



Université Lille 1, Sciences et Technologies
UMR 1281 : Stress Abiotiques et Différenciation des Végétaux cultivés

THÈSE

pour l'obtention du grade de
Docteur en Sciences de l'Université de Lille 1
Stratégies d'Exploitation des Fonctions Biologiques

**Analyse génétique du métabolisme de l'acide
caftarique, chlorogénique et chicorique chez
Cichorium intybus L. :
cartographie des QTL et gènes candidats**

Par
Meriem BAHRI

Directeur de thèse : Pr. Theo Hendriks
Directeur du laboratoire : Pr. Jean-Louis Hilbert

Soutenue le 5 novembre 2010 devant la commission d'examen :

Laurent Legendre	Professeur, Université Jean Monnet, St-Etienne	Rapporteur
Didier Merdinoglu	Directeur de Recherches, INRA, Colmar	Rapporteur
Nadia Mulinacci	Professeur, Université de Florence, Italie	Examineur
Jean-Louis Hilbert	Professeur, Université Lille 1	Examineur
Theo Hendriks	Professeur, Université Lille 1	Examineur

Remerciements

Merci aux professeurs Jean-Louis Hilbert, directeur du laboratoire SADV, et Theo Hendriks, directeur de thèse, de m'avoir accueillie au sein du laboratoire et pour les conseils et corrections apportés au manuscrit. Merci à Marie-Christine Quillet, responsable de l'équipe cartographie et du matériel végétal utilisé.

Je remercie Didier Merdinoglu et Laurent Legendre d'avoir accepté d'être mes rapporteurs afin de juger de cette thèse.

Côté biochimie, je remercie Sébastien Grec, tuteur des premiers temps, Philippe Hance pour la création d'un programme HPLC sur mesure, et David Gagneul avec qui j'ai pu longuement discuter de mes travaux malgré la fraîcheur de son arrivée au sein du laboratoire.

Côté génétique, mille mercis à Monica Mörchen pour son travail concernant les gènes candidats et merci de m'avoir initié au génotypage et d'avoir répondu à mes interrogations face aux profils « bizarres ». Modèle en termes d'écologie et j'espère être sur la bonne voie pour te ressembler dans quelques années... certes, il y a encore du boulot !

J'exprime également toute ma gratitude à Thierry Cadalen avec qui j'ai pu discuter, trouver des réponses ou des idées quant à la réalisation de mon travail côté carte génétique et QTL. Merci pour ta gentillesse, ta patience et les relectures de mon manuscrit.

Côté végétal, merci à Nicolas Henri, Paul Dhallewyn et bien sûr Bruno Desprez, de m'avoir trouvé beaucoup de place pour stocker mes plantes sur le site de Florimond-Desprez. Merci d'ailleurs à Theo pour les nombreux déplacements en voiture avant que je n'obtienne le permis! A l'université de Lille1, merci beaucoup à l'équipe « serre » : Angélique Bourceaux, Cédric Glorieux, Eric Schmitt et Nathalie Faure. Sans eux, que seraient devenues mes chicorées ?

Pour leur sympathie et bonne humeur, en ou hors paillasse, je remercie vivement Christelle Blassiau, Aurélie Bluteau, Nasser Bahrman, Brigitte Huss, Najia Voedts, Marie-Christine Dendeviel, Séverine Six et le professeur Francis Trottin.

Côté enseignements, je tiens à remercier mes « collègues » de T.P. dont j'ai pu attentivement observer les méthodes pédagogiques mais aussi constater leur sympathie: Estelle Goulas, Arnaud Day, Christophe Vuylstekker, Anne-Sophie Blervacq et Marie-Hélène Boucher.

Un grand clin d'œil à tous les « passagers » de courte ou longue durée de la « salle des diplomitifs ». Parmi eux: Lucy, Estelle, Nadim, Aline, Giovanna, Ildéphonse, Sylvain, Matthieu. Tour à tour sérieuse, délirante ou dramatique, l'ambiance du bureau 305 est sûrement ce qui me manquera le plus... même si quelques mois avant la fin, j'ai changé de bureau pour être plus au calme (j'ai quand même emporté Lucy avec moi!). Je n'oublie pas non plus les GEPViens avec qui j'ai bien rigolé tous les midis mais aussi certains soirs: Isa, J-B, Camillo & Camilla, Aude, Fabrice, Laurent, Martin, Romain, Marilyne...

Merci à la chicorée d'être ce qu'elle est! Une plante souvent (pré-)jugée comme « vieillot » mais finalement très intéressante et multi-facettes, bref, pleine de surprises. Par contre, je remercie beaucoup moins chaleureusement les marqueurs moléculaires et l'HPLC qui m'ont bien cassé la tête par moment.

