EU library. The level of informative SSR from the AGR library was similar to the EU library (39% or 155 SSR). This level was reduced in the SW library (28% or 208 SSR) because the mean size of the inserts containing the SSR was significantly lower (246 bp in the SW library vs 500 bp in EU or AGR libraries) thus increasing the probability to obtain SSR truncated sequences or SSR sequences with insufficient flanking sequences to design primers.

Seven hundred and thirty SSR were surveyed for polymorphism on the 6 parental genotypes: K28, K59, MS8, Rubis 118 (F<sub>1</sub>), B, R. Because of PCR amplifications problems or complex SSR profiles, the 499 informative EU SSR was finally restricted to 367 available SSR. Polymorphism survey revealed 258 (73%) polymorphic markers, or, at the population level, 187 (53%), 153 (43%) and 98 (27%) polymorphic markers for K28K59, Rubis118 and BR populations, respectively. The same survey for the 155 available AGR SSR revealed 114 (73%) polymorphic markers, or, at the population level, 79 (49%), 64 (41%) and 55 (35%) polymorphic markers for K28K59, Rubis118 and BR populations, respectively. The proportion of polymorphic markers was lower for the SW SSR (135 markers *i.e.* 65%). The number of informative markers was 97 (46%) for K28K59, 79 (37%) for Rubis 118 and 67 (32%) for BR. In brief, the K28K49 population is the most informative population, most likely due to its F<sub>1</sub> outbred structure, followed by the Rubis118 population and, finally, by the BR witloof population characterized by an 'inbred' structure with less polymorphic markers than the industrial chicory parents.

Eighty-one STS were screened for polymorphism through SSCP analysis. Polymorphism survey on the set of parental genotypes revealed 52 polymorphic STS (64.1%) between the 3 mapping populations. A strong divergence in terms of polymorphic loci appeared between the industrial chicory populations and the witloof population: 39 (48.1%) and 38 (46.9 %) STS were polymorphic for K28K59 and Rubis118, whereas only 12 (14.8%) STS were polymorphic for BR. Tab 1 lists the assignments of 46 STS locus to the 9 Linkage Groups (LG) of the chicory map. By excluding LG7 with only one STS marker, the distribution of STS was dispersed

throughout the genome, ranging from 3 markers (LG5) to 9 markers (LG8) per linkage group.

## Genetic maps for industrial and witloof chicory

Two hundred and ninety one marker loci were genotyped on the 177 progenies of the K28K59 population. After excluding unlinked markers, the genetic map included 274 mapped markers and covered 749.8 cM with an average marker spacing  $s = 2.7$  cM. Loci were distributed over 9 linkage groups (LG) and were represented by 68 AGR SSR, 37 SW SSR, 143 EU SSR and 26 STS (Fig 2). LG9, LG7 and LG4 had the lowest number of markers  $k_i$  ( $19 < k_i < 21$ ). LG7, LG4 and LG2 had the smallest coverage (observed genome coverage  $G_{oi} < 66$  cM). Conversely, LG1, LG3, LG6 and LG8 were top groups for mapped markers ( $31 < k_i < 48$ ), with  $G_{oi}$  ranging from 89.2cM (LG6) to 106.6 (LG1). The majority of markers were informative in only one parent (65% - backcross configuration types +  $F_2$  dominant type). The remaining 35% markers were informative in both parents ( $F_2$  co-dominant type, three and four allele configuration types) and allowed to construct an integrated genetic map for the K28K59 segregating population.

The 96 progenies of the Rubis118 population were genotyped for 302 markers. Twenty markers were excluded during the mapping procedure and remained unlinked. As showed in Fig 2, the genetic map included 282 markers and covered 806.1 cM with an average marker spacing  $s = 2.8$  cM. These loci were represented by 59 AGR SSR, 65 SW SSR, 130 EU SSR and 28 STS. LG2 LG4 and LG7 were characterized by the smallest number of markers ( $19 < k_i < 25$ ). LG2 and LG4 had also the smallest genetic length ( $G_{oi} < 64$  cM). LG1, LG3, LG6 and LG8 had the largest genetic length ( $G_{oi} > 100$  cM) and were constituted by the largest number of markers ( $35 < k_i < 43$ ). Eighty percent markers were co-dominant over the whole map. This proportion ranged from 63% for LG8 to 91% for LG4.

The 145 progenies of the BR population were genotyped for 190 markers. Twelve SSR markers could not be linked. The remaining loci were distributed over 9 LG (37 AGR SSR, 61 SW SSR, 73 EU SSR and 7 STS). The genetic map included 178 markers and covered 891.4 cM with an average marker spacing  $s = 5$  cM (Fig 2). LG2 was the smallest group ( $G_{o2} = 46.1$  cM with  $k_2 = 11$  markers). The number of

markers  $k_i$  was relatively stable for five LG ( $18 < k_i < 23$  for LG1, LG3, LG4, LG7, LG8 and LG9). The highest value of  $k_i$  was for LG6 ( $k_6 = 33$ ), also characterised by a large size ( $G_{06} = 122$  cM). Seventy-two percent markers were co-dominant over the whole map. This proportion ranged from 60% for LG6 to 80% for LG5 and LG8.

The level of recombination seemed more generally better in the BR witloof chicory map compared to industrial chicory maps, as revealed by the differences in LG lengths. Five LG had largest sizes in the BR map. The remaining LG had similar sizes between maps (LG9, LG7) or were less well covered in the BR map (BR2, BR3). The number  $k_i$  was high in the LG6 of the three maps. This number was also remarkably homogeneous for homologous LG4 and LG7 ( $19 < k_i < 22$ ).

### **Bridge markers to construct the consensus map**

Bridge markers or homologous loci were used to identify co-linear regions between maps. In Tab 2, relationships between the three maps are synthesized through the information displayed by bridge markers.

Though bridge markers were more numerous in the K28K59 and Rubis118 maps than in the BR map (154, 161 and 96 respectively), the proportion of bridge marker to the whole set of mapped markers was stable for all maps (53.9 to 57.1% - Tab 2a). The same proportion applied to each LG revealed that at least half of the markers of the homologous LG1 to LG8 were bridge markers (Tab 2a). This situation would facilitate the process of merging maps for the construction of the reference map. Nevertheless, LG9 had a specific situation with less bridge markers than other LG in absolute values and in proportions (<50% -Tab 2a).

Four kind of connections or 'inter-map comparisons' existed when three maps were compared (Tab 2b). A total of 193 connections were present in the whole set of maps. Ninety-seven connections (50.2%) were associated to bridge markers linking the 'K28K59/Rubis 118' maps. As a consequence, this connection was always predominant for all LG, though the distribution of the four connections was more balanced for LG4, LG5, LG6, LG7 and LG9. Values associated to the connections linking industrial chicory maps(s) and the witloof chicory map were relatively close (14 to 19.7% -Tab2b). The tri-map and most informative connection

'K28K59/Rubis118/BR' was present in all LG, ranging from one connection (LG2, LG9) to five connections (LG5, LG6).

Briefly, alignment of homologous LG and comparison of loci orders (Fig 2, Tab2) showed that rich –marker LG (LG1, LG3, LG6, LG8) contained also the most important number of bridge markers and inter-map comparisons, and inversely for less well-covered LG. However, homologous LG7 were an exception. Despite their low  $k_i$  value, all LG7 had a strong proportion of bridge markers (Tab2a).

Co-linear intervals flanked by bridge markers in the three individual maps (K28K58, Rubis118 and BR) were integrated to construct the consensus LG. As showed in Fig 3, the consensus map was constructed with 474 mapped loci dispersed over 9 LG and covered 881.7 cM (average marker spacing  $s = 1.8$  cM). Markers were distributed among 105 AGR SSR, 118 SW SSR, 209 EU SSR and 42 STS.

The genetic map can be divided into two groups of LG. The first one included five LG (LG1, LG3, LG6, LG8 and LG9) who had an observed genetic length  $G_{oi} > 100$  cM. They were also the richest LG for the number of mapped markers ( $50 < k_i < 75$ , Fig 3). For LG9, the predominance of specific-LG markers in K28K59, Rubis118 and BR maps explained its good level of coverage in the consensus map. The second one included four LG (LG2, LG4, LG5, LG7 characterised by  $35 < k_i < 43$  markers and  $G_{oi} < 87$  cM. LG7 had the lower number of mapped markers as most of them were predominantly bridge markers in the three individual maps.

### **Segregation distortion in chicory**

The calculation of a  $\chi^2$  test of goodness-of-fit for all markers revealed a contrasting situation. The K28K59 and BR maps shared a similar proportion of skewed markers (24 markers or 8.7% in K28K59, 15 markers or 8.4% in BR) whereas the Rubis118 map was comparatively characterized by a two-fold increase of the proportion of skewed markers (55 markers or 19.5%). The level of distortion varied from low ( $P < 0.05$ ) to high ( $P < 0.0001$ ) significant  $\chi^2$  test values.

In the Rubis118 map, 51 distorted markers were concentrated in 4 LG (LG4, LG7, LG8 and LG9 – Fig2). They appeared clustered in specific regions (LG4 -7 markers- and LG9 -11 markers-) or represented the majority of the markers in LG8

(18 markers) and LG7 (15 markers *i. e.* 79% of the genetic length). In each of the four potential Segregation Distortion Regions (SDR), the source of the distortion (gametic vs zygotic) was identified separately for skewed markers, using the procedure of Lorieux *et al.* (1995). Zygotic selection factors affected SDR on LG4 and LG8 whereas gametic selection factors affected SDR on LG7 and LG9.

Small clusters of distorted markers existed also in the K28K59 map (LG1, LG2, LG4, LG8 and LG9). Using the Lorieux's procedure for  $F_2$  type configuration markers and the procedure described in the Appendix for tri- and four-allelic markers, zygotic selection factors affected distorted regions in LG4 and LG8, and gametic selection factor(s) affected the SDR in LG9. This latter region in LG9 was homologous to the SDR identified in the Rubis118 map. The same gene(s) affecting the viability of gametes could be postulated and located in this homologous region of K28K59 and Rubis118 maps.

Most of the distorted markers were dispersed in the BR map. Gametic selection factor(s) could be associated to two distorted markers in L7. Zygotic selection factors could be associated to the four distorted markers in LG8.

### **Genome coverage**

Genome coverage was estimated with the Fishman's estimator. In Tab 3, parameters  $G_{ef}$ ,  $G_o/G_{ef}$  and  $P_{LB}$  have to be calculated under the hypothesis that marker loci are evenly distributed along the genome. When using the complete set of marker data, marker clustering violates the initial hypothesis of uniform distribution of markers and genome coverage parameters could be biased. Cervera *et al.* (2001) and Yin *et al.* (2004) proposed to calculate these parameters with a subset of evenly spaced markers also called 'framework' markers. We have also defined a set of 'framework' markers for K28K59, Rubis118 and BR maps using our own criteria of choice (uniform distribution, co-dominance inheritance, low missing data, low segregation distortion) and computed a 'framework' map for the three progenies (data not shown). Fishman *et al.* (2001) proposed to compare the values obtained using framework and complete marker data. Our data were then divided in 2 groups (the complete and the framework set) to express genome coverage parameters under the form of 'range' values.

Tab 3 showed that both industrial chicory maps and, to a lesser degree, the witloof chicory map had relatively similar  $G_{ef}$  values: a range from 800.3 to 986.5 cM for the complete set to a range from 932.2 to 1126 cM from the framework set.  $G_{ef}$  values of the BR maps were the highest following the trend observed for  $G_o$  values.

Genome coverage ( $G_o/G_{ef}$ ) was very closed in each subset of markers across the 3 maps (K28K59, Rubis 118, BR) and ranged from 80-82% (framework set) to 90-94% (complete set). As expected, genome coverage calculated for the reference map was higher (96%).

The probability  $P_{LB}$  was calculated with 2 conditions  $d = 5$  cM or  $d = 10$  cM (Tab 3). These calculations indicated that the three maps (K28K59, Rubis 118, BR) had still significant gaps at  $d = 5$  cM for the framework set ( $0.52 < P_{LB} < 0.61$ ), but had an acceptable coverage for applications such as QTL identification at  $d = 10$  cM ( $0.77 < P_{LB} < 0.84$  –framework set). The reference map was characterized by a dense genetic coverage ( $P_{LB} = 0.99$  for  $d = 5$  cM).

## Marker distribution

Marker distribution among LG (ILGC analysis) was evaluated for the 4 maps by comparing the observed number of markers per LG to an expected number calculated after re-estimating the size of each LG (procedure of Remington et al. 1999). The significant *cpf* values in Tab 3 ranged from 0.975 to 0.99 for LG containing more markers than expected and from 0.025 to 0.0007 for LG containing few markers than expected. The analysis revealed map-specific situation such as the lack of markers for LG9 (K28K59) or LG8 (BR) or the high accumulation of markers in LG8 (K28K59), but most of the LG in all maps were not significant thus revealing a balanced distribution of markers (Tab 3). This situation was confirmed in the reference map where only LG6 had significantly more markers than expected (Tab 3).

A second level of marker distribution was analysed by dividing each LG in segments or windows to assess clustering in specific areas (WLGC analysis). Significant *cpf* values were reported under the form of 'grey patches' on the LG (Fig 2 and 3). Two sizes of windows  $W$  were considered. When both analyses (fixed  $W_i$ , variable  $W_{ij}$ ) were significant or one was significant and the other closed to the  $\square/2$

level, the corresponding regions were represented by dark grey patches (Fig 2 and 3). When only one analysis was significant or close to the  $\alpha/2$  level, the corresponding regions were represented by pale grey patches (Fig 2 and 3). The analysis revealed 63 regions characterised by clustered markers in the four maps. All LG from all maps contained at least one clustered region. The distribution of clustered regions in the complete set of 36 LG was: 20 LG containing 1 or 2 clustered regions, 5 LG with 3 clustered regions and one LG with 4 clustered regions (LG3 – Rubis118).

Homologous clustered regions were then identified through the analysis of co-linear intervals. Six LG (LG2, LG3, LG4, LG6, LG7 and LG9) had homologous clustered regions maintained in all maps. Six out of 9 LG (LG1-2regions-, LG2, LG3, LG5, LG6 and LG8) possessed homologous clustered regions for 3 maps.

Clustered regions involved a significant proportion of mapped markers in several LG (LG2, LG4, LG7, LG8 for K28K59, LG4, LG9 for Rubis118, LG3 for BR - Fig3). Clustering data in the reference map (Fig 4) revealed 4 LG (LG1 LG2, LG3 and LG6) with two clustered regions, the first one in the distal part close to the end of the group, the second one close to the middle of the group, and 5 LG (LG4, LG5, LG7, LG8 and LG9) with only one clustered region.

### **Structuration of the parental genotypes**

Allelic information derived from polymorphism survey of the 6 parental genotypes (MS8, Rubis118, K28, K59, B, R) and mapping of polymorphic markers were extracted from bridge markers which are the most informative set of markers of the three mapping populations. Tab 4 displayed the proportion of homozygous markers for each parent and their two-by-two comparisons. Witloof chicory parents B and R shared a high and identical proportion of homozygous marker (95-97%, Z-test not significant), whereas the same proportion is less than 53% in the industrial parental chicories. As expected, this proportion was very weak (11%) for the  $F_1$  genotype Rubis118, mainly constituted by allelic heterozygous configurations. Interestingly, two couples of industrial parental chicories K28/K59 and MS8/K28 had also a significantly similar proportion of homozygous markers (Tab 4).

In Fig 4, cluster analysis based on genetic similarities (gs) between the 6 parental genotypes separated the group of two witloof parental chicories B and R

( $gs=0.78$ ) to the group of industrial chicories ( $gs=0.775$ ). Maximum  $gs$  between Rubis118 and MS8 was not surprising, as the  $F_1$  Rubis118 shared 50% alleles with the parent MS8, the other parent Cassel being unknown. Genetic similarity between K28 and K59 could be associated to their origin. They were selected from the same improved landrace, and they shared at least one allele in common for 68% of the 130 bridge markers selected for cluster analysis.

Genetic resemblance between witloof parents B and R combined to a strong allelic fixation compared to industrial parents were in accordance with the history of the witloof cultigroup (a narrow genetic base) and the breeder's effort to select inbred material. A more diverse characterization of this cultigroup using RAPD and AFLP dominant markers had already established its low level of diversity (Van Stallen *et al* 2000).

## Discussion

We have constructed a consensus genetic map in chicory based on genetic data from 3 progenies. This map contains 474 molecular markers distributed over 9 LG, which is in agreement with the haploid chromosome number of the genus *Cichorium*. The integrative process of merging genetic maps to construct a reference map has been described for several species (Jeuken *et al.* 2001, Cervera *et al.* 2001, Yu *et al.* 2003, Yan *et al.* 2005, Gonzalo *et al.* 2005), but never for chicory. Two genetic maps have been published for chicory based on AFLP and RAPD markers, respectively de Simone *et al.* (1997) and van Stallen *et al.* (2003), but the lack of co-dominant markers has hampered the construction of an integrated map for this species. The development of a large set of SSR markers from three-enriched libraries that enabling coverage of the genome and can be used in marker-assisted selection was the starting point of our mapping work.

To complete the information given by neutral markers like SSR, a subset of 81 genic sequences of chicory was added to the maps following an SSCP-based protocol which had already allowed the mapping of 15 genes in pea (Aubert *et al.* 2006). The level of conformational polymorphism revealed a 5 times less decrease from the set of industrial chicories, (K28, K59, MS8, Rubis118) to witloof chicory parents (B and R). The high throughput mapping of genes involves strategies such as the production of EST libraries or whole genome sequencing projects aimed to

SNP discovery. In chicory, the number of sequenced and well-characterised genes has been very limited for a long time. The bottleneck constituted by the lack of EST has recently disappeared since two recent publications (3,348 EST from embryogenic and non-embryogenic cDNA libraries - Legrand *et al.* (2007), 12,226 EST from root, leaf and cDNA libraries – Dauchot *et al.* (2009)) and the Genome Mapping Project white paper for the *Compositae* (*Asteraceae*) family (<http://compgenomics.ucdavis.edu/cwp/draft.php>) from which a large set of EST sequences have already been released: 41,747 EST or 22,038 Unigenes for *Cichorium intybus*, and 30,171 EST or 18,951 Unigenes for the related species *Cichorium endivia* (last update Feb-07). The screening of SSR motifs present in EST sequences (EST-SSR) is an efficient alternative to genomic library construction (costly and time-consuming) to identify EST-SSR markers (Varshney *et al.* 2005). In most of the plant species, SSR with a minimum repeat length of 20 bp are present in ~ 5% EST, with trinucleotides as preponderant motifs followed by dinucleotides or tetranucleotides (Varshney *et al.* 2005). Although the level of allelic informativeness was usually thought lower in EST-SSR than in enriched-libraries-derived SSR, a comparison in sugar beet showed more polymorphism for EST-SSR and emphasised their utility in genetic mapping (Laurent *et al.* 2007).

Systematic comparisons of homologous regions in our maps revealed major features: the presence of common SDR regions in four LG and their eventual connection to reproductive barriers and fitness in chicory, a good level of global coverage between maps, homologous regions with localized reduction of recombination.

Mapping distorted markers is often difficult. Unlike Cervera *et al.* (2001) who rejected high distorted loci from the mapping process, our strategy consisted in introducing all markers (distorted and not distorted) and comparing the resulting orders among the three maps. This so-called 'comparative mapping' (Song *et al.* 2006) is an intuitive and simple way already used by Cloutier *et al.* (1997) and Lashermes *et al.* (2001). According to Song *et al.* (2006), segregation distortion of clustered markers can be regarded as the selective effect of specific loci (SDR) which are able to affect the product of the meiosis (gametic selection) or the zygote after fertilization (zygotic selection). Sequential  $\chi^2$  tests described in Lorieux *et al.*

(1995) allowed the distinction between both types of selection. In our data (SDR in LG4, LG7, LG8 and LG9), zygotic and gametic factors explained each half of the segregation distortion observed. Excess of heterozygotes were the main characteristics of markers affected by zygotic selection, as revealed on SDR of LG4 and LG8. Similar results were already found in lettuce (Kesseli *et al.* 1994), coffee (Lashermes *et al.* 2001), alfalfa (Kalo *et al.* 2000) and white clover (Jones *et al.* 2003). In an outbred species like chicory, reducing deleterious alleles and increasing heterozygosity would substantially reduce fitness depression and contribute to the performance of hybrid genotypes.

Lu *et al.* (2002) showed, in  $F_2$  progenies of maize, that a gametophytic factor occurring in gametes of one sex only displayed an excess of the homozygous class that was compensated by the deficiency of the other homozygous class, the heterozygous class frequency remaining close to 0.50. Such segregation affected by gametic selection was observed for markers in LG7 (Rubis118). The gametic SDR of LG9 was more in agreement with selection affecting gametes in both parents.

The origin of selection factors leading to non-Mendelian segregation of SDR have been listed in recent reviews (Korbecka *et al.* 2002, Song *et al.* 2006). A self-incompatibility system and description of pollen competition before fertilization were reported in chicory (Eenink 1982, Desprez *et al.* 1994) and can be considered as potential explanatory factors associated to distorted genomic regions. But, numerous other genetic factors such as meiotic drive, selective germination or seedling death (Korbecka *et al.* 2002), structural chromosomal rearrangements (Fauré *et al.* 1993, Quillet *et al.* 1995) or negative epistatic interactions (Torjek *et al.* 2006) could also explained the presence of SDR.

The expected genome coverage in our data was divided in range of values from the complete set of markers and values from the framework marker set. As explained earlier, we have adopted the principle of framework markers selection that fitted adequately the hypothesis of uniform distribution of markers necessary for the estimation of genome coverage parameters of Chakravarti *et al.* (1991), Fishman *et al.* (2001) and Lange & Boehnke (1982). Our genome coverage estimates for the framework marker set (932-1126 cM,) were different from estimates calculated by de Simone *et al.* 1997 (1405 cM) and van Stallen *et al.* 2003 (730 cM). Such discrepancies in map length were also found in other plants. The related species

*Lactuca* had also variable genome sizes when comparing the map length in Jeuken *et al.* (2001) (854 cM), Truco *et al.* (2007) (1505 cM) or Kesseli *et al.* (1994) (1950 cM).

If we exclude errors due to mis-scoring of markers, these differences are thought to mainly result from variable recombination rates in relation to the genetic context. Further, genome map length could be overestimated because of differences in calculated recombination fractions and the expected number of crossovers (up to 20% in Fishman *et al.* 2001). Finally, Yin *et al.* (2004) and Yan *et al.* (2005) have suggested that the estimation parameter could be biased itself and preferred a procedure of estimation based on multiple LOD cut-offs for the former or a procedure based on repeated marker samplings for the latter. When the maps are densely covered, the finding of the ends of each LG by linking terminal markers with telomere-specific sequences would contribute to solve the question about the limits of the genetic map (Hu 2006).

Clustering of markers in genetic maps have often been reported in many species like lettuce (Jeuken *et al.* 2001), tomato (Areshchenkova & Ganal 1999) or barley (Ramsay *et al.* 2000), although the uneven distribution of markers seemed less evident in sunflower (Tang *et al.* 2002) or melon (Gonzalo *et al.* 2005). We have adapted the method described in Remington *et al.* (1999) to assess the level in the distribution of markers between the LG of the maps (ILGC test) and inside each LG (WLGC test). ILGC test showed a significant level of concentration of markers in a limited number of LG per map. Interestingly, WLGC test revealed a large number of regions with clustered markers, and particularly 13 clustered homologous regions for at least 3 maps dispersed over the 9 LG. According to this last result, chicory should be associated to the group of species inhibiting specific marker clustering. Inhibition of recombination has been described in numerous genetic maps using different marker systems (RFLP, AFLP, SSR). The non random distribution of enriched-libraries SSR compared to EST-SSR was recently published in sugar beet (Laurent *et al.* 2007). Enriched-libraries SSR also tend to be concentrated in retrotransposons and repetitive elements-rich regions in barley (Ramsay *et al.* 1999) and were, more generally, associated to centromeric regions (Areshchenkova & Ganal 1999, Ramsay *et al.* 2000). The association between centromeric regions and inhibition of recombination could be postulated in chicory, particularly for LG2, LG3, LG4, LG6,

LG7 and LG9 that shared common chromosomal regions with homologous clustered markers in all maps. Such hypothesis will have to be tested genetically with the introduction of centromeric-specific sequences in the maps or physically by FISH after selecting one or several SSR as probes.

The reference map of chicory presented here opens new perspectives in different directions such as the knowledge of the chicory genome, genetic relationships with the other *Asteraceae*, gene tagging and genetic analysis of quantitative-inherited traits, marker-assisted selection... The recent advent of several thousands of chicory EST will help us to enrich the map by integrating a large number of sequences representing genes, and initiate gene candidate approaches when genes are co-located with QTL. Further, mapping EST appears very promising since the PCR amplification of Conserved Orthologous Set (COS) of EST from two 'major' species in the *Asteraceae* family (lettuce and sunflower) succeeded for several other *Asteraceae*, such as chicory, thus opening perspectives for comparative genetic analyses (Chapman *et al.* 2007). However, a recent study showed that the transferability of sunflower EST-SSR and INDELs within the *Asteraceae* was limited to 18% when tested in lettuce or prickly lettuce, two species of the *Cichoriodeae* family (Heesacker *et al.* 2008).

The development of a dense genetic map as well as physical resources is also useful to specify the relationships between genetic and physical maps in chicory. van Stallen *et al.* (2003) have taken the observations of Dujardin *et al.* (1979) on the size of chicory chromosomes to establish a relationship with the size of their LG. Briefly, three LG larger than 80 cM were associated to the four larger chromosomes of Dujardin *et al.* (1979), and four LG smaller than 60 cM were associated to the five smaller chromosomes of Dujardin *et al.* (1979). Ge *et al.* (2002) had also identified in the karyotype of the forage chicory Puna, four chromosomes A, B, C, D ranging from 6.33 to 5.67  $\mu\text{m}$  and five chromosomes E, F, G, H, I ranging from 4.62 to 3.04  $\mu\text{m}$ . If the genetic lengths and the physical sizes of chicory chromosomes are correlated, our set of larger LG (LG1, LG3, LG6 and LG8) could agree with A, B C or D chromosomes, and our set of smaller LG (LG2, LG4, LG5, LG7, LG9) could agree with E, F, G, H or I chromosomes. Without any additional information, this relationship is still just a hypothesis. It will be established only when the 9 LG identified in the genetic maps will be assigned to chromosomes through *in situ*

hybridisation protocols using mapped markers as probes combined to protocols for general identification of chromosomes (banding, painting...). Finally, the development of a BAC library in chicory and the alignment of BAC clones with specific regions on the maps will complete EST or QTL mapping through the initiation of map-based gene cloning strategies.