Merci à Annie & Philippe pour leur présence « familiale » pendant ces années de thèse et Maryse et Hugh chez qui les vacances étaient toujours trop courtes. Merci à mon grand bol d'air hebdomadaire à plumes et paillettes, Atabak, ainsi qu'à mes machines à coudre Heidi et Bertha qui m'ont régulièrement changé les idées! Merci aussi aux personnages de Scrubs qui me garantissaient 15 minutes de rire par jour pendant la lourde période de rédaction.

J'exprime ma profonde reconnaissance à ma famille, qui m'a régulièrement aidé lors de la dernière année sans financement. Ma mère, mon père, mon frère, ma grand-mère et mon grand-père. Tous m'ont épaulée d'une façon ou d'une autre... j'ai une famille formidable, que dire de plus? Malheureusement, mes allers-retours en Tunisie n'ont pas été aussi nombreux que je l'aurai souhaité mais chacun d'entre eux était une très chouette bouffée d'oxygène (et de couscous). Deux jours après la fuite du dictateur, toutes mes pensées vont également au courageux peuple tunisien... en espérant que le calme revienne bientôt.

Enfin, je remercie tout particulièrement, le beau, l'intelligent, le gentil, le merveilleux, bref le plus-que-parfait Benjamin, pour son soutien sans faille, pour sa patience infinie et son aide incommensurable dans bien des domaines. Il a même bien voulu devenir mon mari entre temps, c'est pour dire! Et vive la prochaine aventure... à Chicago!

Villeneuve d'Ascq, le 16 janvier 2011

Sommaire

INTRODUCTION BIBLIOGRAPHIQUE	1
I. La chicorée	3
I.1. Classification botanique et cycle de développement	3
I.2. Utilisation de la chicorée, une histoire ancienne.....	3
I.3. Utilisation agro-alimentaire de nos jours	4
II. Polyphénols : caractéristiques et intérêts.	6
II.1. Les différentes classes de polyphénols	6
II.1.A. Lignanes et stilbènes	6
II.1.B. Flavonoïdes	7
II.1.C. Acides phénoliques	7
II.2. Rôle en tant que métabolites secondaires	8
II.2.A. Exemple d'accumulation suite à un stress biotique	9
II.2.B. Exemple d'accumulation suite à un stress abiotique.....	9
II.3. Rôle antioxydant des polyphénols	10
II.4. Rôle préventif supposé des polyphénols chez l'Homme	11
II.5. Polyphénols et chicorée	13
III. Biosynthèse des polyphénols et gènes candidats.....	17
III.1. Biosynthèse de l'acide caféique : enzymes impliquées.....	17
III.2. Biosynthèse des acides chlorogéniques : enzymes impliquées	18
III.3. Biosynthèse des acides chicorique et caftarique : enzymes impliquées.....	21
III.4. Facteurs de transcription régulant la biosynthèse des polyphénols.....	24
IV. Architecture génétique de caractères quantitatifs.....	26
IV.1. Origine de l'analyse QTL.....	26
IV.2. Outils d'une analyse QTL	27
IV.2.A. Une descendance en ségrégation	27
IV.2.B. Les marqueurs moléculaires et la construction d'une carte génétique	28
IV.2.C. Le phénotypage.....	32
IV.2.D. Cartographie de QTL.....	32
IV.3. En complément de l'analyse QTL : cartographie de gènes candidats.....	36
V. Contexte de la thèse de doctorat	38

V.1. Partenariat industriel.....	38
V.2. Contexte scientifique et objectifs	38
CHAPITRE I: Mise en place d'une méthode d'extraction des hydroxycinnamates et collection des données phénotypiques	41
I. Historique de l'utilisation de la chicorée au sein du laboratoire SADV	43
II. Matériels et méthodes	44
II.1. Obtention de la population à phénotyper	44
II.2. Obtention du matériel végétal.....	44
II.2.A. Obtention des plantes mères (octobre 2006).....	44
II.2.B. Obtention des clones	46
II.3. Constitution des blocs de clones	47
II.4. Protocole final : extraction et analyses des cinq blocs (954 chicorées).....	48
II.4.A. Prélèvement et extraction.....	48
II.4.B. Analyse des acides hydroxycinnamiques par CLHP	49
II.4.C. Analyse par CL/MS.....	50
II.4.D. Analyse de la capacité antiradicalaire globale	51
II.5. Traitement statistique.....	52
III. Résultats	53
III.1. Mise en place d'un protocole pour le phénotypage.....	53
III.1.A. Préambule au protocole final.....	53
III.1.B. Mise en place du protocole final.....	56
III.1.C. Adaptation des outils de phénotypages.....	58
III.1.D. Identification des molécules d'intérêt.....	59
III.2. Essais préliminaires sur les parents et la descendance	60
III.3. Phénotypage pour la détection de QTL.....	63
III.3.A. Données CLHP	63
III.3.B. Données activité antiradicalaire (AAR) et analyses statistiques	70
III.3.C. Corrélation entre les différentes données.....	71
IV. Discussion	72
IV.1. Protocole final	72
IV.2. Identification des molécules majoritaires.....	76
IV.3. Phénotypage pour la détection de QTL.....	78
IV.3.A. Obtention des données CLHP	78
IV.3.B. Essais analyses statistiques.....	79

IV.3.C. Résultats des analyses statistiques.....	79
IV.4. Héritabilité au sens large et effet de l'environnement.....	84
V. Conclusion.....	85
CHAPITRE II: Mise en place d'une carte génétique pour la détection de QTL et cartographie de gènes candidats.....	87
I. Historique de la carte génétique de la chicorée (2n=2x=18) au laboratoire SADV..	89
II. Matériels et méthodes, carte « oxychic ».....	92
II.1. L'extraction d'ADN.....	92
II.2. Les marqueurs moléculaires	93
II.2.A. Les marqueurs SSR.....	93
II.2.B. Les marqueurs SSCP (EST)	95
II.2.C. Les marqueurs AFLP	98
II.3. Construction de la carte génétique.....	100
II.4. Distorsion de ségrégation : double test de χ^2	100
III. Matériels et méthodes, gènes candidats.....	101
III.1. Recherche des EST <i>Cichorium</i> à partir de gènes candidats identifiés chez d'autres espèces (en collaboration avec Monika Mörchen).	101
III.2. Définition des amorces	107
III.3. Recherche de polymorphisme	107
III.3.A. Polymorphisme SSCP	107
III.3.B. Polymorphisme HRM.....	108
IV. Résultats	109
IV.1. Recherche de gènes candidats	109
IV.2. Construction de la carte « oxychic »	112
IV.3. Ordre des marqueurs	117
IV.4. Distorsion	118
V. Discussion.....	122
V.1. Région riche/pauvre en marqueurs moléculaires.....	123
V.2. Cas particulier du groupe de liaison 9	123
V.3. Distorsion de ségrégation	124
V.4. Gènes candidats	126
VI. Conclusion.....	127
CHAPITRE III: Détection des QTL impliqués dans l'activité antiradicalaire et le métabolisme des hydroxycinnamates	129

I. Matériels et méthodes	131
I.1. Données utilisées pour la détection de QTL	131
I.2. Tests statistiques pour la détection de QTL	132
I.3. Identification de QTL.....	133
I.4. Calcul des interactions entre marqueurs	134
II. Résultats.....	134
II.1. Détections de QTL.....	134
II.1.A. Test de Kruskal-Wallis (K-W).....	134
II.1.B. Test de permutation.....	136
II.1.C. Interval-Mapping (IM)	137
II.1.D. Multiple QTL Model (MQM).....	140
II.1.E. Effets alléliques	158
II.1.F. Interactions entre marqueurs	161
III. Discussion.....	164
III.1. Méthode de détection et robustesse des QTL.....	164
III.2. Co-localisations	167
III.2.A. Co-localisation des différents QTL des caractères étudiés.....	167
III.2.B. Co-localisation des QTL avec des gènes candidats	170
III.2.C. Co-localisation avec des QTL du poids secs résiduels	175
III.2.D. Travaux parallèles à partir du même croisement.....	175
III.3. QTL à effet interactif.....	176
IV. Conclusion.....	177
CONCLUSION ET PERSPECTIVES	181
REFERENCES BIBLIOGRAPHIQUES	189
ANNEXES	205

INTRODUCTION
BIBLIOGRAPHIQUE

I. La chicorée

I.1. Classification botanique et cycle de développement

La chicorée est une plante herbacée, dicotylédone, de la famille des Astéracées, comme la laitue, le tournesol ou encore l'artichaut. Le genre *Cichorium* regroupe six espèces retrouvées de nos jours sur les cinq continents (Kiers, 2000).

Trois d'entre elles sont présentes en Europe : *Cichorium spinosum* L., *Cichorium endivia* L. et *Cichorium intybus* L. Cette dernière est de type vivace lorsqu'elle pousse dans un environnement sauvage (dans les prés ou le long des routes). Pour le domaine agro-alimentaire, elle est souvent cultivée sous forme bisannuelle. Comme le montre la figure 1, à la fin de la première année, une rosette de feuilles se forme puis une racine tubérisée accumule des réserves azotées et carbonées (Limami *et al.*, 1993). Ces réserves sont remobilisées au printemps suivant, lors de la montaison, pour la formation d'une hampe florale (tige ramifiée florifère). La floraison a lieu entre juillet et septembre. Quinze à vingt fleurs ligulées sont regroupées en une inflorescence bleue/violette appelée capitule qui ne s'épanouit que quelques heures lors d'un seul jour. *C. endivia*, *C. spinosum* et *C. intybus* présentent un mode de pollinisation entomophile. La première est préférentiellement autogame, alors que les deux suivantes sont allogames, l'allofécondation étant favorisée, par la présence d'un système d'auto-incompatibilité sporophytique (Eenink, 1981), des phénomènes de compétition pollinique (Eenink, 1982) et la morphologie florale défavorable à l'autofécondation (Desprez *et al.*, 1994).