A set of SSR markers are available for academic research applications with Materials Transfer Agreement. Applications should be sent to Dr Bruno Desprez,, Florimond Desprez, BP41, 3 rue Florimond Desprez, 59242 Cappelle en Pévèle, France

## Acknowledgements

The present work was granted through a public - private partnership 'Groupement d'Intérêt Scientifique' CARTOCHIC involving Université des Sciences et Technologies de Lille, Florimond Desprez Veuve & Fils SAS, Leroux SAS and SARL Hoquet Endives. We would like to thank Louis Delesalle, Jean-Christophe Lepeltier, Pierre Devaux (Florimond-Desprez) and Pascal Lévêque (Hoquet Endives) who have generated and maintained the Rubis118 and BR populations. We also acknowledge Aude Darracq for the development of a program devoted to the identification of unique SSR sequences. Helpful comments and critical reading of the manuscript by Bruno Desprez and Pierre Devaux have also been greatly appreciated.

## Appendix

- 1) Consider one marker m1 with four alleles (a, b, c, d) in the four-allelic configuration type in the K28K59 population *i.e* ab x cd. Let p, q, r, and s the gametic frequency of a, b, c, and d alleles, respectively.

The  $H_0$  hypothesis  $p = q = r = s$  is tested by:

$$\chi^2_{(3df)} = \frac{(np-n/2)^2 + (nq-n/2)^2 + (nr-n/2)^2 + (ns-n/2)^2}{n/2} \quad (1)$$

n = size of the population

The  $H_0$  hypothesis of random assortment of gametes leading to the following distribution of phenotypes  $pq:qr:ps:qs$  for genotypic classes  $ac:bc:ad:bd$  is tested by:

$$X^2_{(3df)} = \frac{(n_{ac}-npq)^2}{npq} + \frac{(n_{ad}-nps)^2}{nps} + \frac{(n_{bc}-nqr)^2}{nqr} + \frac{(n_{bd}-nqs)^2}{nqs} \quad \text{or}$$

$$X^2_{(3df)} = \frac{n_{ac}^2}{npq} + \frac{n_{ad}^2}{nps} + \frac{n_{bc}^2}{nqr} + \frac{n_{bd}^2}{nqs} - n \quad (2)$$

$n_{ac}$ ,  $n_{ad}$ ,  $n_{bc}$  and  $n_{bd}$  are the observed number of genotypes in the  $ac$ ,  $ad$ ,  $bc$  and  $bd$  classes, respectively

2) Consider one marker  $m_2$  with three alleles ( $a$ ,  $b$ ,  $c$ ) in the three-allelic configuration type in the K28K59 population *i.e*  $ab \times ac$ . Let  $p$ ,  $q$  and  $r$  the gametic frequency of  $a$ ,  $b$  and  $c$  alleles, respectively.

The new expressions of (1) and (2) are:

$$X^2_{(2df)} = \frac{(2np-n)^2}{n} + \frac{(nq-n/2)^2}{n/2} + \frac{(nr-n/2)^2}{n/2} \quad (1)$$

$n$  = size of the population

$$X^2_{(3df)} = \frac{n_{aa}^2}{np^2} + \frac{n_{ab}^2}{npq} + \frac{n_{ac}^2}{npr} + \frac{n_{bc}^2}{nqr} - n \quad (2)$$

$n_{aa}$ ,  $n_{ab}$ ,  $n_{ac}$  and  $n_{bc}$  are the observed number of genotypes in the  $aa$ ,  $ab$ ,  $ac$  and  $bc$  classes, respectively

For all tests, it follows that:

- whatever the significance of (1) is, if (2) is significant, then the selection will be

zygotic

- when (1) is significant and (2) is non significant, then the selection will be gametic.

The type of selection cannot be estimated with previous successive  $\chi^2$  tests for dominant  $F_2$  markers and markers in backcross configuration type, because the selection would always appear to be of gametic type, the test equivalent to expression (2) would be always equal to zero (Lorieux *et al.* 1995).

## References

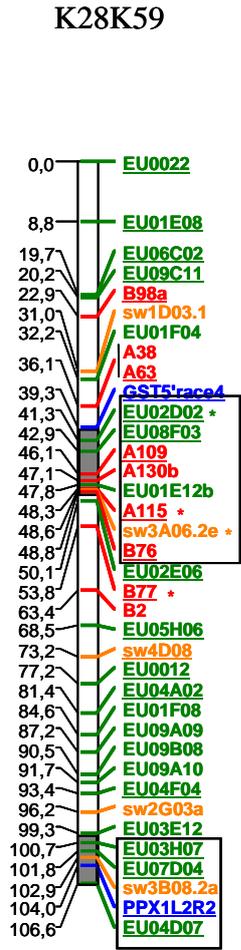
- Andersen PS, Jespersgaard C, Vuust J, Christiansen M, Larsen LA (2003) High-throughput single strand conformation polymorphism mutation detection by automated capillary array electrophoresis: validation of the method. *Human Mutat* 21: 116-122
- Areshchenkova T, Ganai MW (1999) Long tomato microsatellites are predominantly associated with centromeric regions. *Genome* 42
- Aubert G, Morin J, Jacquin F, Loridon K, Quillet M-C, Petit A, Rameau C, Lejeune-Hénaut I, Huguet T, Burstin J (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. *Theor Appl Genet* 112: 1024-1041
- Bannerot H, de Coninck B (1970) L'utilisation des hybrides F<sub>1</sub> : nouvelle méthode d'amélioration de la chicorée de Bruxelles, In: (eds) Section horticole d'Eucarpia. La chicorée de Bruxelles. Gembloux, Belgique, pp 99-112
- Barcaccia G, Pallottini L, Soattin M, Lazzarin R, Parrini P, Lucchin M (2003) Genomic DNA fingerprints as a tool for identifying cultivated types of radicchio (*Cichorium intybus* L.) from Veneto, Italy. *Plant Breeding* 122: 178-183
- Bellamy A, Vedel F, Bannerot H (1996) Varietal identification in *Cichorium intybus* L. and determination of genetic purity of F<sub>1</sub> hybrid seed samples, based on RAPD markers. *Plant Breed* 115: 128-132
- Bremer K (1994) *Asteraceae: Cladistics and classification*, Timber Press, 133 SW South West, suite 450, Portland, OR, 97204-3527, USA, pp 752
- Cervera M-T, Storme V, Ivens B, Gusmao J, Liu BH, Hostyn V, van Slycken J, van Montagu M, Boerjan W (2001) Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. *Genetics* 158: 787-809
- Chakravarti A, Lasher LK, Reefer JE (1991) A maximum likelihood method for estimating genome length using genetic linkage data. *Genetics* 128: 175-182
- Chapman M, Chang J, Weisman D, Kesseli R, Burke J (2007) Universal markers for comparative mapping and phylogenetic analysis in the *Asteraceae* (*Compositae*). *Theor Appl Genet* 115: 747-755
- Cloutier S, Cappadocia M, Landry BS (1997) Analysis of RFLP mapping inaccuracy in *Brassica napus* L. *Theor Appl Genet* 95: 83-91
- Dauchot N, Mingeot D, Purnelle B, Muys C, Watillon B., Boutry M, Van Cutsem P (2009) 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
- de Simone M, Morgante M, Lucchin M, Parrini P, Marocco A (1997) A first linkage map of *Cichorium intybus* L. using a one-way pseudo-testcross and PCR-derived markers. *Mol Breed* 3: 415-425
- Desprez B, Delesalle L, Dhellemmes C, Desprez M (1994) Genetics and breeding of industrial chicory. *C R Acad Agric Fr* 80: 47-62
- Desprez FF, Bannerot H (1980) A study of pollen tube growth in witloof chicory, In: Maxon Smith JW and Langton A (eds) *Proceedings Eucarpia Meeting on Leafy Vegetables*. Littlehampton, UK, pp 47-52
- Dice LR (1945) Measures of the amount of ecologic association between species. *Ecology* 26:297-302
- Dujardin M, Louant BP, Tilquin JP (1979) Determining the caryogram of *Cichorium intybus* L. *Ann Amélior Plantes* 29: 305-310
- Eenink AH (1980) Breeding research on witloof chicory for the production of inbred lines and hybrids, In: Maxon Smith JW and Langton A (eds) *Proceedings Eucarpia Meeting on Leafy Vegetables*. Littlehampton, UK, pp 5-11
- Eenink AH (1981) Compatibility and incompatibility in witloof-chicory (*Cichorium intybus* L.). 2. The incompatibility system. *Euphytica* 30: 77-85

- Eenink AH (1982) Compatibility and incompatibility in witloof-chicory (*Cichorium intybus* L.). 3. Gametic competition after mixed pollinations and double pollinations. *Euphytica* 31: 773-786
- Fauré S, Noyer JL, Horry JP, Bakry F, Lanaud C, Gonzalez de Leon D (1993) A molecular marker-based linkage map of diploid bananas (*Musa acuminata*). *Theor Appl Genet* 87: 517-526
- Fishman L, Kelly AJ, Morgan E, Willis JH (2001) A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. *Genetics* 159: 1701-1716
- Gautschi B, Tenzer I, Muller JP, Schmid B (2000a) Isolation and characterization of microsatellite loci in the bearded vulture (*Gypaetus barbatus*) and cross-amplification in three Old World vulture species. *Mol Ecol* 9: 2193-2195
- Gautschi B, Widmer A, Koella J (2000b) Isolation and characterization of microsatellite loci in the dice snake (*Natrix tessellata*). *Mol Ecol* 9: 2192-2193
- Ge RC, Zhao ML, Gao HW, Zhao BC, Li GL (2002) Study on C-banding and karyotype analysis of Puna chicory. *Acta Agrest Sin* 10:190-193
- Glenn TC, Schable NA (2005) Isolating microsatellite DNA loci. *Methods Enzymol* 395: 202-222
- Gonzalo M, Oliver M, Garcia-Mas J, Monfort A, Dolcet-Sanjuan R, Katzir N, Arús P, Monforte A (2005) Simple-sequence repeat markers used in merging linkage maps of melon (*Cucumis melo* L.). *Theor Appl Genet* 110: 802-811
- Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137: 1121-1137
- Heesacker A, Kishore VK, Gao W, Tang S, Kolkman JM, Gingle A, Matvienko M, Kozik A, Michelmore RM, Lai Z, Rieseberg LH, Knapp SJ (2008) SSRs and INDELS mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *Theor Appl Genet* 117: 1021-1029
- Hu J (2006) Defining the sunflower (*Helianthus annuus* L.) linkage group ends with the Arabidopsis-type telomere sequence repeat-derived markers. *Chrom Res* 14: 535-548
- Jeuken M, van Wijk R, Peleman J, Lindhout P (2001) An integrated interspecific AFLP map of lettuce (*Lactuca*) based on two *L. sativa* x *L. saligna* F<sub>2</sub> populations. *Theor Appl Genet* 103: 638-647
- Jung C, Pillen K, Frese L, Fähr S, Melchinger AE (1993) Phylogenetic relationships between cultivated and wild species of the genus *Beta* revealed by DNA "fingerprinting". *Theor Appl Genet* 86: 449-457
- Jones ES, Hughes LJ, Drayton MC, Abberton MT, Michaelson-Yeates TPT, Bowen C, Forster JW (2003) An SSR and AFLP molecular marker-based genetic map of white clover (*Trifolium repens* L.). *Plant Sci* 165: 531-539
- Kalo P, Endre G, Zimanyi L, Csanadi G, Kiss GB (2000) Construction of an improved linkage map of diploid alfalfa (*Medicago sativa*). *Theor Appl Genet* 100: 641-657
- Kaur N, Gupta AK (2002) Applications of inulin and oligofructose in health and nutrition. *J Biosci* 27: 703-714
- Kenis K, Keulemans J (2005) Genetic linkage maps of two apple cultivars (*Malus x domestica* Borkh.) based on AFLP and microsatellite markers. *Mol Breed* 15: 205-219
- Kesseli RV, Paran I, Michelmore RW (1994) Analysis of a detailed genetic linkage map of *Lactuca sativa* (lettuce) constructed from RFLP and RAPD markers. *Genetics* 136: 1435-1446
- Kiers AM, Mes THM, vanderMeijden R, Bachmann K (2000) A search for diagnostic AFLP markers in *Cichorium* species with emphasis on endive and chicory cultivar groups. *Genome* 43: 470-476
- Koch G, Jung C (1997) Phylogenetic relationships of industrial chicory varieties revealed by RAPDs and AFLPs. *Agronomie* 17: 323-333
- Kosambi DD (1944) The estimation of map distance from recombination values, *A Eug*, 12, 172-175

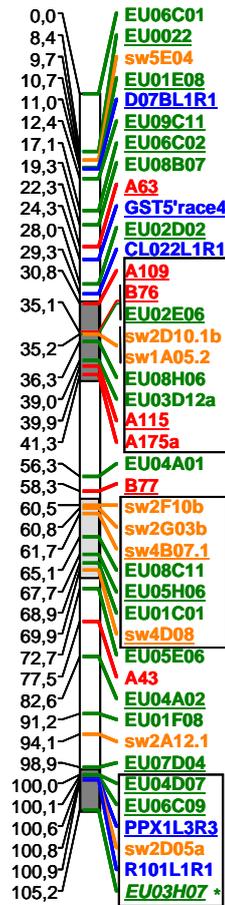
- Korbecka G, Klinkhamer PGL, Vrieling K (2002) Selective embryo abortion hypothesis revisited - A molecular approach. *Plant Biol (Stuttg)*: 298-310
- Lange K, Boehnke M (1982) How many polymorphic genes will it take to span the human genome? *Am J Hum Genet*. 34: 842-845.
- Lashermes P, Combes M-C, Prakash NS, Trouslot P, Lorieux M, Charrier A (2001) Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meioses. *Genome* 44: 589-596
- Laurent V, Devaux P, Thiel T, Viard F, Mielordt S, Touzet P, Quillet M-C (2007) Comparative effectiveness of sugar beet microsatellite markers isolated from genomic libraries and GenBank ESTs to map the sugar beet genome. *Theor Appl Genet* 115: 793-805
- Legrand S, Hendriks T, Hilbert J-L, Quillet M-C (2007) Characterization of expressed sequence tags obtained by SSH during somatic embryogenesis in *Cichorium intybus* L. *BMC Plant Biology* 7: 27
- Li G, Kemp PD (2005) Forage chicory (*Cichorium intybus* L.): A review of its agronomy and animal production, In: Sparks DL (eds) *Advances in Agronomy* vol 88. Academic Press, pp 187-222
- Lorieux M, Perrier X, Goffinet B, Lanaud C, Gonzales de Leon D (1995) Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F<sub>2</sub> populations. *Theor Appl Genet* 90: 81-89
- Lu H, Romero-Severson J, Bernardo R (2002) Chromosomal regions associated with segregation distortion in maize. *Theor Appl Genet* 105: 622-628
- Maliepaard C, Jansen J, van Ooijen JW (1997) Linkage analysis in a full-sib family of an outbreeding plant species : overview and consequences for applications. *Genet Res, Camb* 70: 237-250
- Morgante M, Olivieri A (1993) PCR-amplified microsatellites as markers in plant genetics. *The Plant J* 3: 175-182
- Panero JL, Funk VA (2002) Toward a phylogenetic subfamilial classification for the *Compositae*. *Proc Biol Soc Wash* 115:909-922
- Pécaut P (1962) Etude sur le système de reproduction de l'endive (*Cichorium intybus* L). *Ann Amélior Plantes* 12: 265-296
- Perrier X, Jacquemoud-Collet JP (2006) DARwin software <http://darwin.cirad.fr/darwin>
- Pool-Zobel BL (2005) Inulin-type fructans and reduction in colon cancer risk: review of experimental and human data. *Br J Nutr* 93: S73-S90
- Quillet MC, Madjidian N, Griveau Y, Serieys H, Tersac M, Lorieux M, Bervillé A (1995) Mapping genetic factors controlling pollen viability in an interspecific cross in *Helianthus* sect. *Helianthus*. *Theor Appl Genet* 91: 1195-1202
- Rambaud C, Dubois J, Vasseur J (1993) Male-sterile chicory cybrids obtained by intergeneric protoplast fusion. *Theor Appl Genet* 87: 347-352
- Ramsay L, Macaulay M, Cardle L, Morgante M, Ivanissevich Sd, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J* 17: 415-425
- Ramsay L, Macaulay M, Ivanissevich Sd, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuveesson S, Morgante M, Massari A, Maestri E, Marmioli N, Sjakste T, Ganai M, Powell W, Waugh R (2000) A Simple Sequence Repeat-based linkage map of barley. *Genetics* 156: 1997-2005
- Remington DL, Whetten RW, Liu B-H, O'Malley DM (1999) Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. *Theor Appl Genet* 98: 1279-1292
- Rick CM (1953) Hybridization between chicory and endive. *J Am Soc Hort Sci* 61: 459-466
- Sneath PHA, Sokal RR (1973) *Numerical taxonomy*, WH Freeman and Co, San Francisco
- Snedecor GW, Cochran WG (1989) *Statistical methods*, Eighth Edition, Iowa State University Press, pp 803

- Song X-L, Sun X-Z, Zhang T-Z (2006) Segregation distortion and its effect on genetic mapping in plants. *Chin J Agric Biotechnol* 3: 163-169
- Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? *Mol Ecol* 12: 1339-1348
- Tang S, Yu J-K, Slabaugh M, Shintani D, Knapp S (2002) Simple sequence repeat map of the sunflower genome. *Theor Appl Genet* 105: 1124-1136
- Torjek O, Witucka-Wall H, Meyer R, von Korff M, Kusterer B, Rautengarten C, Altmann T (2006) Segregation distortion in *Arabidopsis* C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. *Theor Appl Genet* 113: 1551-1561
- Truco M, Antonise R, Lavelle D, Ochoa O, Kozik A, Witsenboer H, Fort S, Jeuken M, Kesseli R, Lindhout P, Michelmore R, Peleman J (2007) A high-density, integrated genetic linkage map of lettuce (*Lactuca* spp.). *Theor Appl Genet* 115: 735-746
- van Cutsem P, du Jardin P, Boutte C, Beauwens T, Jacqmin S, Vekemans X (2003) Distinction between cultivated and wild chicory gene pools using AFLP markers. *Theor Appl Genet* 107: 713-718
- van Ooijen JW, Voorrips RE (2001) JoinMap 3.0, software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands
- van Stallen N, Noten V, Demeulemeester M, de Proft M (2000) Identification of commercial chicory cultivars for hydroponic forcing and their phenetic relationships revealed by random amplified polymorphic DNAs and amplified fragment length polymorphisms. *Plant Breed* 119: 265-270
- van Stallen N, Vandenbussche B, Londers E, Noten V, de Proft M (2005a) QTL analysis of important pith characteristics in a cross between two inbred lines of chicory (*Cichorium intybus* var. *foliosum*): a preliminary study. *Plant Breed* 124: 54-58
- van Stallen N, Vandenbussche B, Londers E, Noten V, de Proft M (2005b) QTL analysis of production and taste characteristics of chicory (*Cichorium intybus* var. *foliosum*). *Plant Breed* 124: 59-62
- van Stallen N, Vandenbussche B, Verdoodt V, de Proft M (2003) Construction of a genetic linkage map for witloof (*Cichorium intybus* L. var. *foliosum* Hegi). *Plant Breed* 122: 521-525
- Varotto S, Pizzoli L, Lucchin M, Parrini P (1995) The incompatibility system in Italian red chicory (*Cichorium intybus* L.). *Plant Breed* 114: 535-538
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23: 48-55
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 93: 77-78
- Wittop Koning, D A., Leroux A (1972) La chicorée dans l'histoire de la médecine et dans la céramique pharmaceutique. *Rev Hist Pharm (Suppl)* 215: 28p.
- Yan Z, Denneboom C, Hattendorf A, Dolstra O, Debener T, Stam P, Visser PB (2005) Construction of an integrated map of rose with AFLP, SSR, PK, RGA, RFLP, SCAR and morphological markers. *Theor Appl Genet* 110: 766-777
- Yin T, DiFazio S, Gunter L, Riemenschneider D, Tuskan G (2004) Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. *Theor Appl Genet* 109: 451-463
- Yu J-K, Tang S, Slabaugh MB, Heesacker A, Cole G, Herring M, Soper J, Han F, Chu W-C, Webb DM, Thompson L, Edwards KJ, Berry S, Leon AJ, Grondona M, Olungu C, Maes N, Knapp SJ (2003) Towards a saturated molecular genetic linkage map for cultivated sunflower. *Crop Sci* 43: 367-387
- Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. *Mol Ecol* 11: 1-16

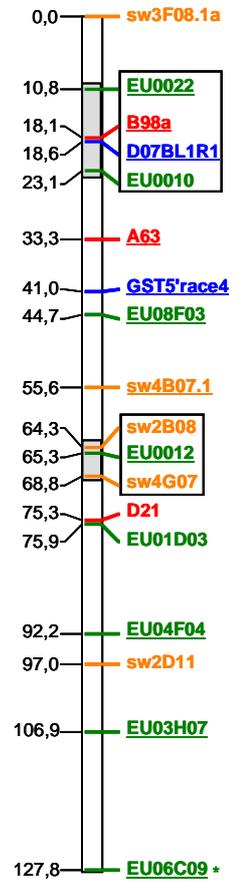
LG1



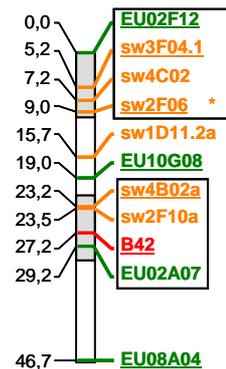
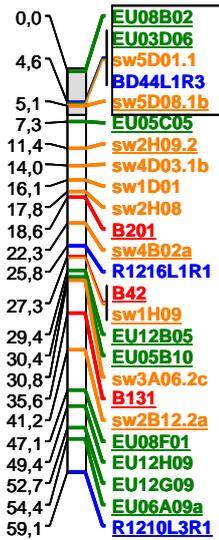
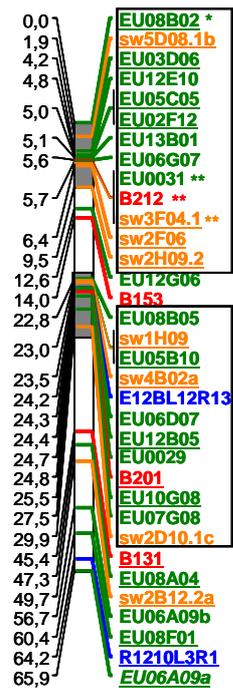
**Rubis 118**

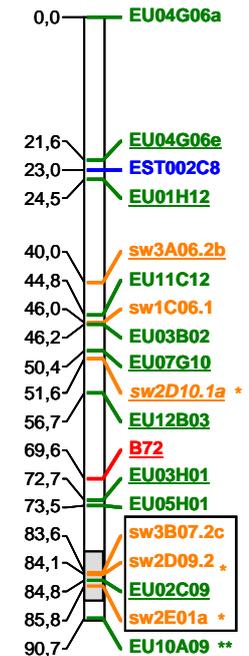
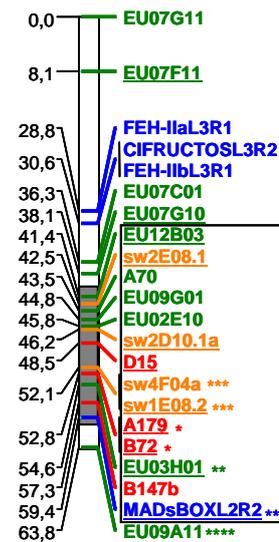
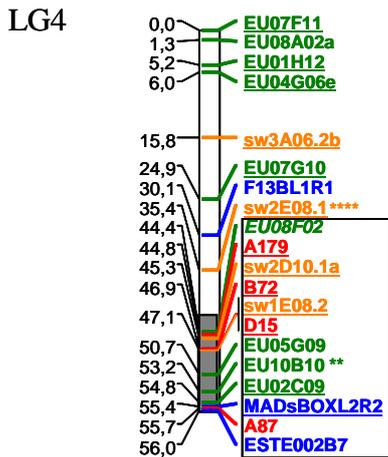
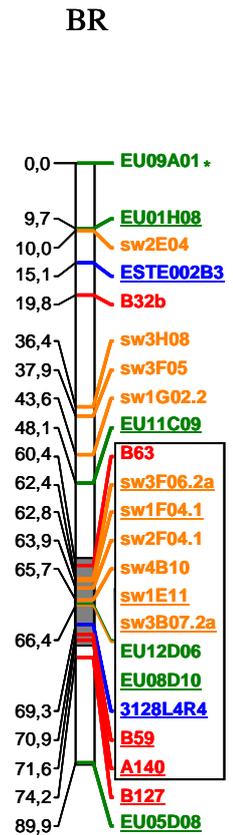
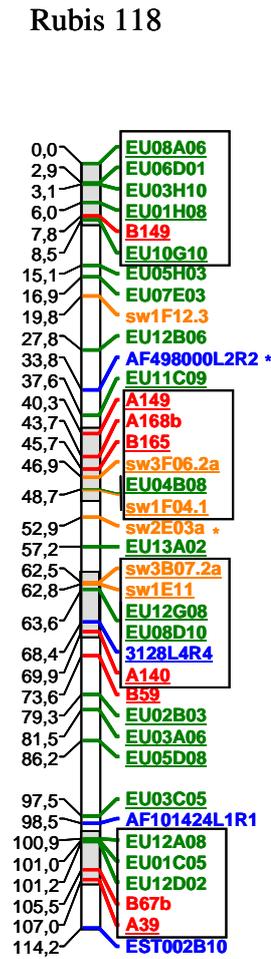
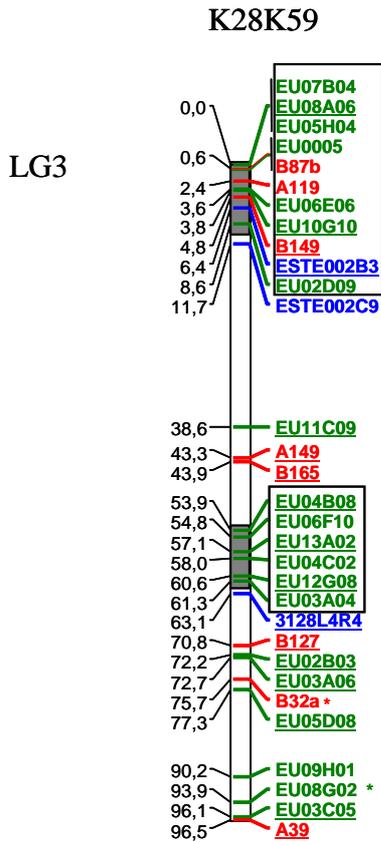


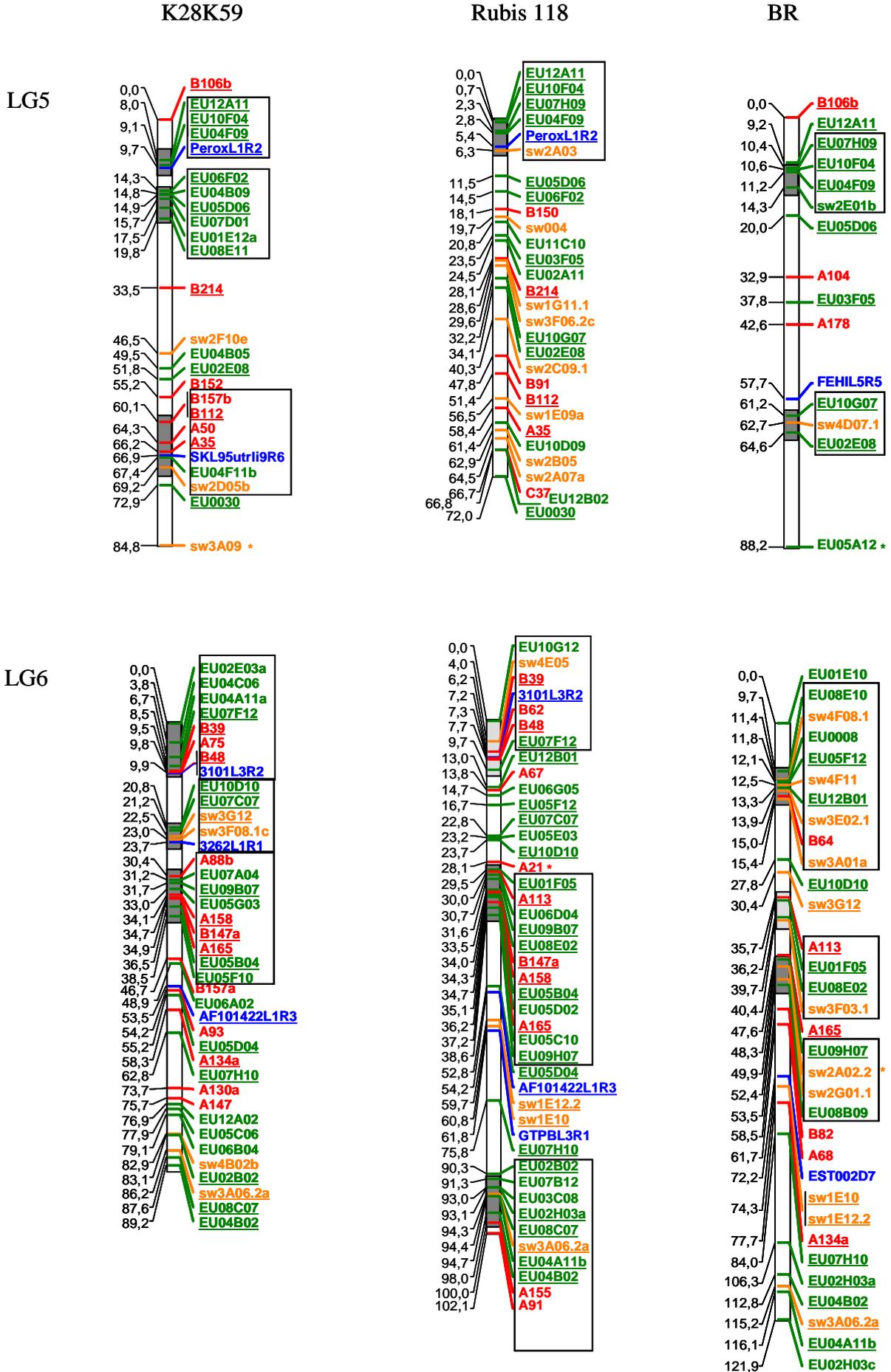
**BR**



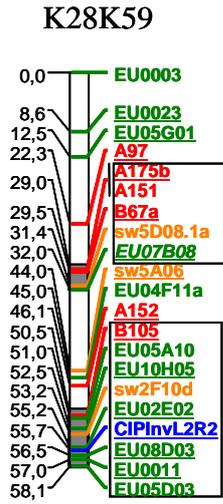
LG2



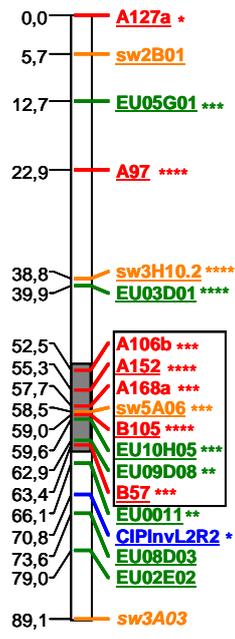




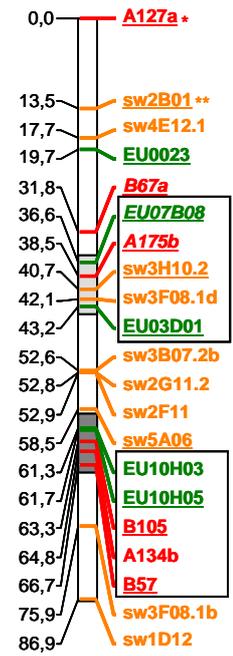
LG7



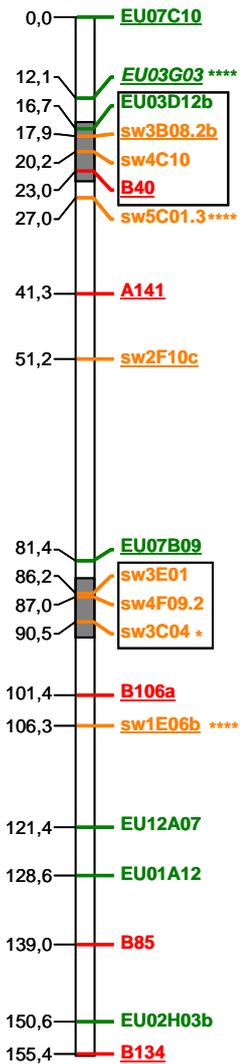
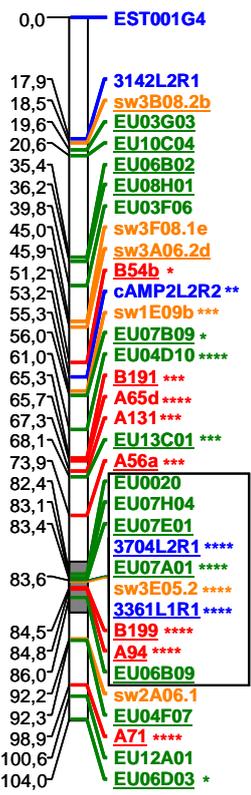
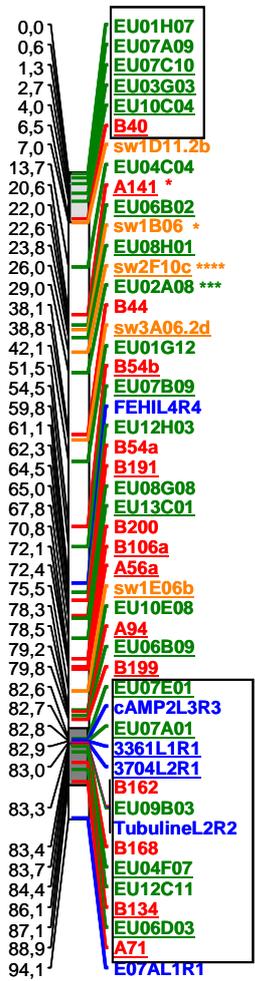
**Rubis 118**

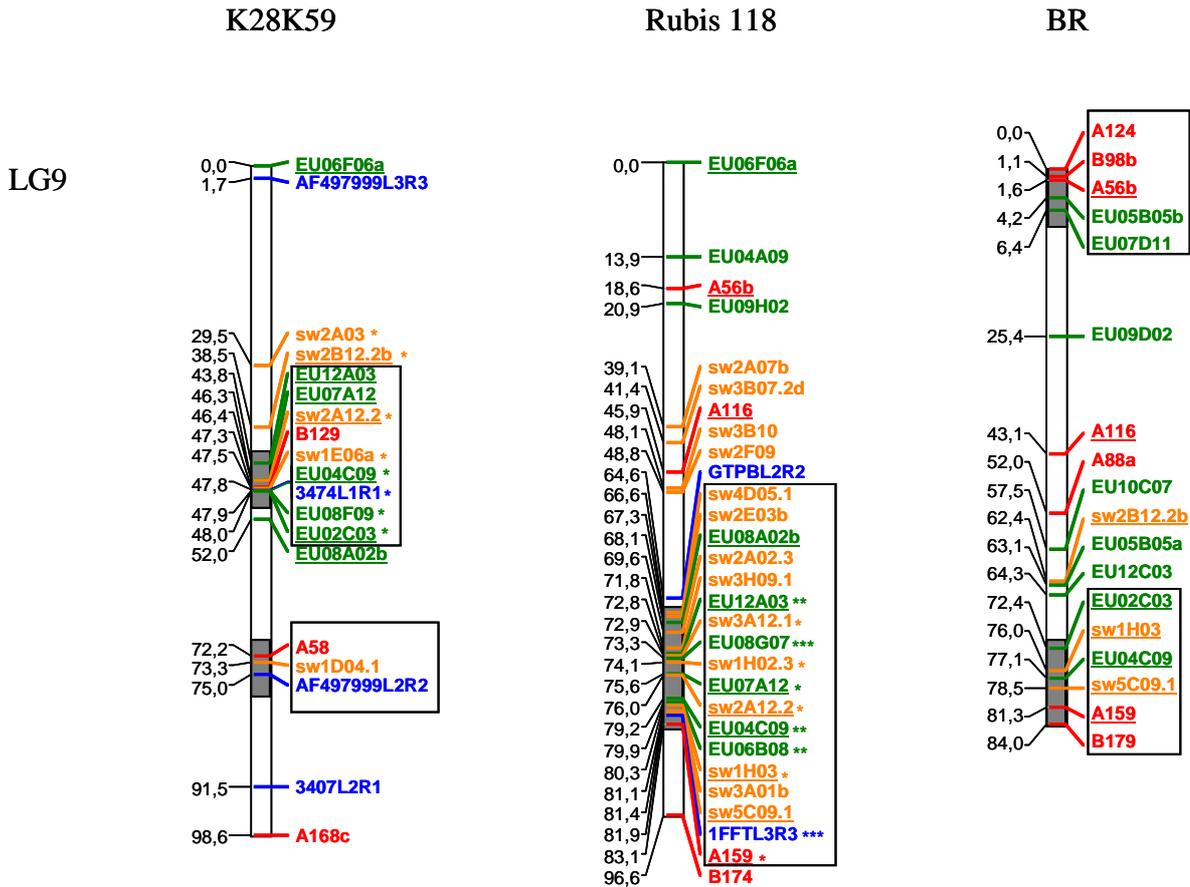


**BR**



LG8





**Figure 2:** Genetic maps of 3 chicory populations (K28K59, Rubis 118, BR). STS markers are blue-coloured, SSR markers are differentially coloured (red for AGR SSR, orange for SW SSR, green for EU SSR). Bridge markers are underlined. Skewed markers are followed by “\*” representing the probability associated to  $\chi^2$  tests of goodness-of-fit (\* P<0.05, \*\* P<0.01, \*\*\* P<0.005, \*\*\*\* P<0.0001). Marker clustering (WLGC) are visualized through grey patches inside LG and rectangles surrounding significant markers (dark grey : significant WLGC test, pale grey: WLGC test just below the threshold)

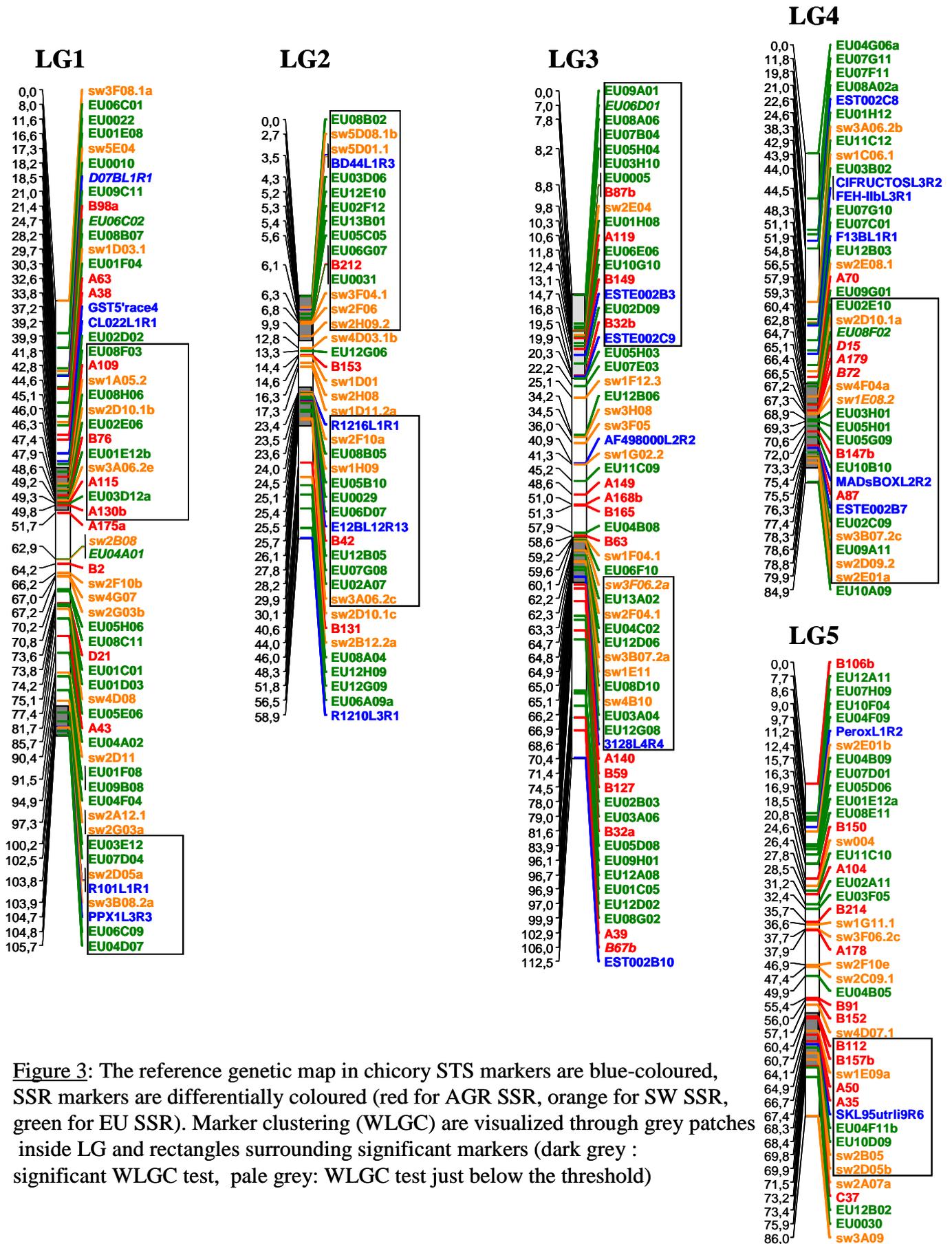
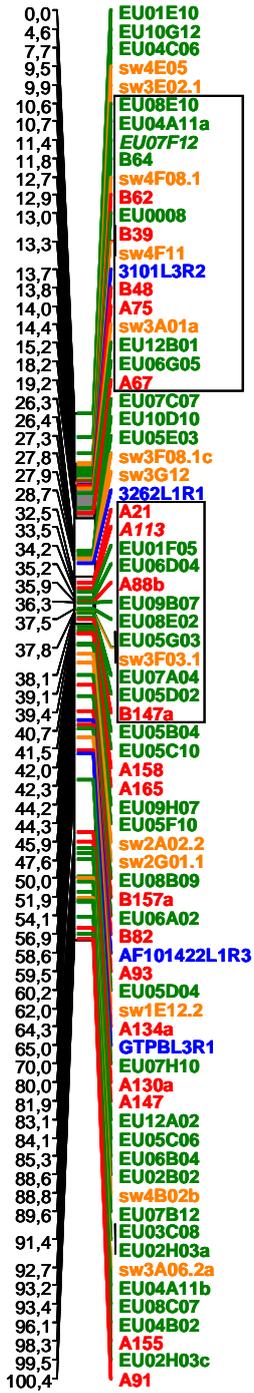
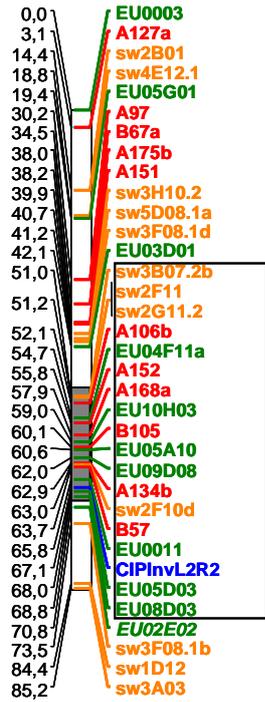


Figure 3: The reference genetic map in chicory STS markers are blue-coloured, SSR markers are differentially coloured (red for AGR SSR, orange for SW SSR, green for EU SSR). Marker clustering (WLGC) are visualized through grey patches inside LG and rectangles surrounding significant markers (dark grey : significant WLGC test, pale grey: WLGC test just below the threshold)

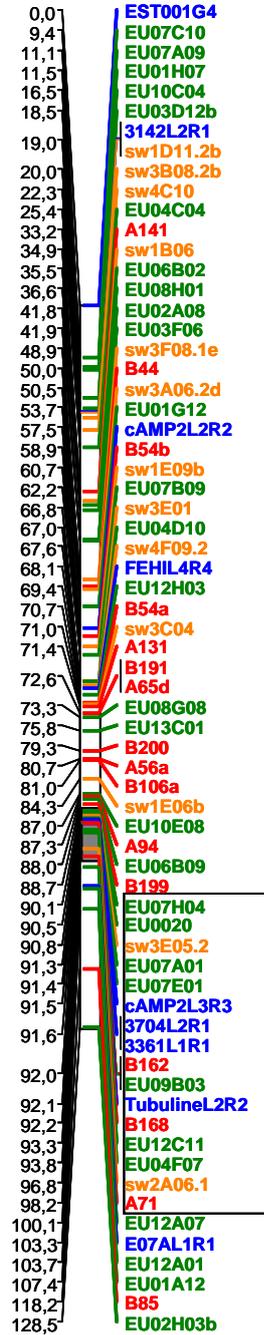
LG6



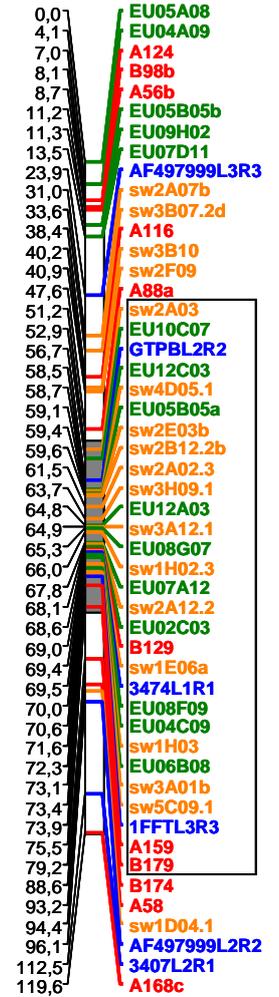
LG7



LG8



LG9



## Annexes

**Table 1:** List of mapped STS sequences. Locus were classified according to their assignment to specific linkage groups (LG). Primer pairs (For=Forward, Rev=Reverse primers) and T<sub>m</sub> are reported. For sequences listed in public databases, the accession number (<http://www.ebi.ac.uk>) is displayed as well as their putative function; The remaining sequences have been obtained in our laboratory and were noted as 'home' sequences.

Locus Name	Accession n° / home	Primer Pairs	T <sub>m</sub>	Function	LG
D07B L1/R1 <sup>c</sup>	home	For: ACGGGTGCTTTAGGAATC Rev: GCACGGGAAAAAAGAAAC	55°C	/	1
GST5'-Race4	home	For: ACTGGTCCACTGCCTC Rev: CAAAGCTCTTCCCATCAAA	53°C	glutathione S transferase	1
CL022 L1/R1	home	For: GTGAAGCGGTTGAAAGAGGA Rev: TTCGCAGCTTTTACATCACG	53°C	kinase protein	1
R101 L1/R1	home	For: TTCGCTTCAATCAATCGTC Rev: ATCCTCCACCCACCTAAAG	55°C	ribosomal protein 2	1
PPX1 L2/R2 <sup>a</sup>	AF160961	For: GTCTGGAATCTGGAGCAAAA Rev: CTCGTGAGCTTCCAACACAA	52°C	protoporphyrinogen IX oxydase	1
PPX1 L3/R3 <sup>a</sup>	AF160961	For: CAATGTCATCCACGAAGTCA Rev: GTCAATGGATCTTCCGCATC	52°C	protoporphyrinogen IX oxydase	1
BD44 L1/R3	home	For: CATCAGGACCACCGATAACC Rev: TCCCATTCTCTTTGCTCCAG	60°C	/	2
E12B L12/R13 <sup>c</sup>	home	For: CTCACAATGAGAAACAAGCA Rev: GCGTTCAAGTAAGTCCATTTC	55°C	/	2
R1216 L1/R1	home	For: GAAGCTCACCGGAGATTGAG Rev: GACACAAATGCCCTTCCATT	55°C	transketolase	2
R1210 L3/R1	home	For: ACAAAACATTGCTGCCAAGC Rev: CACTTTAACCGGGCTTTGAA	55°C	ferredoxin NADP reductase binding protein or oxygen evolving complex	2
ESTE002C9 <sup>b</sup>	DT212606	For: TTCACAAGGTGTGACGGTTG Rev: CAAAGCTTCTTCGGGTGCTA	60°C	haloacid dehalogenase-like hydrolase family protein	3
ESTE002B3 <sup>b</sup>	DT212592	For: GGTCAATAATCCTCGGCATA Rev: CAGCTCCAATGGTGGCTTTA	60°C	GTP binding protein	3
AF498000 L2/R2	AF498000	For: CGTTACTCCTTGGCAGCAAT Rev: TAGCTCGTCAATCGTTCCT	55°C	germacrene A synthase short form	3
3128 L4/R4 <sup>d</sup>	home	For: TCGAGAACCAACAACCCG Rev: AAACCGGATGCAATAACTCAA	60°C	/	3
AF101424 L1/R1	AF101424	For: GCACGAGGATTTCTTAAAGATG Rev: TGCACATGCTTACCATCAAG	50°C	pyrroline-5-carboxylate synthetase	3
ESTE002B10 <sup>b</sup>	DT212598	For: TCCTTCGTCTCACTTCCATGA Rev: ATTGTCCACCACTTCGAACC	60°C	peroxidase, putative	3
FEHIIb L3/R1	AJ295034	For: CTTTCGTCAAATGAAGTCAACC Rev: AAAACCTCTGCGTCTTTCAACT	59°C	fructan 1 -exohydrolase IIb (1-FEHIIb)	4
FEHIIa L3/R1	AJ295033	For: GGATCGACAGAAATGGGAAA Rev: AAACCTCTGCCTCTTTCAACC	59°C	fructan 1 -exohydrolase IIa (1-FEHIIa)	4
CIFRUCTOS L3/R2	Y11176	For: CGACAAGTTATGGTGCATTG Rev: CGAAATTTGACATTTGAGTCC	55°C	fructosidase	4
ESTE002C8 <sup>b</sup>	DT212605	For: TACCGTGGGTAACAAACA Rev: TTCAGCTTTCGTCCTTTGAA	60°C	glycosyl hydrolase family 1 protein	4
F13B L1/R1 <sup>c</sup>	home	For: GCATGTATGGTCTCAAAAA Rev: CACGAGGACATCTCAAATA	55°C	/	4
MADs box L2/R2	AF101420	For: AAAGTTGACGAACCTTCTCTCG Rev: AAGCCTTGAAGGTCCTCTCC	55°C	MADS box protein	4
ESTE002B7 <sup>b</sup>	DT212595	For: TCAACATTTGAGGTCGGTGA Rev: TATGCGCACTTGCAGGTAAC	60°C	amino acid permease 6	4
perox L1/R2	AF101427	For: TGGCTCAGAAACATCTTCCA Rev: ATACAACCTCCAGGGCACTCC	52°C	peroxydase	5
FEH-I L5/R5	AJ242538	For: CTACGGAGCTTTTCTTGATT Rev: AATTCCTTCTCCACCAAAAC	54°C	fructan-exohydrolase I (1-FEH-I) exo-inulinase	5

## Annexes

**Table 1:** (continued)

<b>Locus Name</b>	<b>Accession n° / home</b>	<b>Primer Pairs</b>	<b>Tm</b>	<b>Function</b>	<b>LG</b>
SKL95UTR9L19R6	home	For: TCTGGTTGAGTAATCGATG Rev: TTGCAAAATATTATTCCGATCG	55°C	serine kinase protein	5
3101 L3/R2	home	For: CATCCTACCTGCCACATTG Rev: TCGGATCCATTACAAGGTACA	55°C	/	6
3262 L1/R1	home	For: AGCTGCAAAACCCCATTA Rev: CTAAAGCCCAAGCTTTGCT	55°C	dehydrin	6
GTPB L3/R1 <sup>d</sup>	AJ249865	For: GCGGGTCAAGAAAGGTA Rev: TTTTAGCCCGTGTAATGAA	58°C	GTP binding protein (GTP1)	6
ESTE002D7 <sup>b</sup>	DT212613	For: AGACATGCCTTGCCAAAGATT Rev: CGAGCATACAGAGCTCCTTTC	60°C	phosphatidylinositol-4-phos.5-kinase family protein	6
AF101422 L1/R3	AF101422	For: GTCACACCGTCGAGAAAGG Rev: TATCCAGCAGTCGTTCCAGA	59°C	cytochrome	6
CIP invert L2/R2	Y11124	For: AAGAATGGACGATATGTTGTGCT Rev: AATGGCCAATGTTGGGTAAA	55°C	putative invertase	7
3142 L2/R1	home	For: GCGTTTGCTTCCCTTATTC Rev: TGCTCCATCACCTTCATCAG	52°C	HMG CoA synthase	8
ESTE001G4 <sup>b</sup>	DT212558	For: GTCCGACCCAAAAGATCCA Rev: AGCACCAGAGACTCCCATGT	60°C	transmembrane protein B	8
cAMP2 L2/R2	AF067186	For: GAGCCAACCTTGAGCAAGAGC Rev: GAAGTTCACCATCAGCAGCA	55°C	cAMP responsive element binding protein	8
cAMP2 L3/R3	AF067186	For: CCTGCAGAACGGTGTCTTCT Rev: GCGTCTTCTGCTTGTGTGA	53°C	cAMP responsive element binding protein	8
FEH-I L4/R4	AJ242538	For: GGCTTCCACTTCCAGCCT Rev: GAAATGTTGCCCCAGAGTG	52°C	fructan-exohydrolase I (1-FEH-I) exo-inulinase	8
3704 L2/R1	home	For: TGGAAAATGTGGTGCAACT Rev: TTATGACTTTATTGCCGTGG	56°C	Lipid Transfert Protein 2	8
3361 L1/R1	home	For: GCGGGTATGCTGATGAAAAT Rev: TCGTCTCCCTTTATCCGTTG	53°C	Lipid Transfert Protein 1	8
Tubulin L2/R2 <sup>d</sup>	AF101419	For: CCGAAAATTGTGATTGCTT Rev: CAGAAGTGTCCCATACCAG	55°C	$\pi$ tubulin	8
E07A L1/R1 <sup>c</sup>	home	For: TGCACTATTAATCACCAGCA Rev: GAAACTTCATTCATACCCCA	55°C	/	8
3407 L2/R1	home	For: TTCTTGGTTGGTGGAGCTT Rev: GCATTAGTGCAGAGTTATTGACCA	50°C	triose P isomérase	9
AF497999 L2/R2	AF497999	For: CAAGCCATGAGTCGGTTGTA Rev: ACGTGCCCGAGAGTAATACG	55°C	germacrene A synthase long form	9
GTPB L2/R2 <sup>d</sup>	AJ249865	For: GGAGCTGGTAAATCAAGTCTGGT Rev: GATACCTTCTTGACCCGCC	60°C	GTP binding protein (GTP1)	9
1FFT L3/R3	U84398	For: GGGGTTATGTTGCAGAATCG Rev: CGTCTCTTGTCCACCTCAAA	59°C	fructan-fructan 1-fructosyltransférase (1-FFT)	9
AF497999 L3/R3	AF497999	For: GGGCACGTATCATAGCCACT Rev: TTGAAGCGTGGACACTGAAC	53°C	germacrene A synthase long form	9
3474 L1/R1	home	For: TTCCCAAAATAACCCCAA Rev: TGCCTTCGCCACTGTTCT	55°C	hemoglobine 1	9

<sup>a</sup>: PPX1 L2R2 and PPX1 L3R3 are the same locus in K28K59 and Rubis 118 maps, respectively

<sup>b</sup>: EST sequences issued from chicory cDNA libraries developed by Legrand *et al.* (2007)

<sup>c</sup>: RAPD fragments converted in STS markers. Primer pairs were not designed to flank introns

<sup>d</sup>: Primer pairs were not designed to flank introns

## Annexes

**Table 2:** Synthesis of the relationships between the three maps through bridge marker information.

- Distribution of bridge markers between the three maps. The proportion of bridge markers vs mapped markers for each LG is indicated.
- Distribution of inter-map comparisons. Relationships between three maps through bridge markers imply four types of inter-map comparisons : KK/Ru, KK/BR, Ru/BR, KK/Ru/Br, where KK, Ru and BR mean K28K59, Rubis118 and BR, respectively.

a)

	K28K59		Rubis118		BR	
	Number	bridge vs mapped (%)	Number	bridge vs mapped (%)	Number	bridge vs mapped (%)
LG1	23	62.1	22	52.3	11	61.1
LG2	20	58.8	16	64	7	63.6
LG3	16	51.6	23	60.5	13	56.5
LG4	13	65	11	50	9	47.3
LG5	13	50	15	50	9	56.2
LG6	20	51.2	28	65.1	17	50
LG7	14	66.6	15	78.9	12	57.1
LG8	27	56.2	21	60	10	50
LG9	8	42.1	11	37.9	8	42.1
Total	154	56.2	161	57.1	96	53.9

b)

	inter-map comparisons				Sum	%
	KK/Ru	KK/BR	Ru/BR	KK/Ru/BR		
LG1	15	4	3	4	26	13.5
LG2	14	5	1	1	21	10.9
LG3	13	2	8	3	26	13.5
LG4	6	4	2	3	15	7.8
LG5	6	1	3	5	15	7.8
LG6	13	2	10	5	30	15.5
LG7	7	4	5	3	19	9.8
LG8	18	7	1	2	28	14.5
LG9	5	2	5	1	13	6.7
Sum	97	31	38	27	193	
%	50.2	16.1	19.7	14		

## Annexes

**Table 3** : Genome coverage ( $G_{ef}$ ,  $G_o$ ,  $G_o/G_{ef}$ ,  $P_{LB}$ ) and marker distribution (cpf) for 4 genetic maps (K28K59, Rubis 118, BR, Reference map). Parameters are calculated for the complete set of marker data or after selecting j framework markers (K28K59, Rubis118, BR maps). The cpf (ILGC) values are only given for significant LG tests .

Parameters	K28K59		Rubis 118		BR		Reference
	complete	framework (j = 88)	complete	framework (j = 93)	complete	framework (j = 84)	complete
$G_{ef}$ (cM)	800.7	932.4	859.2	1001.3	986.4	1126	915.8
$G_o$ (cM)	749.8	759.4	806.1	824.6	891.4	908.1	881.7
$G_o/G_{ef}$	0.936	0.814	0.938	0.823	0.90	0.806	0.962
$P_{LB}$ (d= 5 cM)	0.966	0.61	0.962	0.605	0.834	0.526	0.994
$P_{LB}$ (d= 10 cM)	0.998	0.848	0.998	0.844	0.972	0.775	0.999
cpf	0.982 (LG2), 0.993 (LG8) 0.0007 (LG9)		0.998 (LG3) 0.008 (LG7)		0.976 (LG6) 0.025 (LG8)		0.997 (LG6)

**Table 4** : Level of homozygosity between the 6 parents of mapped populations.

- a) the proportion of homozygous markers was based on a selection of 130 bridge markers
- b) two-by two comparison of homozygous markers proportions based on the Z-test for comparison of two proportions issued from independent samples. Z-values (above the diagonal) and associated levels of significance (below the diagonal) were displayed in the 6 x 6 table. Significance to  $\alpha= 0.05$  or  $\alpha =0.001$  were noted \* and \*\*\*, respectively. Non significant-Z-tests were noted 'ns'.

a)	MS8	Rubis118	K28	K59	B	R
Homozygosity (%)	0.532786885	0.11023622	0.415384615	0.361538462	0.951612903	0.976190476

b)	MS8	Rubis118	K28	K59	B	R
MS8		4.562486814	1.404745603	2.195210023	5.044088009	4.113182909
Rubis118	***		3.324063537	2.51659506	8.935664739	7.279142497
K28	ns	***		0.843232403	6.282358831	4.986320084
K59	*	*	ns		6.97685621	5.527707984
B	***	***	***	***		1.291543691
R	***	***	***	***	ns	

**Annexe 2** : Milieux de culture utilisés pour la multiplication clonale des génotypes de la descendance F1', les tests d'embryogenèse somatique et d'organogenèse

Composition des milieux de culture	M17S20	H5	H5solide
solution de macroéléments concentrée 10x	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 <sub>2</sub> P <sub>04</sub>	85 mg	-	-
Solution de Fe-EDTA	5 ml	5 ml	5 ml
Saccharose	20 mg	5 g	5 g
Agar	-	-	6 g
H <sub>2</sub> O	qsp 1 l	qsp 1 l	qsp 1 l

Solution stock macro éléments x10 (mg/l)	M17S20	Heller
MgSO <sub>4</sub> , 7H <sub>2</sub> O	1850	2500
CaCl <sub>2</sub> , 2H <sub>2</sub> O	2200	750
NH <sub>4</sub> N <sub>03</sub>	8250	
KCl	7500	7500
NaN <sub>03</sub>		6000
NaH <sub>2</sub> P <sub>04</sub> , 2H <sub>2</sub> O		1413

Vitamines de Morel et Wetmore (1951) (mg/ml)	
Panthoténate de Ca	1
Thiamine	1
Acide nicotinique	1
Pyridoxine	1
Biotine	0.5

Milieu de Bourgeonnement Bg 0.5	
CaCl <sub>2</sub> , 2H <sub>2</sub> O	4,4 mg
MgSO <sub>4</sub> , 7H <sub>2</sub> O	3,7 mg
NH <sub>4</sub> N <sub>03</sub>	16,5 mg
KN <sub>03</sub>	19 mg
KH <sub>2</sub> P <sub>04</sub>	1,7 mg
Solution de Fe-EDTA	5 ml
Microéléments (MS)	1 ml
Glutamine	100 mg
Vitamines de Morel et Wetmore	10 ml
inositol	100 mg
BAP	0,5 mg
Saccharose	20 g
pH	5,7

Microéléments Heller (1953) (mg/ml)	
ZnSO <sub>4</sub> ,7H <sub>2</sub> O	1
H <sub>3</sub> B <sub>04</sub> 1	1
MnSO <sub>4</sub>	0,075
CuSO <sub>4</sub> , 5H <sub>2</sub> O	0,03
AlCl <sub>3</sub>	0,03
NiCl <sub>2</sub> , H <sub>2</sub> O	0,01
KCl	0,01

Microéléments de Murashigue et Skoog (MS) (1962) (mg/ml)	
H <sub>3</sub> B <sub>03</sub>	620
CoCl <sub>2</sub>	2,5
CUS <sub>04</sub>	2,5
MnSO <sub>4</sub>	1690
KI	83
ZnSO <sub>4</sub> , 7H <sub>2</sub> O	860
Na <sub>2</sub> Mo <sub>04</sub> , 2H <sub>2</sub> O	25

Solution de Fe-EDTA (g/l)	
Na <sub>2</sub> EDTA	7,46
FeSO <sub>4</sub> , 7H <sub>2</sub> O	5,56

## Annexes

### Annexe 3: EST pour lesquels des SNP (*Single Nucleotide Polymorphisme*) a pu être détectés.

Les OC indiqués en rouge ont été cartographiés sur la carte F1'a. Les couples d'amorces précédés d'un symbole \* ont permis de détecter du polymorphisme.

OC	Gène d' <i>Arabidopsis thaliana</i>	Nature/Fonction de la protéine	Amorces définies	Taille de la séquence à amplifier (pb)	Taille de l'intron chez <i>Arabidopsis thaliana</i> (pb)	Taille du produit PCR sur gel d'agarose (pb)
<b>OC0082_a</b>	AT2G27710	60s acidic ribosomal protein [Hyacinthus orientalis]	L mTGCAGATGAAGACAGAATCCA R CCTCTTTCTCTTCAACCTTCTCC	210	0	200
OC0170	AT5G02380	metallothionein 1 [Aster tripolium]	L1 CGGTGTTGCACCAAAGAAG R hACCTGGATGTGGGATGGTC	250	0	500
			L2 TGTCTTGCTGCAATGGAAAG R hACCTGGATGTGGGATGGTC	561	176	400 et 600
<b>OC0183</b>	AT5G53400	salt tolerance protein 5 [Beta vulgaris]	L hAAGGAGAAGCTCGATGCTGA R CGTTCGACTCCTCCATCTTC	184	0	200
			*L2 mGGAGTCGAACGATGATGACA R2 GGGATGTTTATGGTGACCTCTT	599	485	600
<b>OC0206</b>	AT2G17370	hydroxy-methylglutaryl-coenzyme A reductase [Nicotiana tabacum]	L mATACAAGCGGTGGTTTCTGG R CGGAGAACTCCTTTCCATCA	250	0	250
OC0316	AT1G77120	alcohol dehydrogenase (EC 1.1.1.1) [Lactuca sativa]	L hCATGTGCGACCTACTGAGGA R ACAGCTCCCAATCCGAAAAT	285	0	390
			L2 CTCCGTTTCAGCAATCAGGT R2 hTTCGGACTTCCATTTTCTGG	287	131	350
<b>OC0378_a</b>	AT3G25520	60S ribosomal protein L5 [Solanum melongena]	L mGGCTCGTCGAGTTCTGAAAA R AACACGATTGCCAGTTGTTG	160	85	200
OC0466_b	AT5G52640	putative heat shock protein 81-2 (HSP81-2) [Arabidopsis thaliana]	L hGCTTTTCATCAGGCTCGTTC R AGAAACCAACACCGAACTGG	191	85	600
OC0564	AT1G60730	putative auxin-induced protein [Arabidopsis]	L GCGATGTGGTTCCAATACCT R hTTTCCACGATGACAATGGAG	200	0	190
			L2 hTGGTCAAGAGATTTGGAGGA R2 TCCCTCGGTCAAACCTTCTG	209	76	300
<b>OC0603_a</b>	AT3G03780	methionine synthase [Solanum tuberosum]	L mCCAATTCCAGCACCGTATTT R AAAGCGGAACATGCCTTCTA	233	0	200
<b>OC0687</b>	AT2G33790	arabinogalactan protein [Daucus carota]	L mTGACCACGTATGGGTCACAC R TGGACCCACACTGAACACAT	182	0	200

## Annexes

			*L2 GACAAGAACGGGTACTTCTTCA R2 mCATCATTGTCAAATCAGTCCTCA	292	0	300
<b>OC0748</b>	AT3G13580	ribosomal protein L7 [Solanum tuberosum]	*L CCATGGACCCAAAGTCAAAG R mGGCTCAACCCTGTGTAGCAT L2 TCGTTTCGCACTAAATCTTGC R mTTGGGCCCATGGTTACTCTA	115	86	200 et 400
<b>OC1635_a</b>	AT2G45180	lipid transfer protein (LTP) family protein [Arabidopsis thaliana]	L hGATGGGGATTTCTTTCCACA R GAACCCCATCAAAGACTCCA	166	0	180
OC1817_b	AT3G23990	mitochondrial chaperonin HSP60 [Arabidopsis thaliana]	L1 TGCAAACAACAAGACTCAACAG R mTTGGGCCCATGGTTACTCTA	150	0	200
<b>OC1976</b>	AT3G12120	delta 12 fatty acid desaturase [Crepis palaestina]	L GCATCCATCTTTACCGCACT R hCTGAAAAACCGACGTTCCAT	235	0	200

## Annexes

### Annexe 4: EST pour lesquels des amorces ont été définies autour du ou des régions introniques.

Les OC indiqués en rouge ont été cartographiés sur la carte F1'a. Les couples d'amorces précédés d'un symbole \* ont permis de détecter du polymorphisme.

\* amplification d'un intron dans la séquence chicorée

OC	Gène d' <i>Arabidopsis thaliana</i>	Fonction putative	Couples d'amorces	Taille de la séquence à amplifier (pb)	Taille de l'intron chez <i>Arabidopsis thaliana</i> (pb)	Taille du produit PCR sur gel d'agarose (pb)
OC0003_a	AT3G09270	glutathione S-transferase GST 13 [Glycine max]	L hGAGTTCCGATTTTGCCTCAG R CTTGGGGAAAAGCCTTGAAT	253	145	800*
<b>OC0023_a</b>	AT1G02920	glutathione transferase [Hyoscyamus muticus]	L mTGGTGATCTCAAGCTGTTTCG R GCGTAAGCGTGTGCTATGT	167	106	400*
<b>OC0026a2</b>	At3g04120	glyceraldehyde 3-phosphate dehydrogenase [Solanum tuberosum]	L mTGGCGTTTGAGACGATGTTA R CCCACATTAAGGGTGGTGTC	217	94	aucune amplification
			* L3 hATGGGTGCATCTTTGCTAGG R3 mACGACAGTGTTCACGGACAA	250	95	350*
OC0030	AT3G09390	metallothionein-1 like protein [Oenanthe javanica]	L mAGTGTGGCTCAGACTGCAAA R GTGGTGGTCGACGTCTCAA	281	206	200*
<b>OC0031</b>	AT5G05340	peroxidase 3 [Scutellaria baicalensis]	L hTGGTGTGGTTTCTTGTGCAG R TGTTGTGCAATCCCTCCTTC	198	89	200*
OC0075	AT3G21240	4-coumarate--CoA ligase [Petroselinum crispum]	L GAGGATGAAGTTCCCGAATG R hGTCCTGCCTCCGTCATAACC	997	936	100
OC0134	AT5G15230	predicted protein [Magnaporthe grisea 70-15]	L mCAGCGCTTTCATACCGAACT R CTTCAATTCCTCCAGCACT	178	118	400*
OC0138	AT5G25420	unnamed protein product [Debaryomyces hansenii]	L mCAGAAGTATAGCGCAATGTTAAGAG R GTCAAATGGCCAACGAACAT	159	101	150
OC0163	AT1G58030	similar to Methanosarcina mazei Methanosarcina frisia [Dictyostelium discoideum]	R hCACCATGAACGGTGGATATG L hTTAAATTAACACACAAATTGAAGGA	282	91	aucune amplification
OC0178	At4g33865	putative ribosomal S29 protein [Arabidopsis thaliana]	L hGTCCTGGATCCCGCCTCT R CATGGGAGTTGCCACACA	492	450	650*
<b>OC0185</b>	AT1G54290	putative translation factor [Pinus pinaster]	L hCGTGTTCACAACGAAATGG R CCTGAGTAAGAAAGCTGGAGACA	280	81	400*
			L2 hCAACTTCAAGGTGACCAGAGAA R2 CTTGCATATCCATCATAACCACA	227	110	200*

## Annexes

			*L3 mTGACCTCGACCTCCAGATTC R3 AGCACCAGAGTCCTCAGCAT	218	148	350*
OC0242	AT5G60360	putative oryzain gamma chain precursor [ <i>Oryza sativa</i> ]	L hGGTGCCCATATTGGCTCAT R TTGTACGATGCACAAGTTGC	212	95	250*
<b>OC0252_a</b>	AT4G25740	hypothetical protein CPE1384 [ <i>Clostridium perfringens</i> str. 13]	L mAAGTCAGCCAAACCCCTTG R1 TGTCTCCTTCGAATCTTGAG	248	166	250*
			L mAAGTCAGCCAAACCCCTTG R2 CACCACGGTATCCATCCCTA	212	154	200*
OC0289	AT3G54300	synaptobrevin-related family protein [ <i>Arabidopsis thaliana</i> ]	L mTTGCGATCACCATACCTTCA R1 TCAATCTCGTCTGCATGCTC	481	100 + 86	900*
			L mTTGCGATCACCATACCTTCA R2 TTGCCAACTGATTCTTTTGC	178	100	600*
OC0311a	At3g11250	acidic ribosomal protein P0 [ <i>Glycine max</i> ]	L mACCTCTGGGCTGAAGACTGA R AAATCATCACTCCCGTCGAG	244	93	300*
OC0314_b	AT1G60710	auxin-induced protein [ <i>Helianthus annuus</i> ]	L hTGTTAGCGACGGAGTTTGTG R CCAGCTCCAGACTGCTACAAT	244	182	aucune amplification
			L2 mGGAGCTGTAGCCTCGTTCAC R2 AGAACGCCTCTCGTTCCTGT	244	159	aucune amplification
OC0347	At4g02390	poly(ADP-ribose) polymerase [ <i>Zea mays</i> ]	L hAGCTCAGCCATCTCTCCAAG R TGCAAATATTGCTGTGCATC	187	103	350*
OC0353	AT4G35090	catalase 3 [ <i>Helianthus annuus</i> ]	L ACGCAAAGATTCTCCCTTGA R TTTGCAACCAGTTGGACGTA	247	82	142
<b>OC0390</b>	AT2G41840	40S ribosomal protein S2 [ <i>Arabidopsis thaliana</i> ]	*L hCCTCCAGAAATCAGGAGTCAG R CCTCCAGAAATCAGGAGTCAG	600	525	600*
OC0394	AT5G47770	farnesyl diphosphate synthase [ <i>Parthenium argentatum</i> ]	L mGCAACCGGAATATAAAATGAGT R ATCTATTCATCGCCGGATTG	182	118	150*
			L2 mCAATCCGGCGATGAATAGAT R2 TCTGTTCAATGAGGTGGAATTT	192	76	180*
OC0423	AT4G24830	argininosuccinate synthase-like protein [ <i>Arabidopsis thaliana</i> ]	L mGAATTGGGCGAATCGACAT R TTGAATTGGATGGCTTCTCG	277	130	140*
OC0494	AT5G60360	pre-pro-cysteine proteinase [ <i>Lycopersicon esculentum</i> ]	L mCACAACTCACCGATGCTGT R CACAATTCCTGTTTTCTCCA	198	138	148*
<b>OC0565a</b>	AT5G48230	acetoacetyl-CoA thiolase [ <i>Arabidopsis thaliana</i> ]	L1 hGGCGCAGGAATCCCTAATAC R1 TTTATGCCCAATTGGATGCT	201	94	240*
			*L2 mTGGCATGGAAAGCATGTCTA R2 CATCCCACAAACCGTCTTTC	192	80	210*
<b>OC0581</b>	AT5G10360	S6 ribosomal protein [ <i>Asparagus officinalis</i> ]	L hTAATGGAGATGCCCTTGGAG R CTCCTTCCATACCCACGAAA	352	93 + 97	500*
			*L2 mCTGGATGCCAGAAGAAGCTC R2 CTCCAAGGGCATCTCCATTA	391	292	450*

## Annexes

OC0633	AT4G16260	ss-1,3-glucanase [Cichorium intybus x Cichorium endivia]	L mTCTGTTGGGCAGTTCCTTTC R TAGATGTTGGCAAGCATTGG	238	118	117
<b>OC0686</b>	AT2G36630	unknown protein [Arabidopsis thaliana]	L hACTTGAGATCGGCGTCATTC R GGATTGGGAACCGATTGAGT	204	90	178*
OC0688	AT4G24830	putative argininosuccinate synthase [Arabidopsis thaliana]	L hTGTGGCAAATGAAGTTGGAG R GCCACAACCTGAAGATCAGGA	206	88	189*
OC0696	AT3G16240	aquaporin [Helianthus annuus]	L hTTGATACGGCAACAAACAGG R CAGCAGATGCTGCCCTAGAT	488	439	aucune amplification
OC0749	AT1G12950	unknown protein [Arabidopsis thaliana]	L hTCAGAGGGTCATGTGGTGAA R ACCAATGCCATGAAATACCA	233	100	500*
OC0755	AT1G05010	1-aminocyclopropane-1-carboxylate oxidase [Pelargonium x hortorum]	L mCCAGTTGTGAGCATGGAGAA R TCGCCACCATTTCTTAAAC	271	72	400*
OC0812	AT3G19130	RNA Binding Protein 47 [Nicotiana glauca]	L GGATTCGGTGCTATTCCTCA R hCTGCTTGCGAATTGAACAAA	271	83	300*
OC0813	AT4G24830	argininosuccinate synthase-like protein [Arabidopsis thaliana]	L mGGGTGTGAGGGTTGTCTGCTT R CACTTGCCTTAGCCTTTGCT	192	106	525*
<b>OC0818</b>	AT5G04140	glutamate synthase (ferredoxin) [Nicotiana tabacum]	L mGATGATACCCTCATTCCTCAAG R AGTAGCGCCTTTGGTGCTC	300	75 + 91	aucune amplification
			*L2 mAGAAACGTGGCAGCTGGTAT R2 TGAGGCTCTTGAGCTGCAT	217	75	250*
<b>OC0839</b>	AT1G28330	dormancy-associated protein [Codonopsis lanceolata]	L mTCGATCAGATGTGGGATGAA R GCGGCATCGATAGAGACTTC	235	91	600*
OC0840	AT5G47770	farnesyl diphosphate synthase [Parthenium argentatum]	L hATCTTTTCGACCCATTGACG R GCACCGATCTGAAGTCCAAA	219	106	190*
			L2 hCAAGAAATTTGGACTTCAGATCG R2 TGATACTTCGCCATTTTATCG	186	118	400*
OC1011	AT3G63150	GTP-binding protein-related [Arabidopsis thaliana]	L mACCGGGGCTTCTACAAGATT R AGAACCCAGGAAGTCGGAAG	350	213	1200*
OC1044	AT4G02120	CTP synthase-like protein [Arabidopsis thaliana]	L hTGCCCTTCATAGAGGCTTTG R GGAACCTGAGAGAGGTTTCTGCT	463	97+65+69	600*
			L2 hTTCTTGATGATGGGGGAGAG R2 TTCCCGGTGGTTATGTTGTT	178	80	200*
OC1051	AT4G30210	NADPH-ferrihemoprotein reductase [Helianthus tuberosus]	L mGGAAGTTTAGGGTCGGAAGG R CCGATCTTTATCAGAACGTCAA	180	125	150*
<b>OC1104</b>	AT5G57660	zinc finger protein [Pinus radiata]	L hTCTCAATCCGTCAGCCAAAG R AATTGGTGTTGGTGTTGTT	205	89	210*
OC1111	AT3G63150	P protein [Flaveria pringlei]	L mATGCTGGGCTCATGTTATC R TGATCTCCTCTTGCCATGTG	249	200	aucune amplification
			L2 mTTCGTTCTCTTTGCAACCAA R2 mTGATCTCCTCTTGCCATGTG	291	200	aucune amplification

## Annexes

OC1121	AT5G16440	isopentenyl pyrophosphate:dimethylallyl pyrophosphate isomerase [Lactuca sativa]	L hTGCATTTTGGTGGATGAGAA R AAACACACTGAATGCTCTGTGAA	265	151	400*
			L2 hAACGGTCAGAAACGAAGGTG R2 CCTTTGTGCAGCATTCTGA	215	90	150
			L3 hTCACAGAGCATTCAAGTGTGTTTT R3 ACCTTCGTTTCTGACCGTTG	382	307	190
OC1129	AT4G05460	F-box protein family, AtFBL20 [Arabidopsis thaliana]	L hAGTTCTTAGCATCGGGCAAC R TTTCCATTATTGGAGGAACTGA	487	386	190
<b>OC1138</b>	AT1G76490	3-hydroxy-3-methylglutaryl-coenzyme A reductase [Hevea brasiliensis]	L mTCACGATGATGGAAGCTGTT R CTGTGATGCCAATTGTGTCC	198	91	200*
OC1143	AT3G27060	ribonucleotide reductase R2 [Nicotiana tabacum]	L mTTGAGACGATTCCGTGTGTC R CAACCCACGTTTCTTCAACC	249	88	aucune amplification
OC1149	AT1G22710	sucrose transporter [Glycine max]	L hAATCTTGCGATCGTCATTCC R ATCAGGAGGTGAAGGAAGCA	248	89	aucune amplification
			L2 hGCCACCAAACCTCCGGTATAA R2 CGCAAGGGACACTGAAAGTT	234	145	aucune amplification
OC1165	AT5G26751	Petunia Shaggy kinase 4 [Petunia x hybrida]	L hCACCGTGACATTAAGCCTCA R CCAGAACTTTTGCACTTCCA	184	96	350*
			L2 hGGTCAGTCGGTTGTGTCCCTT R2 TTGGTCAACTCCACTTTCACC	161	87	180*
OC1225_b	AT3G01280	34 kDA porin [Solanum tuberosum]	L mAGCTTTTCCAATCCGACCT R GTCCCAACCGTGATTGTGTT	250	86	250*
OC1307	AT2G31740	putative spermine/spermidine synthase [Glycine max]	L mTGGATCTAATTCAACTGCCTCA R ATTGGTCTTGGTGTGGACT	158	74	190*
			L2 mGACCAATGACAACCTGCCCTAA R2 GCAAGCAATAATCCAATGTGA	249	107	300*
OC1308	AT3G13580	60S ribosomal protein L7 [Arabidopsis thaliana]	L hGTGCCGCTTCCAAGATTT R ACGGACTCTGGAACAAGCTG	700	631	600 et 1500*
OC1393	AT3G02780	isopentenyl pyrophosphate:dimethylallyl pyrophosphate isomerase [Lactuca sativa]	L hGATGGAAAGTGGGGAGAACA R TCATCAGGGTTTGGGTTTAC	172	92	1200*
<b>OC1394</b>	AT5G53400	salt tolerance protein 5[Beta vulgaris]	L mCTGAAACCCGATCCACAGTT R CTCATCACTTGTGGGCAATC	169	92	190*
<b>OC1412</b>	AT5G65650	unknown [Arabidopsis thaliana]	L hCATGCTTTTCTGCTTCTC R GAAATCGCCAAGATGTCGTT	169	85	200*
OC1419	AT2G41420	No hits found	L mTTCAAATTTCTTCAATTCATGGT R TCTGTGTTGTTGCTGCCTTC	241	145	180*
OC1447	AT4G38220	aminoacylase, putative / N-acyl-L-amino-acid amidohydrolase [Arabidopsis thaliana]	L mTTGTTCCCGATGAGGAAATC R AACCACATCGGACACCTCTC	231	71	180
OC1461	AT1G76490	3-hydroxy-3-methylglutaryl coenzyme A reductase [Solanum tuberosum]	L mAAGCTGACTTCCCCGACAT R CACGGCTGCTGGTTTCTTAT	235	161	190*

## Annexes

			L2 hCCCATGCAAGTAACATCGTG R2 AGCCAATTGAGTTCCACCAC	273	91	400*
<b>OC1517</b>	AT1G60420	unknown protein PDI-like protein [Arabidopsis thaliana]	L hACAAACCCGAAATACACATCA R GAGAAGGGGAAGATGGAAGG	215	90	190*
<b>OC1528</b>	AT5G18170	NADH-glutamate dehydrogenase [Nicotiana glauca]	L hTTCTGGTCAACGGTTTGTCA R TCCGTTCTTGCCCTTATCG	206	76	250*
<b>OC1531_a</b>	AT4G25740	putative ribosomal protein S10 [Arabidopsis thaliana]	L mTCCATCTGACATCGTTCTCTG R TCGGTCTATCTCCTTCAAATCT	285	166	250*
OC1539	AT4G11290	bacterial-induced class III peroxidase [Gossypium hirsutum]	L hTAATTAAGCCAGCCGGGACT R GCAACCTTAACCACCTTAGCA	383	244	200*
OC1546	AT3G02260	auxin transport protein [Arabidopsis thaliana]	L mCGGGATGAAATGAAACGACT R GTTCTTGCCTCATTCCAAGC	161	88	180*
OC1554_a	AT1G44010	No hits found	L hTCCAACCAAAAACCTCACAA R CGAACAGTGGCAACAACAAC	203	87	1500*
<b>OC1580</b>	AT4G31700	40S ribosomal protein S6 [Cicer arietinum]	L CGCAGAACCTTCACAAACAA R hAATTCTTGCCCTTCTCTTTG	196	94	200*
OC1582	AT1G75750	GAST1-like protein [Arabidopsis thaliana]	L mAAGGGCGCTTCTTTTGT R GGGACATGCTGTGGAAGGT	287	93+82	150
OC1590	AT5G60660	plasma intrinsic protein 2,2 [Juglans regia]	L mCCTAGGCATTGCTCCTGAAG R GCCTTCTACCACCAGTTCTG	272	199	100
			L2 mTGCACCAAACTACGAGCAG R2 CTTCCAATCGGATTTGCTGT	187	97	200*
OC1608	AT1G54290	translation initiation factor B04 [Helianthus annuus]	L1 GAAAGAATTCTGCTGCAATGG R mGAGTCAAGAACGCAGAGACG	185	81	100
			L2 mTCCAGATCCCATCTGCTTTC R2 CTTGGTCAACGTCGAAAAAT	432	148	500*
OC1623	ATG19130	DNA/RNA binding protein-like [Arabidopsis thaliana]	L mTTACCTTCAGTCGTGCTTCG R TTGACCAGTTTGCTTGTTC	484	410	400*
OC1643	AT1G24020	vegetative storage protein, VSP [Cichorium intybus]	L hTTGGAACCATCAAGAGCATT R TGCTTTGAGCTTGTGAATGG	396	338	200*
OC1649	AT3G10370	putative glycerol-3-phosphate dehydrogenase [Arabidopsis thaliana]	L hGGAACGGATCATAGGAGCAA R AAAGGTCCACCAGCATTAC	182	89	400*
OC1658	AT2G05510	No hits found	L hGGCGTGGAGGATACAACAAT R GTATCCACGTCCACCACCAT	215	69	aucune amplification
<b>OC1696</b>	AT2G26080	P protein [Flaveria pringlei]	L hTGGAGCGAATATGAATGCAC R1 TGGAATGCAAAATGTCTTGTG	172	75	250*
			*L hTGGAGCGAATATGAATGCAC R2 CTGCTGAAATGGTGCCAAGT	187	97	250*
OC1741	AT4G34110	poly(A)-binding protein [Nicotiana tabacum]	L mGCTCTGGTTTTGTTCCTTT R TGCAACCATCTTTCCATTCA	152	75	180

## Annexes

			L2 hTGGCACTTGCACAAAGAAAG R2 ATCCGAGGTGCAACAGTAGG	367	267	250*
OC1752	AT4G17260	lactate dehydrogenase-2 [Solanum tuberosum]	L mTCTCAATGCTCAGGATGTGC R TGTTCTCCGACGATGTATGC	233	191	200*
OC1785a	At2g24520	H(+)-transporting ATPase [Phaseolus vulgaris]	L hGACGTGCTGAAGTTGCAAGA R TCAAGCCCCTTAAGTTAAACCA	281	199	180*
<b>OC1821</b>	AT3G48930	40S ribosomal protein S11 [Euphorbia esula]	L hCTTCAGAGCTCGCAATCCAT R GTTGCTTCAAAAACGCCTTC	176	111	180*
<b>OC1822_b</b>	AT1G08360	putative ribosomal protein L10a [Oryza sativa]	L mTCAAGCAAATACCCCGTCTC R TTCACTGTTGCCTTTGTTCA	293	182	200*
<b>OC1833_b</b>	AT1G61580	ribosomal protein L3A [Nicotiana tabacum]	L hAAGTTCGAGCACCCCAGAC R CCAGCCTTGTAACCCAAGAA *L2 hCAAGGATGACCCAACCAAAAC	282	139	300*
			R2 GGAGGTGTTTCGATGATGGT	233	83	400*
OC1839_b	AT1G10750	unnamed protein product [Debaryomyces hansenii]	L mTTTAGGTAACGCTCGGCTTG R TTGTAACGACCCGGCTCTC	191	102	190*
			R2 mTTTGTAAACGACCCGGCTCTC			aucune amplification
			L2 mTTTAGGTAACGCTCGGCTTG	191	102	
<b>OC1929</b>	AT1G48630	G protein beta subunit [Nicotiana glauca]	L mAAATGGCAGCTGAAGACACC R TCCATCTGCACTCCAGCTTA	400	311	350*
<b>OC1952</b>	AT4G30210	NADPH-ferrihemoprotein reductase [Helianthus tuberosus]	L hAAAGGAATCTGTCAACCTGGA R GGGTTTGTAGGAAGTCTGAAGTTT	247	134	200*
<b>OC2006</b>	At1g48410	putative leaf development protein Argonaute [Arabidopsis thaliana]	L hTCAACCGGTCATAAACCACA R TCCATTTTCATTCAGCAACACA	174	85	180*
OC2012_b	AT3G52580	putative ribosomal protein S14 [Oryza sativa (japonica cultivar-group)]	L mTTGCTGCACAAGATGTTTCC R TGGTCTTGTTCCTCCAGTT	186	96	200*
			L2 mACGTCACTCTTGACCTGCT R2 ACCATGGTTTCTCTCCAGA	198	77	400*
OC2022	AT4G34670	ribosomal protein S3a [Catharanthus roseus]	L hCAATCAAGCCAGATTGTC R AAACTTGGGTGCCTTCAAGA	293	92	1000*
<b>OC2060</b>	AT2G38540	lipid transfer protein [Helianthus annuus]	L hAAATGCGGTGTCTCAATTCC R TCGTCTCCCTTATCCGTTG	226	119	500*
<b>OC2141</b>	AT1G12950	unknown protein	L mGCGGTTTATGCTCTCACTCC R AATAACCGCTTGCCATCCT	203	98	400*
OC2155a	At1g72370	P40-like ribosomal protein [Daucus carota]	L mAACAGACACCCAATGCTGTG R CCATCAAGGAAGCTGCACTA	231	104	300*

## Annexes

**Annexe 5:** EST pour lesquels des alignements des séquences de chicorée avec les séquences laitue *sp.* ont été réalisés.

\* amplification d'un intron dans la séquence chicorée

OC	Gène d' <i>Arabidopsis thaliana</i>	Nature/Fonction de la protéine	Amorces définies	Taille de la séquence à amplifier (pb)	Taille de l'intron chez <i>Arabidopsis thaliana</i> (pb)	Taille du produit PCR sur gel d'agarose (pb)
OC0168	AT1G62380	leaf senescence protein-related [ <i>Arabidopsis thaliana</i> ]	L mCATCAACTTGGGTGATCAACTT R CACACTCTTGACCTTCCATTTG	248	193	600*
OC0745	AT4G16260	ss-1,3-glucanase [ <i>Cichorium intybus</i> x <i>Cichorium endivia</i> ]	L CCATGGCAACAATGCTCATA R AAAGGCTCACCACGTCTTGT	184	108	268*
<b>OC1359_a</b>	AT1G04480	60S ribosomal protein L17 [ <i>Oryza sativa</i> ]	L hATCGTCAGACAACGCAAGC R CTTGGCCATAAATCAGCACA	698	604	1000*
OC1568	AT4G35300	putative sugar transporter protein [ <i>Arabidopsis thaliana</i> ]	L hGTATTGCTGTGGCAATGAGG R CACTTCCCACCACCAAGATT	202	94	500*
OC1867_b	AT5G56010	heat shock protein 90 [ <i>Nicotiana tabacum</i> ]	L hATCGCGGAAAACAAAGAAGA R mGTCAGATTGGCCTTCCTTCA	192	99	300*

## Annexes

**Annexe 6:** EST pour lesquels la prédiction de jonctions « intron-exon » a été réalisée.

\* amplification d'un intron dans la séquence chicorée

OC	Gène d' <i>Arabidopsis thaliana</i>	Nature/Fonction de la protéine	Amorces définies	Taille de la séquence EST (pb)	Taille du produit PCR sur gel d'agarose (pb)
OC0067	AT1G76490	3-hydroxy-3-methylglutaryl coenzyme A reductase [ <i>Malus x domestica</i> ]	L mAAGGTGGTGGATGCAGTCTC R CGACGAACAAGACGATAGCA	178	200
OC1056	AT3G12120	delta-12 oleate desaturase [ <i>Helianthus annuus</i> ]	L hACTGATCGTGAACGGGTTTC R AATGAGGCATCGTCGAGAAC	203	300*
OC1824	AT2G45180	hybrid proline-rich protein (extensin) [ <i>Catharanthus roseus</i> ]	L hCCGAATAAGCCTCTGCAATC R CACCGAAAACCTCCAACACCT	211	aucune amplification
<b>OC1932</b>	AT4G12520	hybrid proline-rich protein [ <i>Catharanthus roseus</i> ]	L hTCGATGAGGCTACAACATGG R GCACTAGCAAGTGTTGCAC	179	200

## Annexes

### Annexe 7: EST pour lesquels des amorces ont été définies au niveau des exons

\* amplification d'un intron dans la séquence chicorée

OC	Gène d' <i>Arabidopsis</i> <i>thaliana</i>	Nature/Fonction de la protéine	Amorces définies	Taille de la séquence à amplifier (pb)	Taille du produit PCR sur gel d'agarose (pb)
OC0044	AT5G24260	putative dipeptidyl peptidase IV [ <i>Oryza sativa</i> ]	L hTACACCGATTGCTTCACACC R GATTCTGCGGGACAAATTA	247	250
<b>OC0081</b>	AT4G02510	putative chloroplast outer membrane protein [ <i>Oryza sativa</i> ]	L mGGAATGTCCGGTCACGTTTCT R GAGAAGGAATAACCGCCACA	389	400
OC0098	AT4G05400		L hGGAACGGCTAAAGGAAAAGG R AACCTTCCGGTCCGATAAAC	327	350
OC0203	AT4G34110	poly(A)-binding protein [ <i>Arabidopsis thaliana</i> ]	L mTTGCGCAAACAAGAAAGAGA R GACCATTAAGCCCTTCCACA	235	250
OC0225	AT4G33010	P protein [ <i>Flaveria pringlei</i> ]	L hGAATCCCTGCTCAATTTCCA R AAATCGCTAGCCATGACGAC	374	450*
OC0283	AT3G12120	delta12-fatty acid acetylenase [ <i>Dimorphotheca sinuata</i> ]	L mTTCGCTTGACAACGATGAAG R CCTAACACGGGAACTCCGTA	336	400*
<b>OC0344</b>	At3g04700	unknown [ <i>Arabidopsis thaliana</i> ]	L hGTGAGGGCCATCTAACGAAA R mACGGTCTCGAAATCATCGTC	108	200*
OC0453	AT1G03220	unknown protein [ <i>Arabidopsis thaliana</i> ]	L hCTCTTGGGCAGGCAAACACTAC R GAAAATTGTTGCCTTTCGTCA	237	350*
OC0502	AT5G03140	Integration host factor alpha-subunit [ <i>Rickettsia rickettsii</i> ]	L mAATCTCCCCAAGTCCAGGTT R CCGCAAGTATGGGAATGTCT	152	300*
OC0970	AT1G03220	putative xyloglucanase inhibitor [ <i>Solanum tuberosum</i> ]	L hGTGCCTTCGATTGACCTTGT R ACATGGCAAGGCCATTTATT	325	Aucune amplification
OC1013	AT1G03220	hypothetical protein [ <i>Oryza sativa</i> (japonica cultivar-group)]	L mAACCTTCTGCATCGATGGTC R GAAAATGGGTTTTTGGGTGA	119	200*
OC1068	At2g44840	leaf senescence protein-related (YLS7) [ <i>Arabidopsis thaliana</i> ]	L hAATCGCGGTTGTTTTTATG R mACCCTCGAAATCCCTCCTTA	215	300*
<b>OC1145</b>	AT3G12120	delta-12 fatty acid desaturase [ <i>Borago officinalis</i> ]	L hTGCATGGTAATGAGGTCTGG R ATGGTTTTGCCAACCATTTTC	377	400
OC1191	AT2G38870	protease inhibitor 2 [ <i>Zinnia elegans</i> ]	R CATGTCTTCCCTCAAAACATCA L mAGAGTTCATGGTCGGAGCTG	313	400*
OC1445	AT3G12120	delta 12 fatty acid desaturase [ <i>Crepis palaestina</i> ]	L hGGGTTTTTCGCAAAGCAATAA R ATCCTCCATTCCGCCTTACT	364	450*

## Annexes

OC1648	AT2G30490	Cinnamic acid 4-hydroxylase [ <i>Zinnia elegans</i> ]	L mCTAGTCCCGCACATGAACCT R CTA AAATCGTTCCCGTTTGC	188	200
OC1744	AT2G38540	lipid transfer protein [ <i>Helianthus annuus</i> ]	L hTGTGTAGTGCTGGCTTGCAT R GCACTCTTCAAACAGCCACA	209	Aucune amplification
<b>OC1796b</b>	AT1G74780	vegetative storage protein, VSP [ <i>Cichorium intybus</i> ]	L hTATGCCATTGCTCATCCTGA R CAACAGCGATACAATCCGACT	186	200
OC1968	AT4G15780	Vesicle-associated membrane protein 724 [ <i>Arabidopsis thaliana</i> ]	L mTCGGTCACTGGGTTTTCTCT R TGAAAGATTCTGACCCATGA	100	150
OC2008	AT1G05010	1-aminocyclopropane-1-carboxylate oxidase [ <i>Actinidia chinensis</i> ]	L mGCATTTAAGGCCATGGAAGA R GCAAGACACCGAAAACACAA	186	300*
OC2013_c	AT5G39580	Cationic peroxidase 2 precursor [ <i>Arachis hypogaea</i> ]	L hAATGCGCTACAAACCCTACG R AACCTGCAACTCGAACATT	158	250*
<b>OC2147</b>	AT3G10370	putative glycerol-3-phosphate dehydrogenase [ <i>Arabidopsis thaliana</i> ]	L hCTGAGGTGGCGTATTCTGCT R GACTACCCGCTGCATACCAT	359	450*

Annexes

Annexe 8: Liste des ESTs cartographiés dans la carte génétique F1'a

Numéro accession	OC (Original Cluster)	Nature de la protéine traduite	Amorces sens (5'-3')	Amorces anti-sens (5'-3')	Groupe de liaison
DT211185	OC0023_a	Gluthatione transferase [H. muticus]	TGGTGATCTCAAGCTGTTCG	GGCGTAAGCGTGTGCTATGT	8
DT212554	OC0026a	glyceraldehyde 3-phosphate dehydrogenase [Solanum tuberosum]	ATGGGTGCATCTTTGCTAGG	ACGACAGTGTTCACGGACAA	3
CONS	OC0031	peroxidase 3 [Scutellaria baicalensis]	TGGTGTGCGTTTCTGTGCAG	TGTTGTGCAATCCCTCCTTC	8
DT210904	OC0081	putative chloroplast outer membrane protein [Oryza]	GAGAAGGAATAACCGCCACA	GGAATGTCGGTACGTTTCT	1
DT213148	OC0082_b	60s acidic ribosomal protein [Hyacinthus orientalis]	TGCAGATGAAGACAGAATCCA	CCTCTTTCTCTTCAACCTTCTCC	2
DT211083	OC0183	salt tolerance protein 5[Beta vulgaris]	GGAGTCGAACGATGATGACA	GGGATGTTTATGGTGACCTCT	9
DT211088	OC0185	putative translation factor [Pinus pinaster]	TGACCTCGACCTCCAGATTC	AGCACCAGATCCTCAGCAT	1
DT211131	OC0206	hydroxy-methylglutaryl-coenzyme A reductase [Nicotiana tabacum]	ATACAAGCGGTGGTTTCTGG	CGGAGAACTCCTTTCCATCA	3
DT212813	OC0311	Acidic ribosomal protein P0 [G. max]	ACCTCTGGGCTGAAGACTGA	AAATCATCACTCCCGTCGAG	1
DT210805	OC0344	Unknow protein	GTGAGGGCCATCTAACGAAA	ACGGTCTCGAATCATCGTC	7
DT211391	OC0347	poly(ADP-ribose) polymerase [Zea mays]	AGCTCAGCCATCTCTCCAAG	TGCAAACTATTGCTGTGCATC	3
DT212642	OC0378_a	60S ribosomal protein L5 [Solanum melongena]	GGCTCGTCGAGTTCTGAAAA	AACACGATTGCCAGTTGTTG	1
DT213495	OC0390	40S ribosomal protein S2 [Arabidopsis thaliana]	CCAAAACCTCTGGAAACTTCG	CCTCCAGAAATCAGGAGTCAG	4
DT211413	OC0565_a	acetoacetyl-CoA thiolase [Arabidopsis thaliana]	TGGCATGGAAAGCATGTCTA	CATCCCACAAACCGTCTTTTC	8
DT211818	OC0581	S6 ribosomal protein [Asparagus officinalis]	CTGGATGCCAGAAGAAGCTC	CTCCAAGGGCATCTCCATTA	1
DT212807	OC0603_a	Cobalamine-independent methionine synthase [S. scutellarioides]	CCAATTCGAGCACCCTATTT	AAAGCGGAACATGCCTCTA	9
DT211599	OC0686	unknown [Arabidopsis thaliana]	ACTTGAGATCGGCGTCATTC	GGATTGGGAACCGATTGAGT	2
DT213316	OC0687	Arabinogalactan protein [D. carota]	TGACCACGTATGGGTACAC	TGGACCCACACTGAACACAT	6
DT212101	OC0748	ribosomal protein L7 [Solanum tuberosum]	CCATGGACCCAAAGTCAAAG	GGCTCAACCCTGTGTAGCAT	8
DT212197	OC0812	RNA Binding Protein 47 [Nicotiana plumbaginifolia]	GGATTCCGGTGCTATTCCTCA	CTGCTTGCGAATTGAACAAA	8
DT212209	OC0818	glutamate synthase (ferredoxin) [Nicotiana tabacum]	AGAAACGTGGCAGCTGGTAT	TGAGGCTCTTGAGCTGCAT	8
DT212243	OC0839	dormancy-associated protein [Codonopsis lanceolata]	TCGATCAGATGTGGGATGAA	GCGGCATCGATAGAGACTTC	5
DT212682	OC1104	zinc finger protein [Pinus radiata]	TCTCAATCCGTCAGCCAAAG	AATTGGTGTGGTGGTGGTT	1
DT212732	OC1138	3-hydroxy-3-methylglutaryl-coenzyme A reductase [Hevea brasiliensis]	TCACGATGATGGAAGCTGTT	CTGTGATGCCAATTGTGTCC	3
DT212742	OC1145	delta-12 fatty acid desaturase [Borago officinalis]	ATGGTTTTGCCAACCATTTC	TGCATGGTAATGAGTCTGG	7
DT213159	OC1394	salt tolerance protein 5[Beta vulgaris]	CTGAAACCCGATCCACAGTT	CTCATCACTTGTGGGCAATC	1
DT210783	OC1359_a	60S ribosomal protein L17 [Oryza sativa (japonica cultivar-group)]	ATCGTCAGACAACGCAAGC	CTTGCCATAAATCAGCACA	2
DT213182	OC1412	unknown [Arabidopsis thaliana]	CATGCTTTTCTGCTTCTCTC	GAAATCGCCAAGATGTCGTT	4
DT213333	OC1517	putative trypanothione-dependent peroxidase [Oryza sativa]	ACAAACCCGAAATACACATCA	GAAAGGGGAAGATGGAAGG	7
DT213358	OC1528	NADH-glutamate dehydrogenase [Nicotiana plumbaginifolia]	TTCTGGTCAACGGTTTGTCA	TCCGTTCTTGCCCTTATCG	2
DT213363	OC1531_a	putative ribosomal protein S10 [Arabidopsis thaliana]	TCCATCTGACATCGTTCTCTG	TCGGTCTATCTCCTTCAAATCT	1
DT213102	OC1580	40S ribosomal protein S6 [Cicer arietinum]	CGCAGAACCTTCACAAACAA	AATCTTGCCCTTCTCTTTG	9
DT212954	OC1635_a	lipid transfer protein (LTP) family protein [Arabidopsis thaliana]	GATGGGGATTCTTTCCACA	GAACCCATCAAAGACTCCA	5
DT212327	OC1696	P protein [Flaveria pringlei]	TGGAGCGAATATGAATGCAC	TGGAATGCAAAATGTCTGTG	7
DT210860	OC1796b	vegetative storage protein, VSP [Cichorium intybus]	CAACAGCGATACAATCCGACT	TATGCCATTGCTCATCCTGA	4
DT212942	OC1821	40S ribosomal protein S11 [Euphorbia esula]	CTTCAGAGCTCGCAATCCAT	GTTGCTTCAAAAACGCCTTC	7
DT213592	OC1822_b	putative ribosomal protein L10a [Oryza sativa (japonica cultivar-group)]	TCAAGCAAATACCCCGTCTC	TTCACTGTTGCCTTTGTTTCA	8
DT213720	OC1833_b	ribosomal protein L3A [Nicotiana tabacum]	CAAGGATGACCCAAACCAAC	TGGAAGTGTTCGATGATGGT	4
DT213070	OC1929	G protein beta subunit [N. plumbaginifolia]	AAATGGCAGCTGAAGACACC	TCCATCTGCACTCCAGCTTA	7
DT212755	OC1952	NADPH-ferrihemoprotein reductase [Helianthus tuberosus]	AAAGGAATCTGTTCAACCTGGA	GGGTTTGTAGGAAGTCTGAAGTTT	3
DT211080	OC1932	hybrid proline-rich protein [Catharanthus roseus]	TCCGATGAGGCTACAACATGG	GCACTAGCAAGTGGTTGCAC	7
DT213958	OC1976	delta 12 fatty acid desaturase [Crepis palaestina]	GCATCCATCTTTACCGCACT	CTGAAAAACCGAGTTCCAT	8
DT213997	OC2006	putative leaf development protein Argonaute [Arabidopsis thaliana]	TCAACCGGTCAATAACCCACA	TCCATTTTCATTAGCAACACA	2

## Annexes

DT213069	OC2060	lipid transfer protein [ <i>Helianthus annuus</i> ]	AAATGCGGTGTCTCAATTCC	TCGTCTCCCTTTATCCGTTG	8
DT214165	OC2141	TSB2 tryptophan synthase beta-subunit [ <i>Arabidopsis thaliana</i> ]	GCGGTTTATGCTCTCACTCC	AATAACCGCTTGCCATCCT	2
DT214172	OC2147	putative glycerol-3-phosphate dehydrogenase [ <i>Arabidopsis thaliana</i> ]	CTGAGGTGGCGTATTCTGCT	GACTACCGCTGCATACCAT	3

## Annexes

**Annexe 9 :** Couples d'amorces définies autour des InDel afin de détecter du polymorphisme au sein du gène *CiSTM*. Les couples d'amorces indiqués en rouge ont révélé du polymorphisme

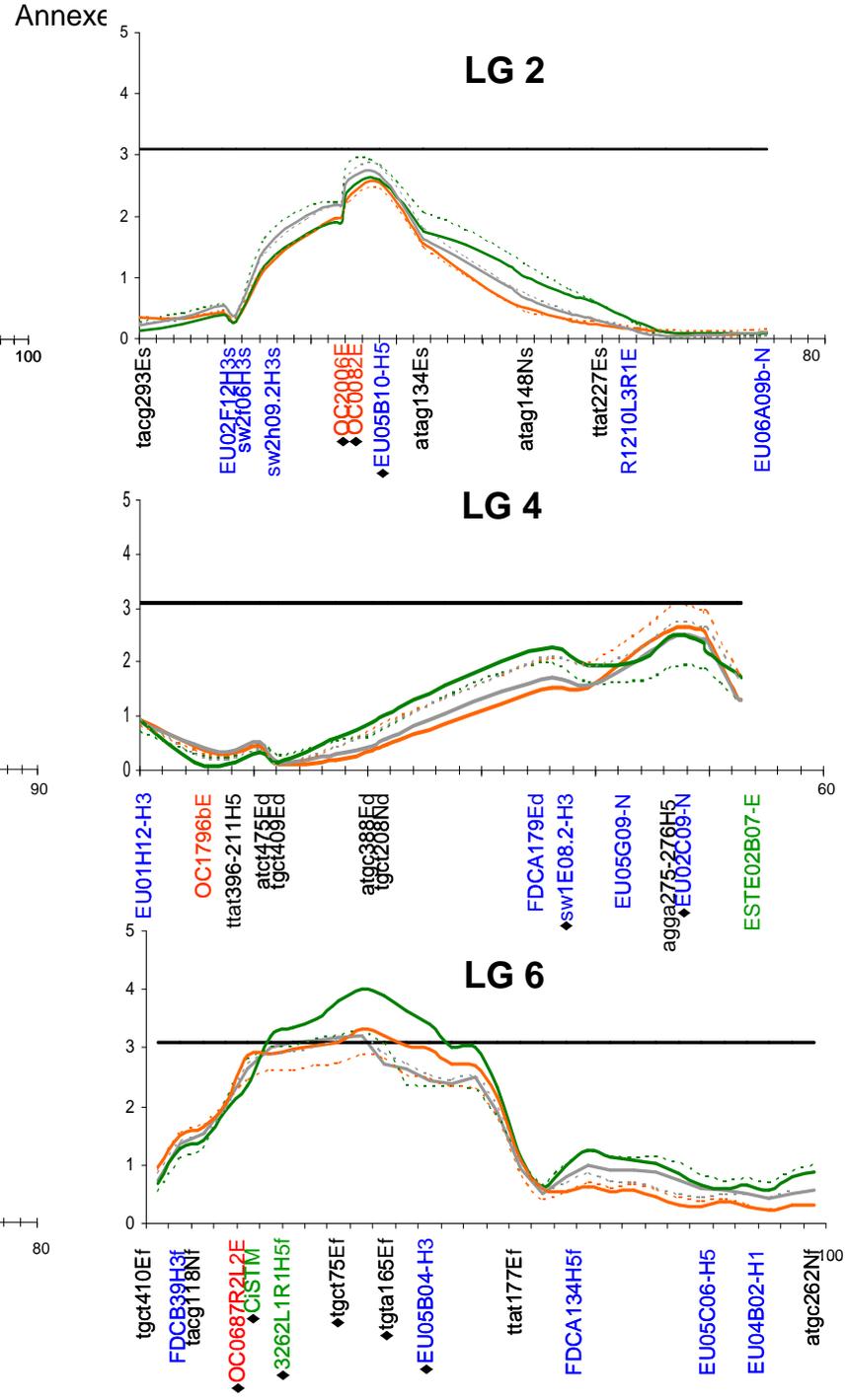
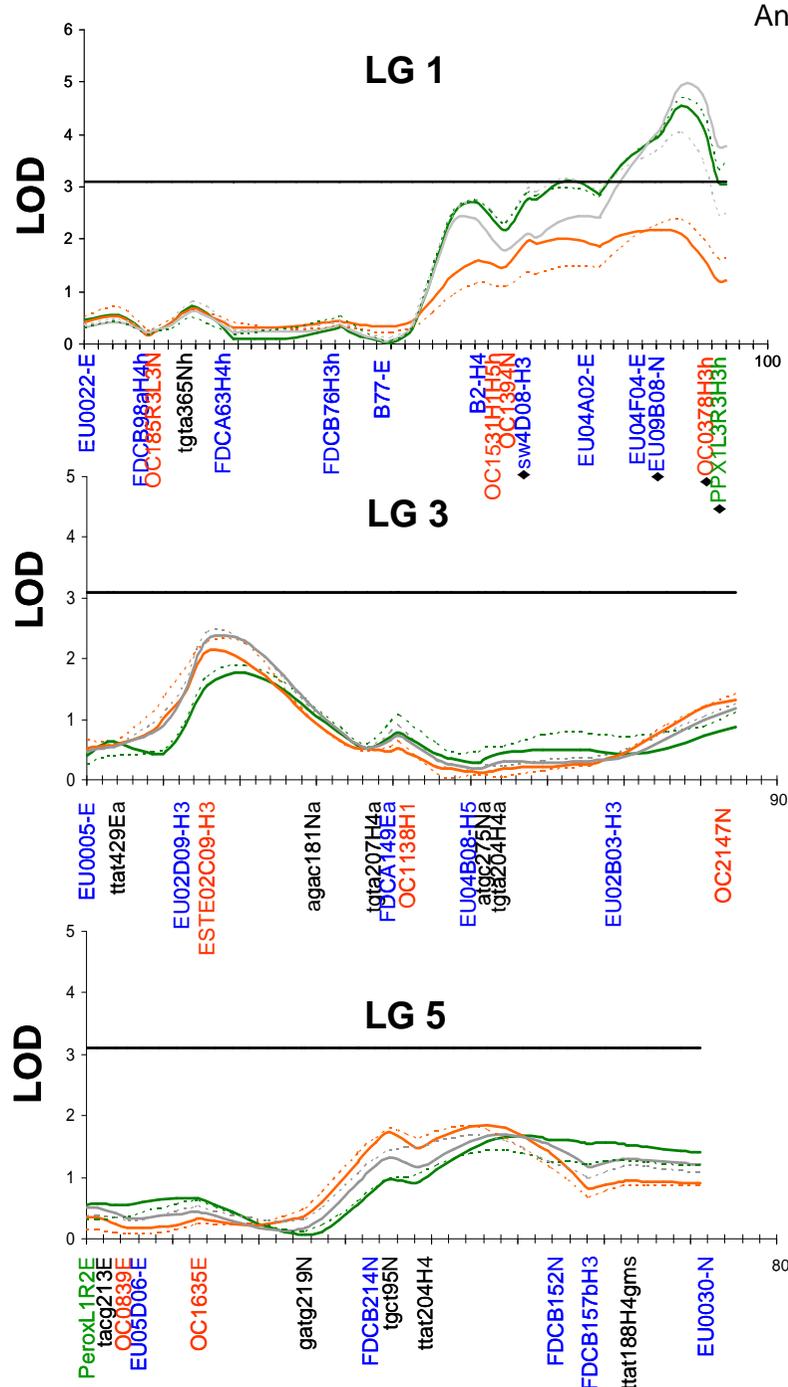
CiSTM	Amorces sens (5'-3')	Amorces anti-sens (5'-3')
CiSTM(GA)	L1 AACCAAAGGAAGCGACTG hex	R1 CAGCAGCATGCATACACTAACA
<b>CiSTM(CTAG)</b>	L1 ACCTTGCCATATCACGCATT hex	R1 TGTTGGCATAGCTGATGTGAT
<b>CiSTM(ACA)</b>	L1 TCCACTCAGGACCATCATCA fam	R1 CCCTACCTTTTGGCAATTCA

**Annexe 10:** Analyse IM (Interval Mapping). Identification des QTL liés à la formation des plantules (PL), des structures chlorophylliennes unipolaires (SH) et des événements totaux (ET) dans les sous-descendances F1'a, F1'b et la descendance F1'a+b. L'analyse a été réalisée pour chaque paramètre d'analyse ( $_{90}$ , intensité moyenne de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires et  $_{50}$ , intensité maximale de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires). La position du QTL (cM), le LOD score (LOD) et le pourcentage de variation expliquée ( $R^2$ ) sont indiqués pour chaque QTL. Les groupes de liaison non liés significativement à un QTL sont indiqués par –

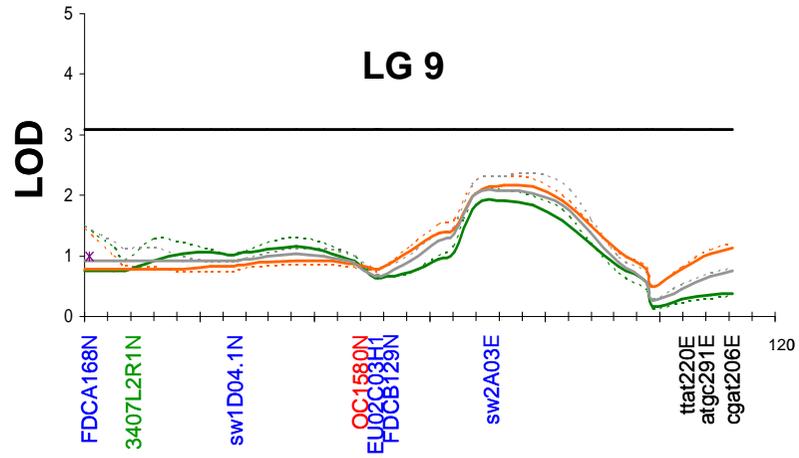
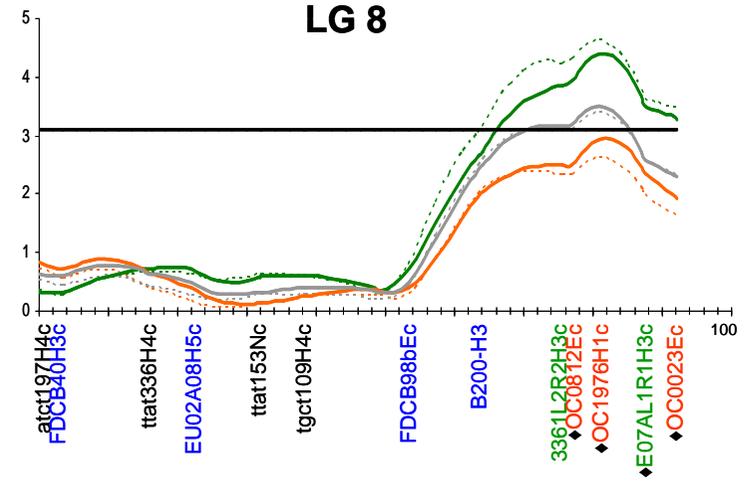
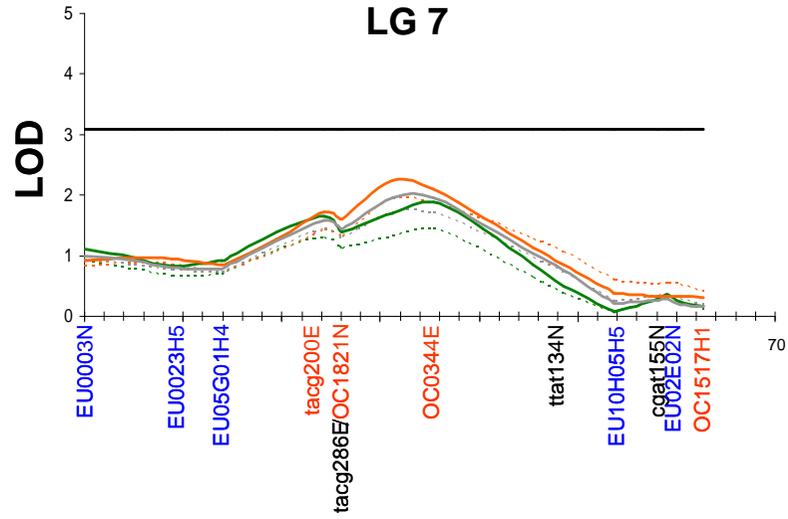
## Annexes

QTL	LG	F1'a			F1'b			F1'a+b		
		POSITION (cM)	LOD	R <sup>2</sup> %	POSITION (cM)	LOD	R <sup>2</sup> %	POSITION (cM)	LOD	R <sup>2</sup> %
<b>PL<sub>90</sub></b>	1	81-93 (86)	3.75	15.3	-	-	-	-	-	-
	2		-	-	21-25 (23)	3.4	14.5	17-38 (23)	5.29	10.4
	3		-	-	40-51 (47)	3.61	15		-	-
	6	20-27 (24)	3.31	12.6		-	-	14-36 (28)	4.04	8.1
	7		-	-		-	-		-	-
	8		-	-	72-92 (80)	3.99	19.1	71-92 (80)	5.46	12.4
				<b>27.9</b>			<b>48.6</b>			<b>30.9</b>
<b>PL<sub>50</sub></b>	1	75-93 (84)	3.96	16.2	-	-	-	-	-	-
	2		-	-		-	-	17-38 (23)	5.09	10.2
	3		-	-	40-45 (43)	3.15	14.6		-	-
	6	11-23 (26)	3.22	11.7		-	-	13-35 (28)	3.56	7.9
	7		-	-		-	-		-	-
	8		-	-	78-82 (80)	3.14	15.3	71-93 (80)	4.61	10.7
				<b>27.9</b>			<b>29.9</b>			<b>28.8</b>
<b>SH<sub>90</sub></b>	1	75-93 (84)	4.97	19.7	-	-	-	76-92 (84)	4.93	11.8
	2		-	-	18-39 (23)	4.96	20	21-34 (23)	4.11	8.1
	3		-	-	34-57 (46)	4.51	18.5	30-43 (21)	3.62	7.9
	6	13-33 (24)	4.04	14.7		3.15	13.6	10-46 (24)	6.13	13.9
	7		-	-		-	-	20-42 (34)	4.11	11.4
	8	66-92 (81)	3.41	17.8	68-92 (80)	5.55	26.5	44-93 (80)	8.44	18.1
				<b>52.2</b>			<b>78.6</b>			<b>57.3</b>
<b>SH<sub>50</sub></b>	1	75-93 (84)	5.02	19.8	-	-	-	79-91 (84)	4.55	10.8
	2		-	-	19-38 (23)	4.48	18.2	21-35 (23)	3.93	8
	3		-	-	39-54 (46)	3.75	16		-	-
	6	13-26 (25)	3.30	12.2		-	-	9-43 (24)	5.4	12.4
	7		-	-		-	-	21-44 (34)	4.46	12.2
	8	64-92 (81)	4.62	18.2	66-92 (80)	5.81	27	44-93 (80)	8.58	18.1
				<b>50.2</b>			<b>61.2</b>			<b>61.5</b>
<b>ET<sub>90</sub></b>	1	77-93 (86)	5.12	20.4	-	-	-	76-92 (84)	4.96	11.1
	2		-	-	19-39 (23)	4.34	19.4	20-35 (23)	4.49	9.1
	3		-	-		-	-		-	-
	6	20-24 (24)	3.19	12.3		-	-	9-43 (24)	5.45	12.9
	7		-	-		-	-	23-39 (34)	3.84	11.4
	8	70-84 (81)	3.50	14.4	76-92 (80)	3.95	19.9	68-93 (80)	6.57	14.4
				<b>47.1</b>			<b>39.3</b>			<b>58.9</b>
<b>ET<sub>50</sub></b>	1	62-93 (84)	4.72	18.6	-	-	-	76-89 (84)	4.03	9.5
	2		-	-	20-38 (23)	4.11	17.1	20-35 (23)	5.05	10.2
	3		-	-	38-50 (44)	3.79	17.9		-	-
	6	13-25 (24)	3.28	13.1		-	-	9-43 (24)	5.21	11.7
	7		-	-		-	-	23-25 (34)	3.1	6.9
	8	70-82 (81)	3.38	13.8	68-92 (80)	4.93	23	68-93 (80)	7.34	15.8
				<b>45.5</b>			<b>58</b>			<b>54.1</b>

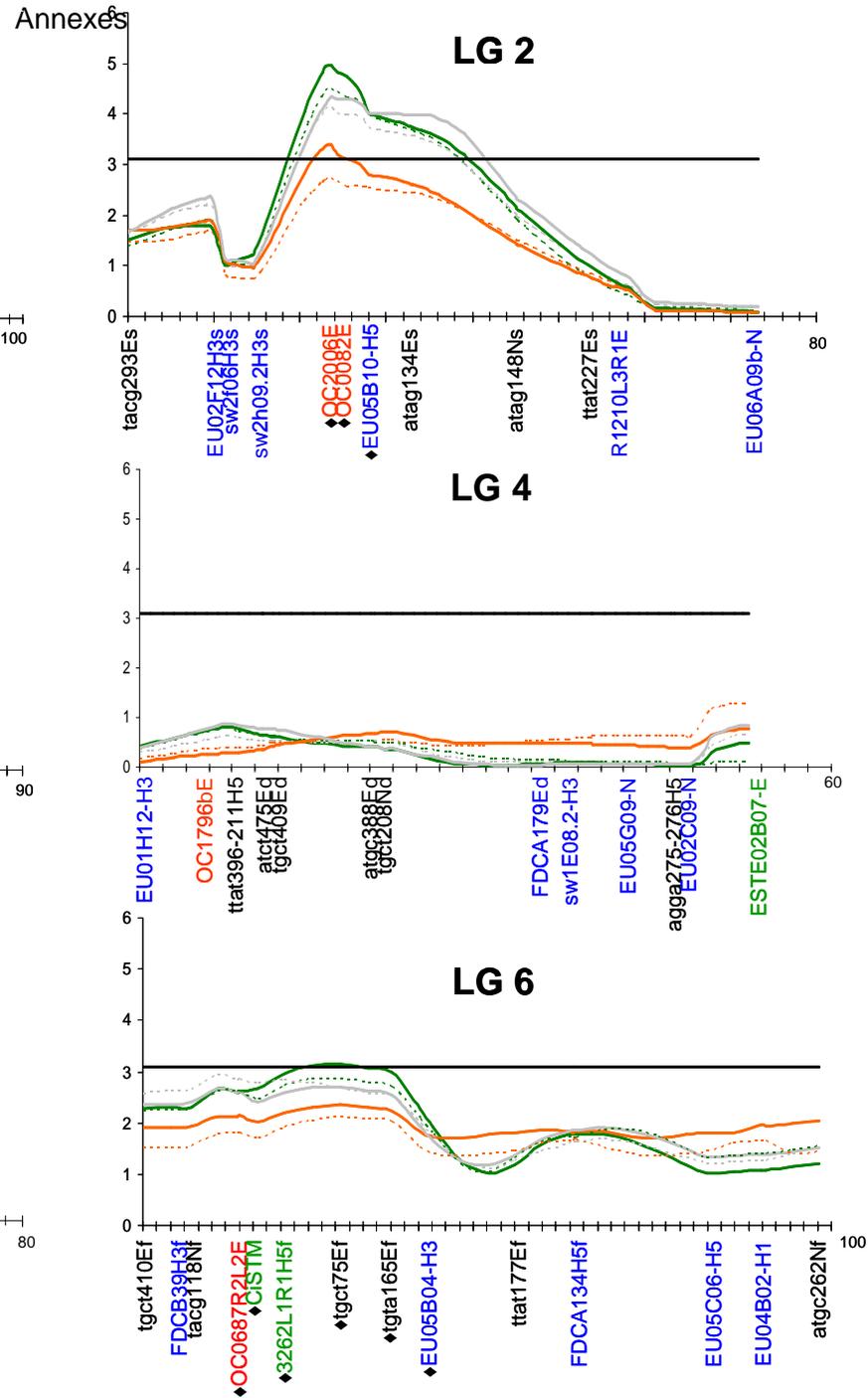
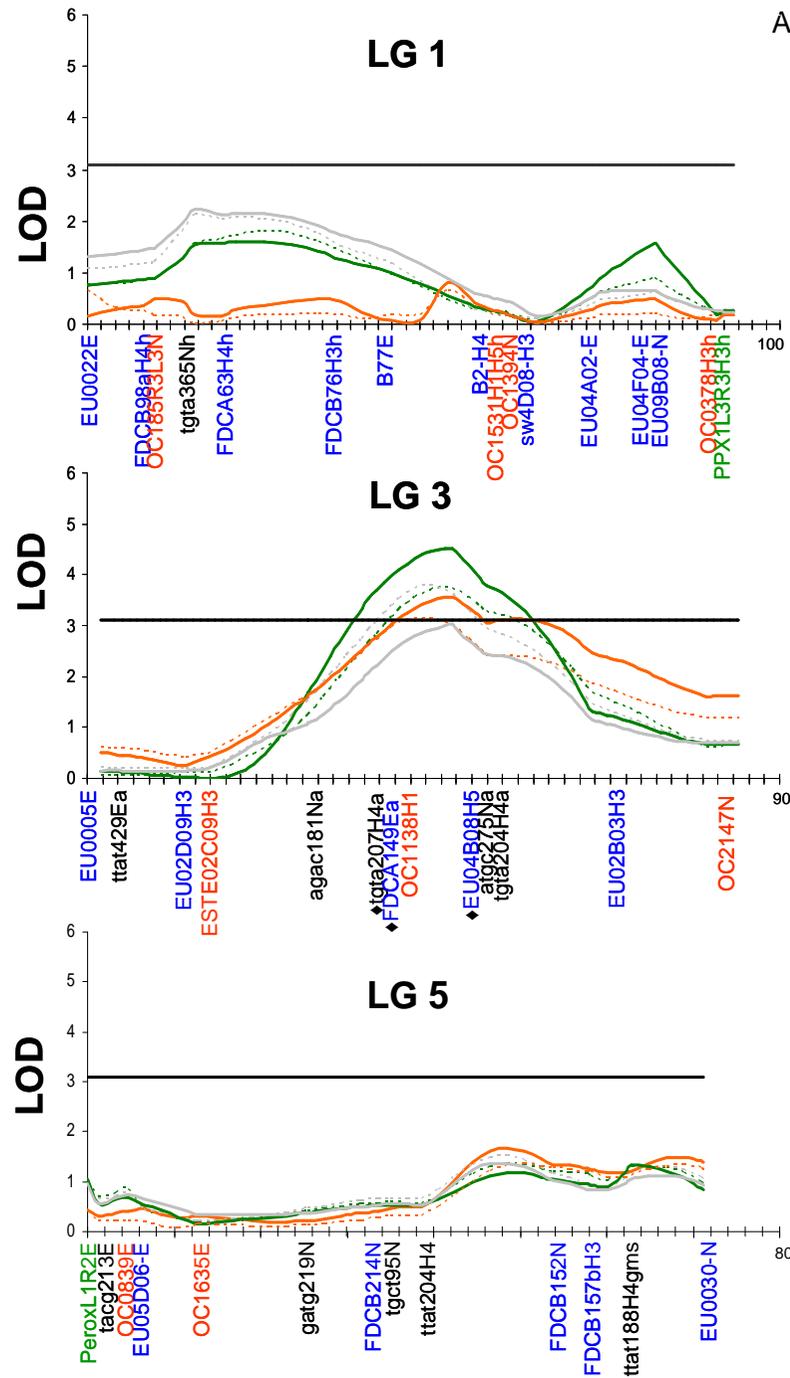
**Annexe 11** : Courbes LOD score obtenues après analyse IM des données plantules (PL ; en orange), structures chlorophylliennes unipolaires (SH ; en vert) et évènements totaux (ET ; en gris) dans la sous-descendance F1'a. L'analyse a été réalisée pour chaque paramètre d'analyse ( $_{90}$ , intensité moyenne de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait en pointillé) et  $_{50}$ , intensité maximale de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait plein). Les régions liées aux QTL PL, SH et ET sont indiquées par un cadre noir. Les marqueurs significatifs aux tests de Kruskal Wallis ( $p < 0.01$ ) sont indiqués par un symbole (♦).

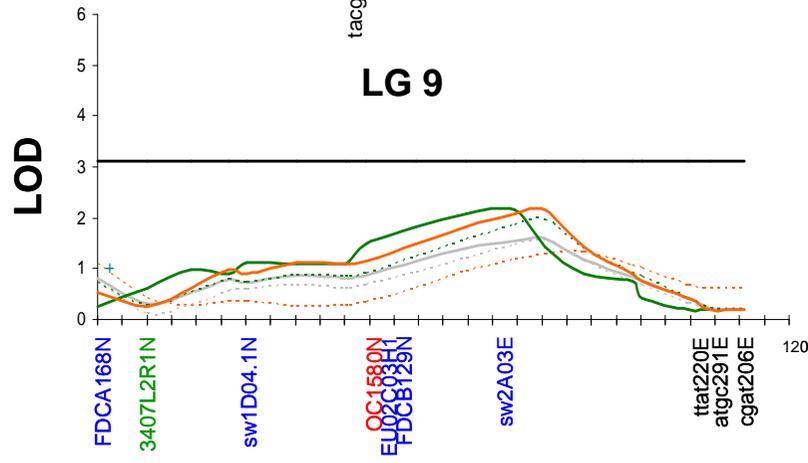
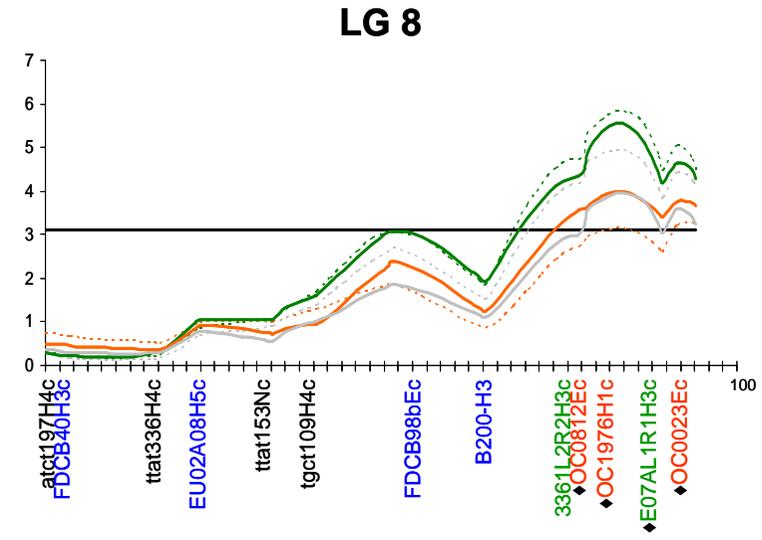
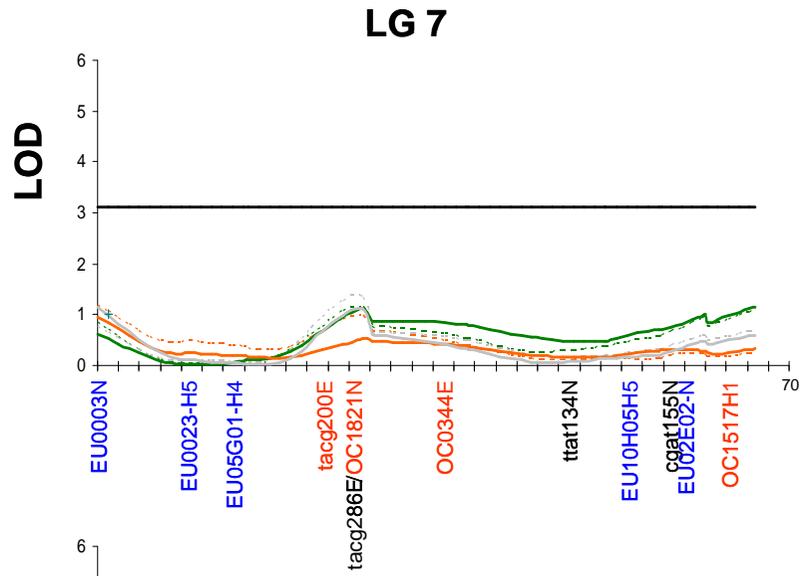


Annexes

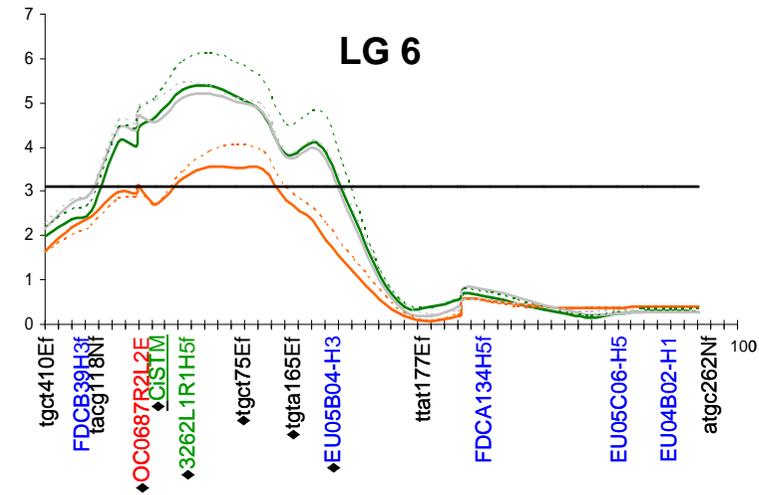
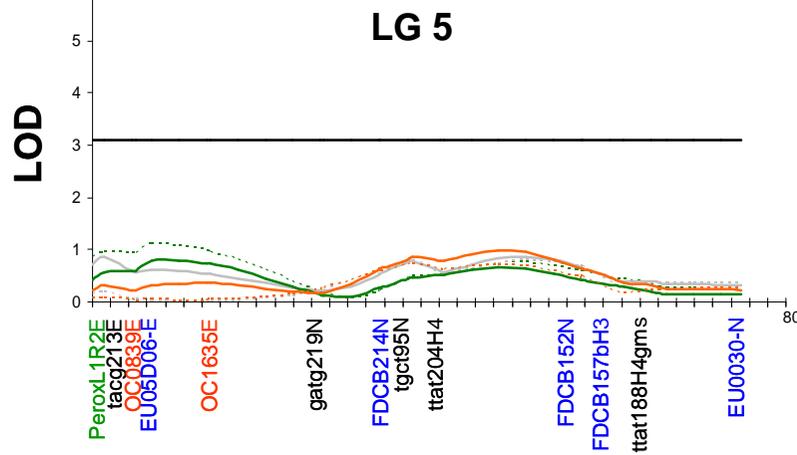
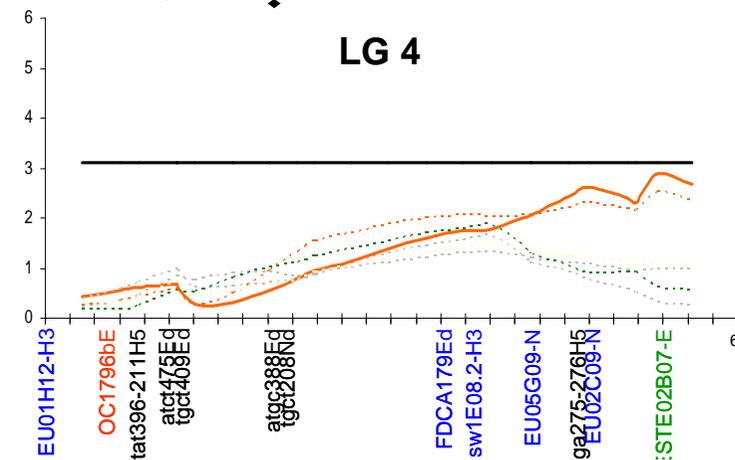
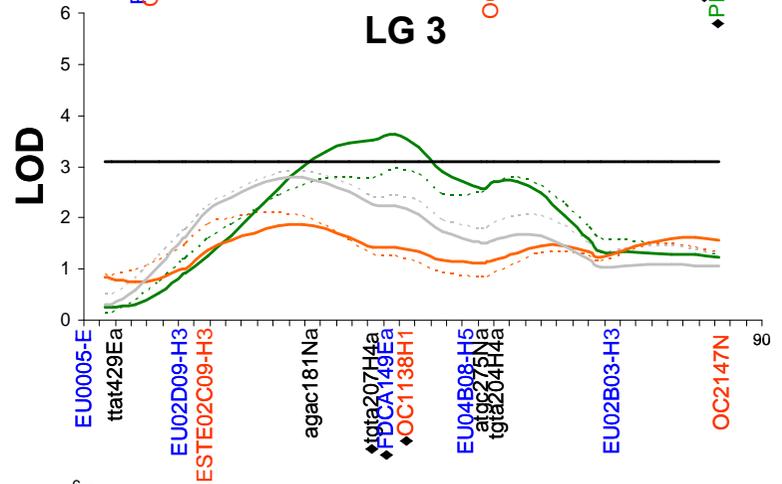
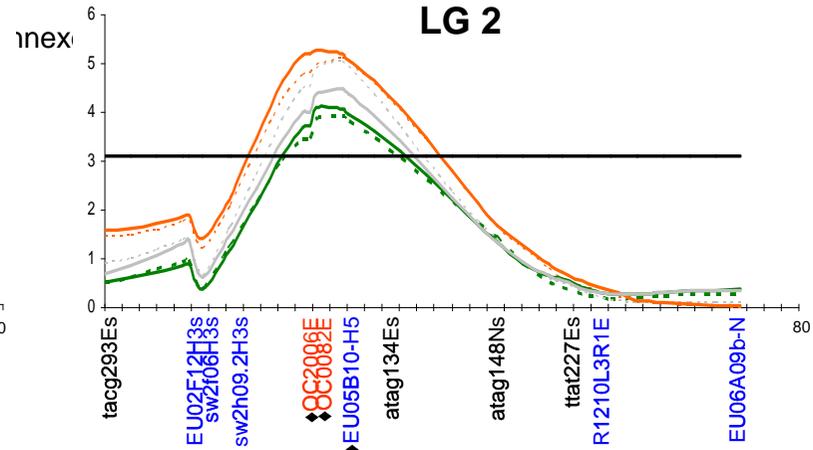
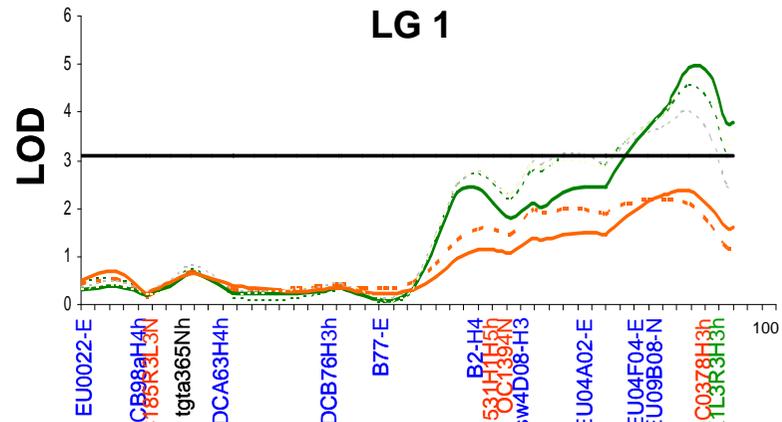


**Annexe 12** : Courbes LOD score obtenues après analyse IM des données plantules (PL ; en orange), structures chlorophylliennes unipolaires (SH ; en vert) et événements totaux (ET ; en gris) dans la sous-descendance F1'b. L'analyse a été réalisée pour chaque paramètre d'analyse ( $\mu_{90}$ , intensité moyenne de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait en pointillé) et  $\mu_{50}$ , intensité maximale de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait plein). Les régions liées aux QTL PL, SH et ET sont indiquées par un cadre noir. Les marqueurs significatifs aux tests de Kruskal Wallis ( $p < 0.01$ ) sont indiqués par un symbole (♦).

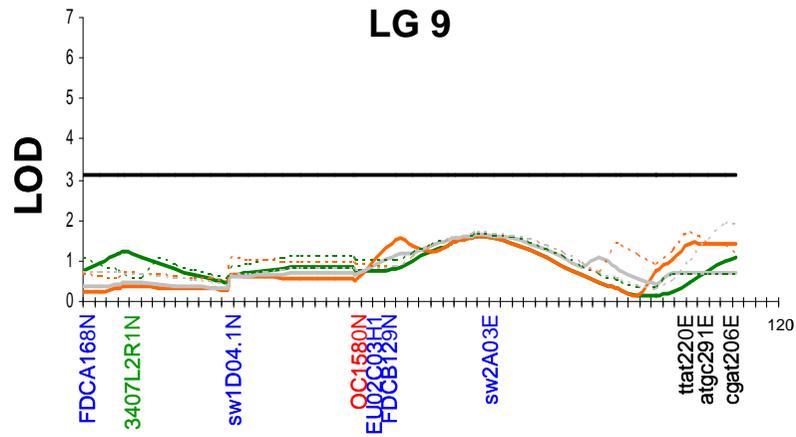
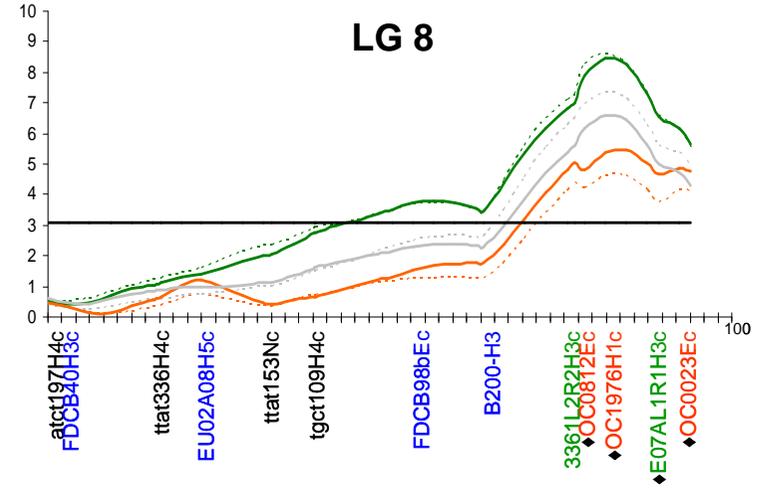
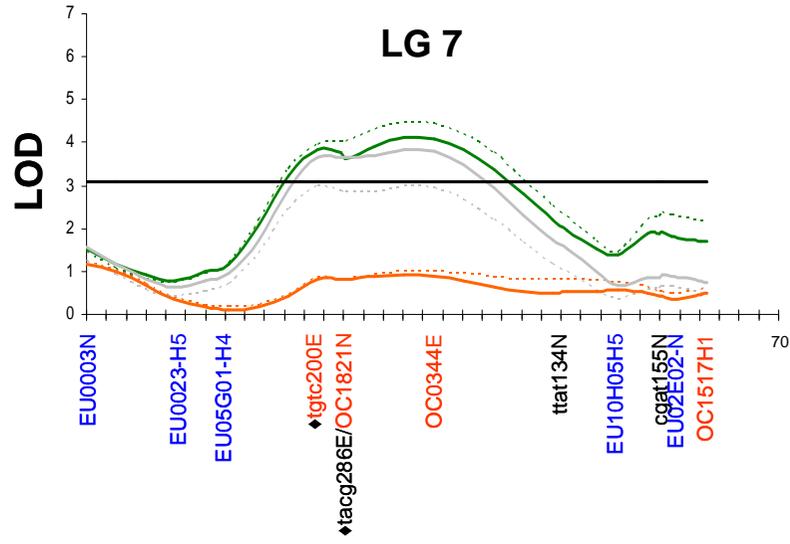




**Annexe 13** : Courbes LOD score obtenues après analyse IM des données plantules (PL ; en orange), structures chlorophylliennes unipolaires (SH ; en vert) et événements totaux (ET ; en gris) dans la descendance F1'a+b. L'analyse a été réalisée pour chaque paramètre d'analyse ( $\mu_{90}$ , intensité moyenne de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait en pointillé) et  $\mu_{50}$ , intensité maximale de réponse liée à l'embryogenèse somatique et à la formation des structures chlorophylliennes unipolaires (trait plein). Les régions liées aux QTL PL, SH et ET sont indiquées par un cadre noir. Les marqueurs significatifs aux tests de Kruskal Wallis ( $p < 0.01$ ) sont indiqués par un symbole (♦).



Annexes



**Annexe 14:** Effets alléliques observés pour les marqueurs liés significativement aux QTL impliqués dans la formation des plantules (PL), des structures chlorophylliennes unipolaires (SH) et des événements totaux (ET) pour chaque paramètre d'analyse (PL<sub>50</sub>, PL<sub>90</sub>, SH<sub>50</sub>, SH<sub>90</sub>, ET<sub>50</sub>, ET<sub>90</sub>)

F1'a					F1'b			
LG1 sw4D08-H3 <efxeg> <K59xK28>								
	ee	ef	eg	fg	ee	ef	eg	fg
PL50	1.57	1.22	1.56	1.78	1.27	1.23	1.30	1.26
PL90	1.34	1.00	1.35	1.51	1.00	0.98	1.06	1.00
SH50	1.45	1.08	1.51	1.63	1.86	1.82	1.96	1.91
SH90	1.21	0.88	1.32	1.40	1.62	1.55	1.74	1.62
TOT50	1.83	1.43	1.84	2.03	1.95	1.94	2.04	2.00
TOT90	1.57	1.20	1.61	1.78	1.81	1.75	1.91	1.87

F1'a			F1'b	
LG2 OC0082 <lmxll> <K59xK28>				
	ll	lm	ll	lm
PL50	1.35	1.69	1.17	1.41
PL90	1.11	1.47	0.89	1.18
SH50	1.20	1.59	1.75	2.07
SH90	1.00	1.37	1.47	1.86
TOT50	1.57	1.96	1.85	2.16
TOT90	1.32	1.73	1.70	2.01

F1'a			F1'b	
LG3 FDCA149 <lmxll> <K59xK28>				
	ll	lm	ll	lm
PL50	1.35	1.61	1.16	2.04
PL90	1.11	1.39	0.89	2.09
SH50	1.20	1.50	1.80	2.10
SH90	1.00	1.30	1.51	2.06
TOT50	1.57	1.87	1.88	2.17
TOT90	1.32	1.63	1.76	2.12

F1'a			F1'b	
LG4 FDCA179 <lmxll> <K59xK28>				
	ll	lm	ll	lm
PL50	1.63	1.41	1.30	1.29
PL90	1.41	1.16	1.04	1.04
SH50	1.48	1.33	1.89	1.94
SH90	1.26	1.11	1.65	1.69
TOT50	1.87	1.66	1.99	2.02
TOT90	1.63	1.42	1.85	1.87

## Annexes

F1'a

LG6 CiSTM <lmxll> <K59xK28>

	ll	lm
<b>PL50</b>	1.76	1.40
<b>PL90</b>	1.55	1.16
<b>SH50</b>	1.62	1.29
<b>SH90</b>	1.42	1.06
<b>TOT50</b>	2.01	1.65
<b>TOT90</b>	1.76	1.41

F1'b

	ll	lm
	1.37	1.19
	1.11	0.94
	2.02	1.76
	1.78	1.50
	2.10	1.87
	1.96	1.71

F1'a

LG7 tacg286 <lmxll> <K59xK28>

	ll	lm
<b>PL50</b>	1.44	1.65
<b>PL90</b>	1.19	1.45
<b>SH50</b>	1.33	1.52
<b>SH90</b>	1.10	1.32
<b>TOT50</b>	1.68	1.90
<b>TOT90</b>	1.43	1.68

F1'b

	ll	lm
	1.38	1.27
	1.13	1.01
	1.93	1.91
	1.71	1.65
	2.04	1.99
	1.90	1.84

F1'a

LG8 OC1976H1 <hkxhk> <K59xK28>

	hh	hk	kk
<b>PL50</b>	1.69	1.60	1.29
<b>PL90</b>	1.50	1.35	1.05
<b>SH50</b>	1.63	1.50	1.08
<b>SH90</b>	1.45	1.25	0.88
<b>TOT50</b>	1.98	1.85	1.49
<b>TOT90</b>	1.77	1.60	1.24

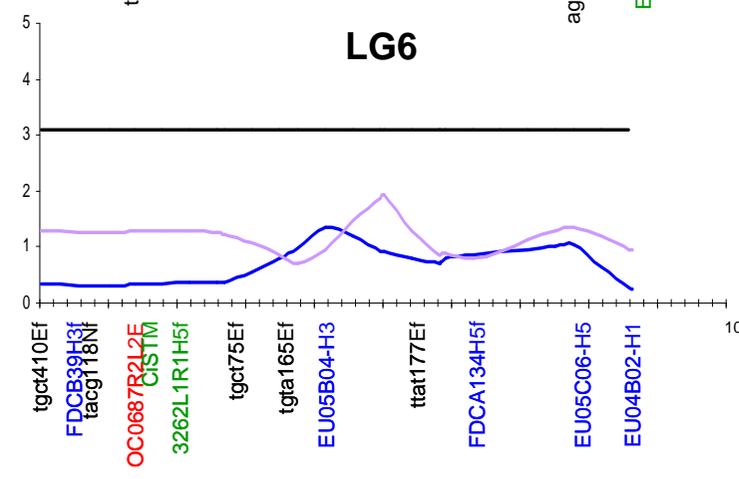
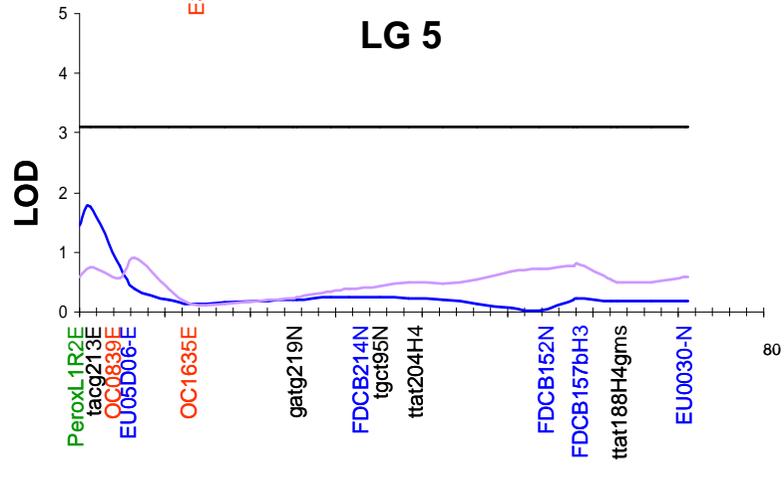
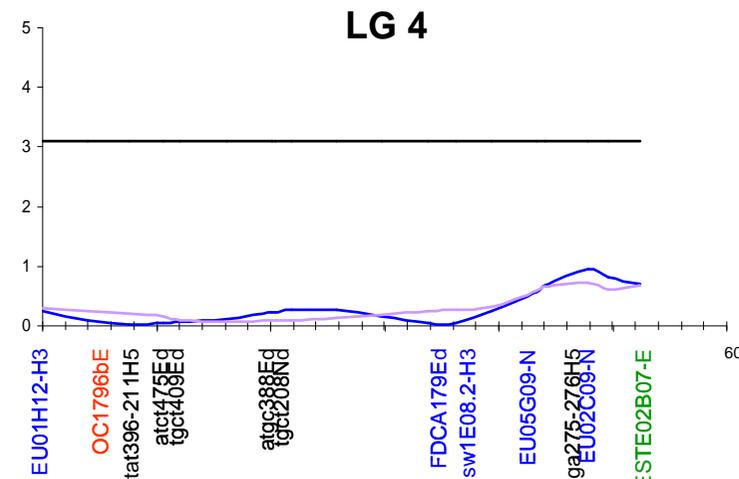
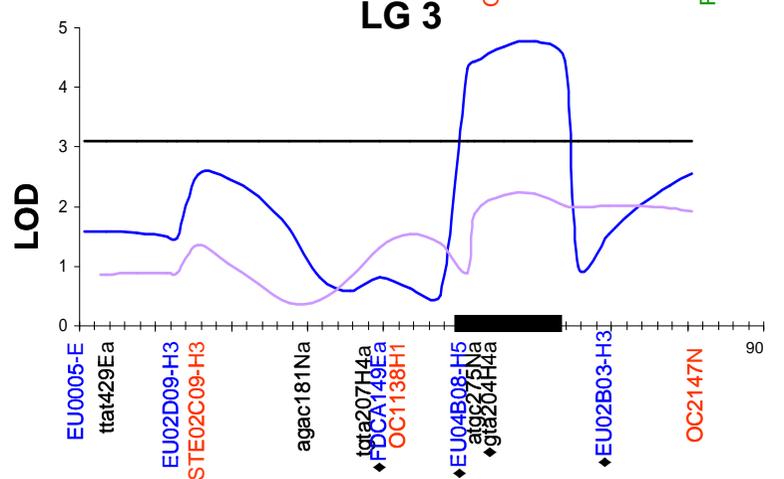
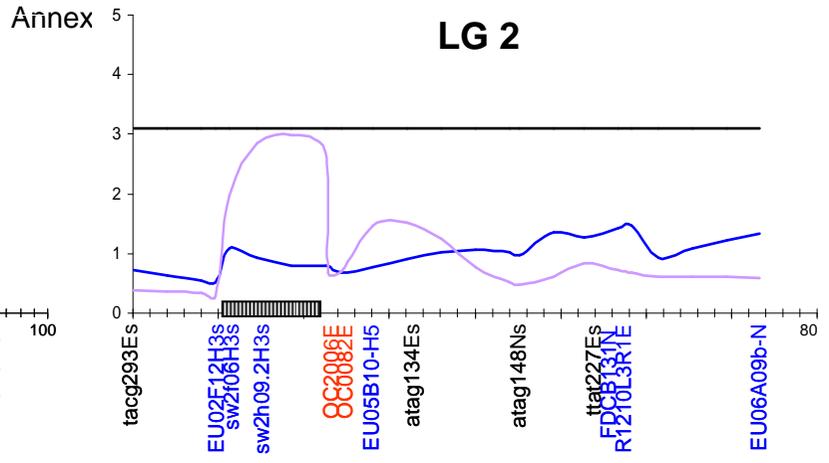
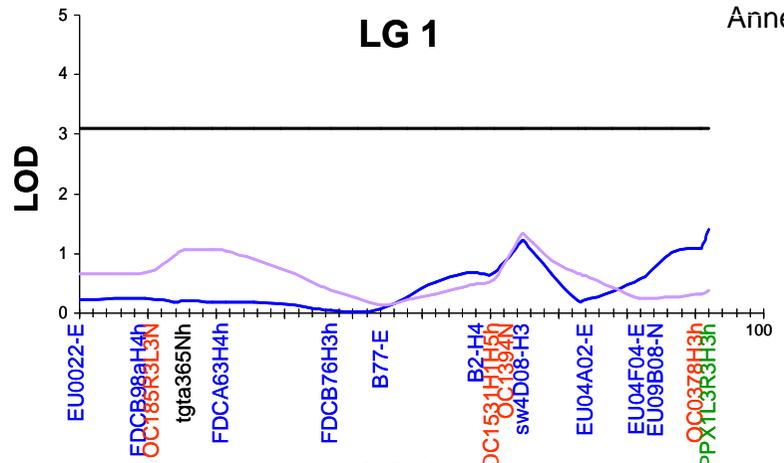
F1'b

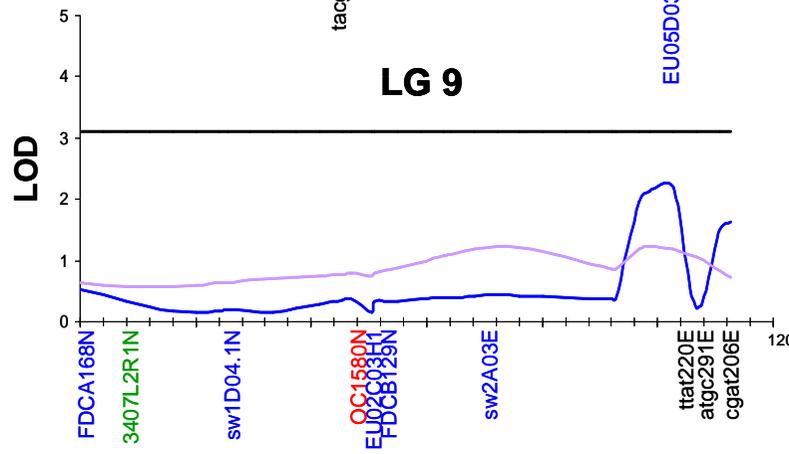
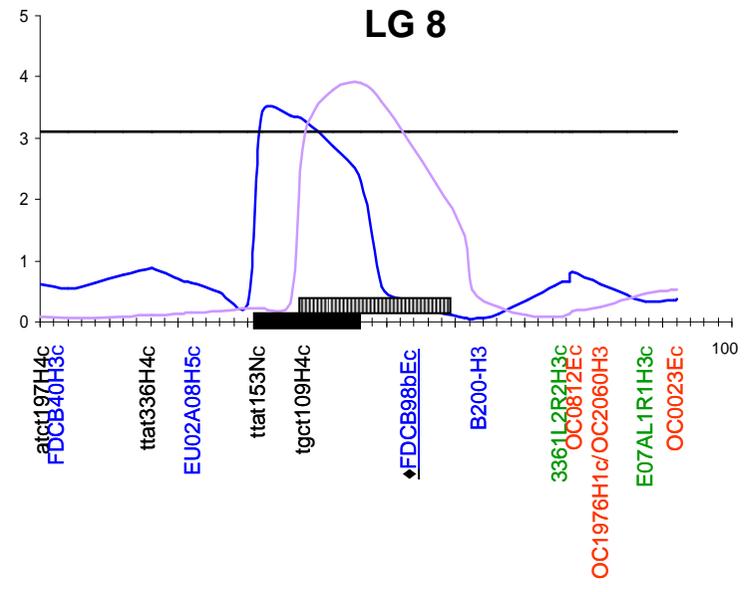
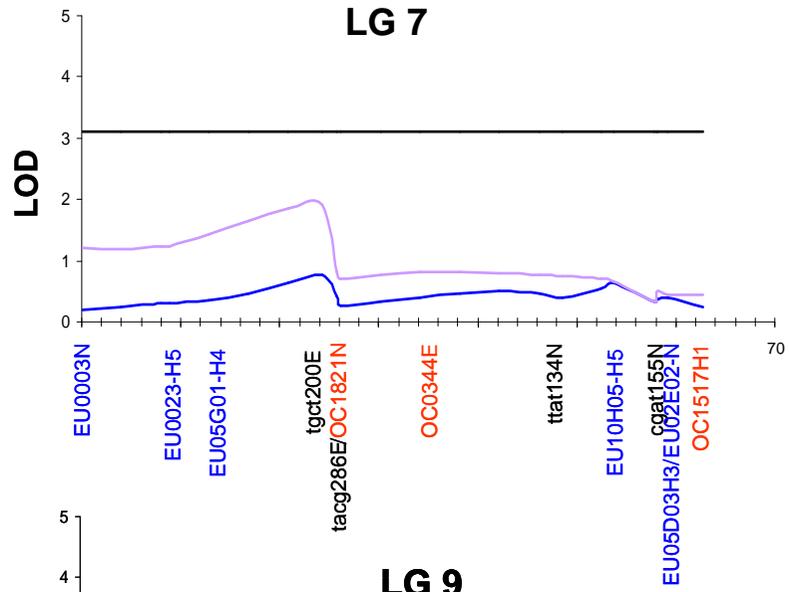
	hh	hk	kk
	1.46	1.32	1.00
	1.26	1.05	0.70
	2.11	1.94	1.55
	1.91	1.68	1.27
	2.21	2.02	1.66
	2.03	1.87	1.55

**Annexe 15:** Analyse IM (Interval Mapping). Identification des QTL liés à la formation des cals (Cal) et des bourgeons (Bg) dans la sous-descendance F1'b. La position du QTL (cM), le LOD score (LoD) et le pourcentage de variation expliquée ( $R^2$ ) sont indiqués pour chaque QTL. Les groupes de liaison non liés significativement à un QTL sont indiqués par –

QTL	LG	POSITION (cM)	LOD	R <sup>2</sup> %
<b>Cal</b>	8	46-53 (49)	3.07	15.7
				<b>15.7</b>
<b>bg</b>	3	51-66 (56)	3.51	18
				<b>18</b>

**Annexe 16** : Courbes LOD score obtenues après analyse MQM des données cals (Cal ; en mauve) et bourgeons (Bg ; en bleue) dans la sous-descendance F1'b. Les régions liées aux QTL Cal et Bg sont indiquées par un cadre rayé et noir, respectivement.





Annexes

**Annexe 17 :** Effets additifs des QTL majeurs (QTL2, QTL6 et QTL8) liés à la formation des plantules (PL90). Les géotypes de la sous-descendance F1'a et F1'b sont indiqués en noir et rouge, respectivement.

Géotypes	QTL2	QTL6	QTL8	PL90 (log10)
	OC0082	C/STM	OC1976	
B666				0.00
B477				0.00
B315				0.11
B672				0.13
B265				0.14
B650				0.19
B317				0.20
B634				0.24
B648				0.25
B656				0.28
B668				0.29
B675				0.29
B611				0.30
B511				0.30
B659				0.31
B605				0.32
B319				0.36
B307				0.37
B614				0.41
B621				0.43
B626				0.43
B642				0.47
B607				0.49
B212				0.49
B514				0.49
B658				0.49
B311				0.49
B604				0.52
B647				0.52
B281				0.53
B409				0.55
B632				0.55
B638				0.56
B260				0.56
B449				0.56
B615				0.57
B296				0.58
B269				0.59
B651				0.60
B623				0.62
B205				0.64
B214				0.66
B655				0.68
B235				0.70
B676				0.70
B627				0.70
B671				0.71
B679				0.72
B649				0.73
B360				0.76
B622				0.77
B678				0.82

## Annexes

B209		0.83
B351		0.88
B266		0.89
B343	████████	0.89
B630	████████	0.90
B289	████████	0.90
B283	████████	0.91
B222	████████	0.93
B624	████████	0.96
B359		0.97
B234	████████	0.98
B681	████████	0.98
B407		1.00
B255		1.01
B206		1.02
B301		1.03
B629	████████	1.04
B496	████████	1.05
B342	████████	1.05
B618	████████	1.06
B233		1.06
B674	████████	1.08
B637		1.08
B646	████████	1.09
B362		1.10
B272	████████	1.10
B669	████████	1.10
B644	████████	1.11
B401	████████	1.11
B428	████████	1.11
B613	████████	1.11
B253	████████	1.11
B241	████████	1.11
B497	████████	1.11
B643	████████	1.12
B601	████████	1.15
B525	████████	1.15
B476	████████	1.16
B340	████████	1.18
B361	████████	1.19
B504	████████	1.19
B252		1.20
B431	████████	1.20
B324	████████	1.20
B308	████████	1.21
B619	████████	1.21
B314	████████	1.22
B641	████████	1.22
B628	████████	1.23
B434	████████	1.23
B635	████████	1.23
B375	████████	1.24
B261	████████	1.24
B663	████████	1.29
B602	████████	1.29
B410	████████	1.30
B325	████████	1.30
B295	████████	1.31
B616	████████	1.32
B620	████████	1.33
B316	████████	1.33
B667	████████	1.33
B402	████████	1.35

## Annexes

B612		1.35
B617		1.35
B606		1.36
B285		1.37
B221		1.38
B298		1.39
B445		1.40
B240		1.41
B322		1.42
B631		1.42
B657		1.44
B680		1.44
B350		1.45
B374		1.45
B267		1.46
B417		1.47
B216		1.47
B480		1.47
B633		1.47
B609		1.48
B207		1.48
B219		1.49
B364		1.51
B238		1.53
B608		1.53
B223		1.54
B677		1.54
B664		1.55
B330		1.56
B660		1.56
B274		1.57
B603		1.57
B670		1.58
B328		1.59
B438		1.60
B682		1.60
B639		1.60
B329		1.61
B523		1.62
B653		1.64
B299		1.65
B358		1.65
B625		1.66
B341		1.66
B610		1.66
B673		1.66
B371		1.66
B229		1.67
B304		1.68
B377		1.68
B213		1.69
B356		1.70
B640		1.70
B662		1.71
B332		1.73
B661		1.73
B645		1.73
B215		1.75
B522		1.75
B665		1.76
B236		1.76
B279		1.76
B320		1.76

## Annexes

B338	[REDACTED]	1.76
B337	[REDACTED]	1.78
B334	[REDACTED]	1.78
B273	[REDACTED]	1.78
B218	[REDACTED]	1.78
B654	[REDACTED]	1.79
B237	[REDACTED]	1.80
B326	[REDACTED]	1.80
B239	[REDACTED]	1.80
B228	[REDACTED]	1.81
B354	[REDACTED]	1.84
B302	[REDACTED]	1.84
B312	[REDACTED]	1.84
B263	[REDACTED]	1.85
B652	[REDACTED]	1.87
B331	[REDACTED]	1.87
B349	[REDACTED]	1.87
B357	[REDACTED]	1.88
B347	[REDACTED]	1.91
B278	[REDACTED]	1.94
B297	[REDACTED]	1.98
B378	[REDACTED]	2.01
B202	[REDACTED]	2.02
B248	[REDACTED]	2.03
B372	[REDACTED]	2.05
B327	[REDACTED]	2.07
B226	[REDACTED]	2.09
B293	[REDACTED]	2.15
B294	[REDACTED]	2.16
B370	[REDACTED]	2.19
B276	[REDACTED]	2.20
B277	[REDACTED]	2.22
B251	[REDACTED]	2.24
B246	[REDACTED]	2.25
B258	[REDACTED]	2.30
B286	[REDACTED]	2.32
B379	[REDACTED]	2.35
B352	[REDACTED]	2.40
B310	[REDACTED]	2.42
B243	[REDACTED]	2.42
B344	[REDACTED]	2.48
B323	[REDACTED]	2.50
B300	[REDACTED]	2.51
B245	[REDACTED]	2.52
B242	[REDACTED]	2.59
B264	[REDACTED]	2.60
B291	[REDACTED]	2.63
B262	[REDACTED]	2.65
B201	[REDACTED]	2.70

---

Annexes

**Annexe 18 : Alignements des séquences Argonaute d'*Arabidopsis* et la séquence chicorée (Ci. CSA1.6010).**

		10	20	30	40	50	60	
At1g48410	AGO1	1	PPWLAQMCYV	SGTSKLSQGK	EIDLLIVILP	-DNNGSLYGD	LKRICETELG	IVSQCCCL---
At1g31280	AGO2	1	APSLIDRCWR	LGASRKHGGA	RPTLVLCAMS	--RKDDGYKT	LKWI AETKLG	LVTQCFLTGP
At1g31290	AGO3	1	APTLIERCKG	LGASHNHGGA	CPTLVLCAMT	--GKHDGYKT	LKWI AETKLG	LVTQCFLTIS
At2g27040	AGO4	1	APKLIKIGGS	KGIQSKLPGV	P-QFILCVLP	DKKNSDLYGP	WKKKNLTFEG	IVTQCMAG---
At2g27880	AGO5	1	PPMLIGMCVS	KG IHKRAPG-	-LQLLIVILP	-DVTG-SYGK	IKRICETELG	IVSQCCQ---
At2g32940	AGO6	1	PPMLISCGMR	KGMKLFKFPD	P-HFILCILP	ERKTSDIYGP	WKKICLTFEG	IHTQCCIC---
At1g69440	AGO7	1	PPKFINELTQ	KCLKETIQRAA	SNNLQLIICV	MEKKHKGYGD	LKRISETRIG	VVTQCCCL---
At5g21030	AGO8	1	TPKLTRCGKM	KGLQSI LGEV	PPKFLLCILE	-KKNSDVY--	--EKSCSMWN	--CECIV---
At5g21150	AGO9	1	TPMLIKCGRE	KGIKSKLPKP	P-LFLLCILA	ERKNSDVYGP	WKKKNLVDLG	IVTQCIA---
At5g43810	AGO10/PINH-ZLL	1	APWLGQMCVE	SGMNKTG-GK	ELELLAILP	-DNNGSLYGD	LKRICETELG	LISQCCCL---
Ci.CSA1.6010: 3 to 1196: Frame		1	PLVIQSQCIA	Q-LKTVPAGN	VLQLLFVILP	-EAKG-TYAR	IKRVCETELG	IVSQCCCK---

		70	80	90	100	110	120	
At1g48410	AGO1	57	TKHVPKMSKQ	YMANVALKIN	VKVGGRNIVL	VDALSRRIIP-	-LVSDRPTII	FGADVTHPHF
At1g31280	AGO2	59	ATKGG---DQ	YRANLALKMN	AKVGGSNVEL	MD-TFSFFK-	---KEDEVMF	IGADVNHPAA
At1g31290	AGO3	59	AIKGETVSDQ	YLANLALKIN	AKVGGTNVEL	VDNIFSFFK-	---KEDKVMF	IGADVNHPAA
At2g27040	AGO4	57	PTRQ--PNDQ	YLTNLLKIN	AKLGGLSNML	SVERTPAFT-	-VISKVPTII	LGMDVSHGSP
At2g27880	AGO5	54	PRQVNKLNKQ	YMENVALKIN	VKTGGRNIVL	NDAIRRNIP-	-LITDRPTII	MGADVTHPQP
At2g32940	AGO6	57	PIK---ISDQ	YLTNVLKIN	SKLGGINSLL	GIEYSYNIP-	-LINKIPTLI	LGMDVSHGPP
At1g69440	AGO7	58	YPNITKLSQ	FVSNLALKIN	AKIGGSMTEL	YNSIPSHIPR	LLRPDEPVIF	MGADVTHPHF
At5g21030	AGO8	51	PPQN--LNDQ	YLTNLLKIN	AKLGGLSNVL	DMELSGTMP-	-LVMRVPTII	IGMDVSHGSP
At5g21150	AGO9	57	PTR---LNDQ	YLTNVLKIN	AKLGGLSNLL	AMERSPAMP-	-KVTQVPTII	VGMDVSHGSP
At5g43810	AGO10/PINH-ZLL	56	TKHVPKISKQ	YLANVSLKIN	VKMGGRNIVL	VDAISCRIP-	-LVSDIPTII	FGADVTHPEN
Ci.CSA1.6010: 3 to 1196: Frame		55	PQHVPKKSPO	YFENVALKIN	VKVGGRNIVL	AAALANRLP-	-YITDRPTII	FGADVTHPSF

		130	140	150	160	170	180	
At1g48410	AGO1	115	GEDS-SPSIA	AVVASQDWPE	ITKYAGLVCA	QAHRQELIQD	LFKEWKDPQK	GVVTGGMIKE
At1g31280	AGO2	111	RDKM-SPSIV	AVVGTLNWPE	ANRYAARVIA	QPHRKEETIQG	FG-----	-----DACLE
At1g31290	AGO3	115	HDNM-SPSIV	AVVGTLNWPE	ANRYAARVKA	QSHRKEETIQG	FG-----	-----ETCWE
At2g27040	AGO4	113	GQSD-VPSIA	AVVSSREWPL	ISKYRASVRT	QPSKAEMIES	LVKKNG----	TEDD-GIIKE
At2g27880	AGO5	112	GEDS-SPSIA	AVVASMDWPE	INKYRGLVSA	QAHREETIQD	LYKLVQDPQR	GLVHSGLIIE
At2g32940	AGO6	112	GRAD-VPSVA	AVVGSKCWPL	ISRYRAAVRT	QSPRLEMIDS	LFQPIENT--	EKGDNMIMNE
At1g69440	AGO7	117	-FDDCSPSVA	AVVGSINWPE	ANRYVSRMRS	QTHRQETIQD	LD-----	-----LMVKE
At5g21030	AGO8	107	GQSDHIPSIA	AVVSSREWPL	ISKYRACVRT	QSPKVEMIDS	LFKPVV----	DKDDQGIMRE
At5g21150	AGO9	112	GQSD-IPSI	AVVSSRWPL	ISKYKACVRT	QSRKMEMIDN	LFKPVN----	GKDE-GMFRE
At5g43810	AGO10/PINH-ZLL	114	GEES-SPSIA	AVVASQDWPE	VTKYAGLVCA	QAHRQELIQD	LYKTWQDPVR	GTVSGGMIRD
Ci.CSA1.6010: 3 to 1196: Frame		113	GEDS-IPSI	AVVASMDWPQ	VTKYKALVSA	QPHRQETIED	LYSTTSDAKR	GVVHSGLIIE

		190	200	210	220	230	240	
At1g48410	AGO1	174	LLIAFRRSTG	H-KPLRIIFY	RDGVSEGOFY	QVLLYELDAI	RRACASLEAG	YQPPVTFVVV
At1g31280	AGO2	157	LVKAHVQATG	K-RPNKIVIF	RDGVSDAQFD	MVLNVELLDV	KLTFEKNG--	YNPKITVIVA
At1g31290	AGO3	161	LIEAHSQAPE	K-RPNKIVIF	RDGVSDGQFD	MVLNVELQNV	KDVFAKVG--	YNPQITVIVA
At2g27040	AGO4	167	LLVDFYTSSN	KRKPETHIIF	RDGVSESQFN	QVLNIELDQI	IEACKLLDAN	WNPKFLLLVA
At2g27880	AGO5	171	HFIAFRRATG	Q-IPQRIIFY	RDGVSEGOFS	QVLLHEMTAI	RRACNSLOEN	YVPRVTFVIV
At2g32940	AGO6	169	LFVEFYRTSR	ARKPKQIIFY	RDGVSESQFE	QVLKIEVDQI	IKAYQRLGES	DVPKFTVIVA
At1g69440	AGO7	164	LLDDFYKAVK	K-LPNRIIFF	RDGVSETQFK	KVLQEELQSI	KTACSKFQD-	YNPSITFAVV
At5g21030	AGO8	163	LLLDFHSSSG	K-KPNHIIFF	RDGVSESQFN	QVLNIELDQM	MQ-----	-----
At5g21150	AGO9	166	LLLDFYYSSE	NRKPEHIIFF	RDGVSESQFN	QVLNIELDQM	MQACKFLDDT	WHPKFTVIVA
At5g43810	AGO10/PINH-ZLL	173	LLISFRKATG	Q-KPLRIIFY	RDGVSEGOFY	QVLLYELDAI	RRACASLEPN	YQPPVTFIVV
Ci.CSA1.6010: 3 to 1196: Frame		172	LLISFKKSTG	H-KPHRIIFY	RDGVSEGOFN	DVLLNEMDKI	RRACASLEEG	YLPVTFVVVV

## Annexes

		250	260	270	280	290	300
At1g48410 AGO1	233	QKRHHTRLFA	QNHNDRHSVD	RSGNILPGTV	VDSKICHPTTE	FDLYLCSHAG	IQTSRPAHY
At1g31280 AGO2	214	QKRHQTRFFP	ATNN-DGSDK	--GNVPSGTV	VDTKVIHPYE	YDFYLCSHHG	GIGTSKPPTHY
At1g31290 AGO3	218	QKRHQTRFFP	ATTSKDGRAK	--GNVPSGTV	VDTTIIHPFE	YDFYLCSHHG	AIGTSKPPTHY
At2g27040 AGO4	227	QKNHHTKFFQ	PTSPE-----	---NVPPGTI	IDNKICHKPN	NDFYLCAHAG	MIGTTRPPTHY
At2g27880 AGO5	230	QKRHHTRLFP	EQHGNRDMTD	KSGNIQPGTV	VDTKICHPTNE	FDLYLNSHAG	IQTSRPAHY
At2g32940 AGO6	229	QKNHHTKLFQ	AKGPE-----	---NVPAGTV	VDTKIVHPTN	YDFYMCASHAG	KIGTSRPAHY
At1g69440 AGO7	222	QKRHHTRLFR	CDPDH-----	--ENIPPGTV	VDTVIITHPKE	FDLYLCSHLG	VKGTSRPPTHY
At5g21030 AGO8	203	-INHHTKFFQ	TESPN-----	---NVLPGTI	IDSNICHQHN	NDFYLCAHAG	KIGTTRPPTHY
At5g21150 AGO9	226	QKNHHTKFFQ	SRGPD-----	---NVPPGTI	IDSQICHPRN	FDLYLCAHAG	MIGTTRPPTHY
At5g43810 AGO10/PINH-ZLL	232	QKRHHTRLFA	NNHRDKNSTD	RSGNILPGTV	VDTKICHPTTE	FDLYLCSHAG	IQTSRPAHY
Ci.CSA1.6010: 3 to 1196: Frame	231	QKRNHTRFFP	MKHGDRATTD	RSGNILPGTV	VDTKICHPTTE	NDFYLCSHAG	IQTSRPPTHY

		310	320	330	340	350	360
At1g48410 AGO1	293	HVLWDENNFT	ADGLQSLTNN	LCYTYARCTR	SVS-----	-IVPPAYYAH	LAAFRRARYM
At1g31280 AGO2	271	YTLWDELGFT	SDQVQKLIFE	MCFTFTRCTK	PVS-----	-LVPPVYYAD	MVAFRGRMYH
At1g31290 AGO3	276	YVLSDEIGFN	SNQIQKLIFD	LCFTFTRCTK	PVA-----	-LVPPVSYAD	KAASRGRVYY
At2g27040 AGO4	279	HVLYDEIGFS	ADELQELVHS	LSYVYQRSTS	AIS-----	-VVAPICYAH	LAAAQLGTFM
At2g27880 AGO5	290	HVLDEENGFT	ADQLQMLTNN	LCYTYARCTK	SVS-----	-IVPPAYYAH	LAAFRRARYM
At2g32940 AGO6	281	HVLDEIGFS	PDDLQNLHS	LSYKLLNSIF	NVSSLLCVFV	LSVAPVRYAH	LAAAQVAQFT
At1g69440 AGO7	275	HILWDENEFT	SDELQRLVYN	LCYTFVRCTK	PIS-----	-IVPPAYYAH	LAAVRGRLYI
At5g21030 AGO8	255	HVLYDEIGFD	TDQLQELVHS	LSYVYQRSTT	AIS-----	-LVAPICYAH	LAAAQMAMATAM
At5g21150 AGO9	278	HVLYDEIGFA	TDDLQELVHS	LSYVYQRSTT	AIS-----	-VVAPVRYAH	LAAAQMGTVM
At5g43810 AGO10/PINH-ZLL	292	HVLWDENNFT	ADGIQSLTNN	LCYTYARCTR	SVS-----	-IVPPAYYAH	LAAFRRARYL
Ci.CSA1.6010: 3 to 1196: Frame	291	HVLHDENKFT	ADGLQNLTNS	LCYTYQRCTR	SVS-----	-IVPPAYYAH	LAAFHARYYM

## Résumé

### Introduction

L'embryogenèse somatique (ES) est une voie de la reproduction asexuée aboutissant à la formation d'embryons à partir de cellules somatiques par un processus qui ressemble à l'embryogenèse zygotique. Chez la chicorée (*Cichorium intybus* L.), une variabilité génétique pour la capacité à former des embryons somatiques *in vitro* a été mise en évidence et deux génotypes contrastés ont été sélectionnés : K59, embryogène et K28, peu embryogène. A partir du croisement K28\*K59, deux descendance F1'a (n=123) et F1'b (n=104) ont été obtenues en 2001 et 2004, respectivement. La variabilité pour l'ES a été exploitée afin d'identifier les régions chromosomiques (QTL) et les gènes candidats impliqués dans ce processus.

### Résultats

Après 7 jours de culture des explants racinaires en condition d'ES et 30 jours de développement, les plantules (PL) issues du développement des embryons somatiques et les structures chlorophylliennes (SH) ont été comptabilisées. Les données phénotypiques ont montré une capacité embryogène plus élevée dans la descendance F1'a comparée à F1'b. Néanmoins, dans les 2 descendance, les caractères PL et SH présentent une distribution normale et continue après transformation en log10 et une héritabilité supérieure à 63%.

Une carte génétique issue du croisement K28\*K59 (F1'a) a été utilisée pour cartographier 47 gènes candidats identifiés auparavant par une expression différentiel au 4<sup>ème</sup> jour de culture en condition d'ES. A partir de cette carte, 124 marqueurs ont été sélectionnés (52 SSR, 33 STS et 39 AFLP) qui couvrent les 9 groupes de liaisons pour une analyse QTL sur les descendance F1'a et F1'b.

En considérant les deux caractères quantitatifs individuellement, nous avons identifié 6 QTL pour PL et SH, expliquant plus de 23% et 44% de variation phénotypique totale de ces caractères, respectivement. Parmi ces QTL, 2 sont spécifiques de la descendance F1'a et 1 de F1'b. Les QTL spécifiques de F1'a et F1'b pourraient être liés aux différences phénotypiques observées, mais aussi à la présence d'une région en distorsion sur le groupe de liaison 7 dans F1'b. Tous les QTL détectés pour PL et SH semblent présents sur les mêmes régions chromosomiques, ce qui suggère que la formation des plantules et les structures chlorophylliennes sont sous un contrôle génétique commun et que ces dernières représentent les embryons malformés.

Parmi les gènes candidats cartographiés, 16 co-localisent avec les QTL pour PL et SH. Vu leur implication dans la maintenance de l'état dédifférencié des cellules souches dans le méristème caulinaire, la co-localisation de gènes homologues à *SHOOT MERISTEMLESS* (*STM*) et *ARGONAUTE* (*AGO*) d'*Arabidopsis* avec les QTL6 et QTL2, respectivement, ce résultat est particulièrement intéressant.

### Conclusion

Avec la détection des QTL pour l'ES, les résultats de cette étude ont fourni pour la première fois des éléments sur le contrôle génétique de l'ES chez la chicorée. Ces résultats ouvrent la possibilité de l'introgression des QTL par sélection assistée par marqueurs dans les cultigrupes commerciaux de la chicorée généralement récalcitrants à l'ES. D'un point vu fondamental, les co-localisations des gènes candidats avec les QTL renforcent leurs implications putatives dans l'ES, et fournissent des critères additionnels dans la hiérarchisation des futures recherches.

**Mots clés :** chicorée (*Cichorium intybus* L.), embryogenèse somatique, contrôle génétique, analyse QTL, gènes candidats

## Abstract

### Introduction

Somatic embryogenesis (SE) is an asexual reproduction pathway in which somatic cells form embryos in a process that resembles zygotic embryogenesis. In chicory (*Cichorium intybus* L, 2n=18) a genetic variability in the capacity of somatic embryo formation was found, and two contrasting genotypes were selected: K59, embryogenic, and K28, hardly embryogenic. From the cross K28\*K59 two progenies, F1'a (n=123) and F1'b (n=104), were obtained in 2001 and 2004, respectively. Variability for somatic embryo formation in both progenies was exploited to identify chromosomal regions (QTL) and candidate genes implicated in this process.

### Results

After 7 days of culture of root explants under SE-inducing conditions and 30 days of development, the number of plantlets (PL) and shoot-like structures (SH) obtained were counted. Inspection of the phenotypic data showed that in contrast to the total number of events (PL+SH), the progenies F1'a and F1'b differed in the capacity of plantlet formation, F1'a clearly possessing a higher plantlet formation capacity than F1'b. Nevertheless, in both progenies the traits PL and SH showed continuous and normal distributions after log<sub>10</sub> transformation, and heritabilities superior to 63%.

A genetic map for the K28\*K59 progenies (F1'a) was used to map 47 candidate genes previously identified as being differentially expressed on the 4<sup>th</sup> day of culture under SE-inducing conditions. Using this map, 124 markers (52 SSR, 33 STS, and 39 AFLP) that covered the 9 linkage groups were selected for a QTL analysis in the progenies F1'a and F1'b.

For each of the two quantitative traits analyzed, 6 QTL were detected for both PL and SH that together explained more than 23% and 44% of the phenotypic variation for these traits, respectively. Amongst these QTL, 2 seemed specific for the progeny F1'a, and 1 for F1'b. These progeny-specific QTL may be related to the observed phenotypic differences, but also to a difference in the segregation of markers in a region of linkage group 7, that was distorted in progeny F1'b but not in F1'a. All QTL for PL and SH seemed to map to the same chromosomal regions, suggesting that the formation of plantlets and shoot-like structures is under common genetic control, and that the latter represent malformed embryos.

Amongst the mapped candidate genes, 16 co-localised with QTL for PL and SH. In view of their implication in the maintenance of the undifferentiated state of the stem cells in shoot apical meristems, the co-localisation of genes homologous to *SHOOT MERISTEMLESS* (*STM*) and *ARGONAUTE* (*AGO*) in *Arabidopsis* with QTL6 and QTL2, respectively, is a particularly interesting result.

### Conclusion

With the detection of QTL for SE, the results of this study have for the first time revealed elements in the genetic control of SE in chicory, and open the possibility of the introgression by marker assisted selection of QTL in commercial chicory cultigroups which are generally recalcitrant for SE. From a fundamental view point, the co-localisations of candidate genes with QTL enforce their putative implication in SE, and provide additional arguments in prioritizing future research.

**Key words:** chicory (*Cichorium intybus* L), somatic embryogenesis, genetic control, QTL analysis, candidate genes