I.2. Utilisation de la chicorée, une histoire ancienne

La chicorée est depuis longtemps utilisée comme plante médicinale. Le document le plus ancien la mentionnant est le papyrus Ebers, vieux de 6000 ans. Il nous apprend que les égyptiens la consommaient en apéritif pour son effet digestif. Plus tard, les grecs et les romains voyaient en elle une alliée pour guérir les maladies du foie ou de l'estomac. Chez les arabes, elle était employée comme remède ophtalmologique. Dans l'iconographie chrétienne du Moyen-âge, elle symbolisait la persévérance dans le martyre. Au cours des siècles qui suivirent, elle fut préconisée dans beaucoup de cas : par exemple, contre la gale, la jaunisse, la

fièvre, les dermatoses, les rhumatismes. Même Louis XIV soigna ses nausées à l'aide d'un remède à base de sirop de chicorée.

Les hollandais semblent être les premiers à mettre au point la torréfaction de la chicorée de façon industrielle (XVII^{ème} siècle). Ce procédé gagna ensuite l'Allemagne, la Belgique et la France. Son utilisation est d'autant plus importante sous la Révolution française puis lors du Blocus Continental décrété par Napoléon, du fait de ses avantages économiques (culture locale) par rapport au café (importé). A partir de cette période, son usage alimentaire prit le pas sur son utilisation médicinale (Kandeler & Ullrich, 2009).

I.3. Utilisation agro-alimentaire de nos jours

Parmi les trois espèces de *Cichorium* retrouvées en Europe, l'espèce *C. spinosum* pousse essentiellement en climat méditerranéen ; elle est peu cultivée étant donné son faible intérêt agronomique. L'espèce *C. endivia* est cultivée de façon annuelle ; ses feuilles sont consommées en salade sous le nom de scarole et chicorée frisée.

L'espèce *C. intybus*, à vocation agro-alimentaire, peut se décomposer en plusieurs variétés : *foliosum*, *sativum* et *sylvestre*.

La variété *sylvestre* réunit un nombre important de chicorées « salade ». « Chioggia » et « Grumolo » sont disponibles en feuilles rouges ou vertes. « Verona » est une chicorée rouge. « Castelfranco » et « Treviso » présentent des feuilles panachées vertes et rouges ou blanches et rouges, respectivement.

La variété *foliosum* réunit la chicorée « Spadona », la « Catalogna » appréciée après cuisson ou encore, la chicorée « Witloof », appréciée aussi bien fraîche que cuite. Cette dernière est cultivée en deux ans. Les graines sont semées en mars-avril et les plantes récoltées en septembre. Les feuilles sont alors coupées au niveau du collet et les racines sont soumises au forçage (figure 1). Dans l'obscurité et avec une température avoisinant les 16-18°C sous humidité saturante, l'activité végétative reprend et génère des bourgeons étiolés au bout de trois semaines. Ces chicorées aux feuilles blanches sont commercialisées ensuite sous le nom (impropre!) d'endives, l'appellation chicon étant plus spécifique au Nord de la France et à la Belgique francophone.

A contrario, la variété *sativum* voit son cycle naturel arrêté au bout d'une année (figure 1). Elle est aussi appelée chicorée « Magdebourg » ou « Industrielle ». Elles ont été sélectionnées pour la grosseur de leur racine dont l'inuline (polyfructosanes) (Meyer &

Stasse-Wolthuis, 2009) ou le fructose sont extraits (Mutanda *et al.*, 2009). Une farine de racines de chicorées est également commercialisée comme améliorant de panification, apportant du moelleux aux produits cuits à base de céréales. La torréfaction des racines permet la production d'une boisson nommée « chicorée », consommée en complément ou succédané de café. Vendue en granulés, en poudre ou liquide, elle est aussi utilisée comme colorant, arôme et exhausteur de goût.

En France, c'est le Sud-Est qui représente la majorité de la production nationale en chicorée salade, c'est-à-dire scarole et frisée (données 2008, Ministère de l'alimentation, de l'agriculture et de la pêche). Par contre, le Nord de la France, avec le groupe Leroux, détient le monopôle mondial de la production de chicorée industrielle à torréfier. La production de chicorée industrielle est encore plus importante en Belgique. Outre ces deux pays, les gros producteurs de chicorées sont les Pays-Bas et l'Inde.

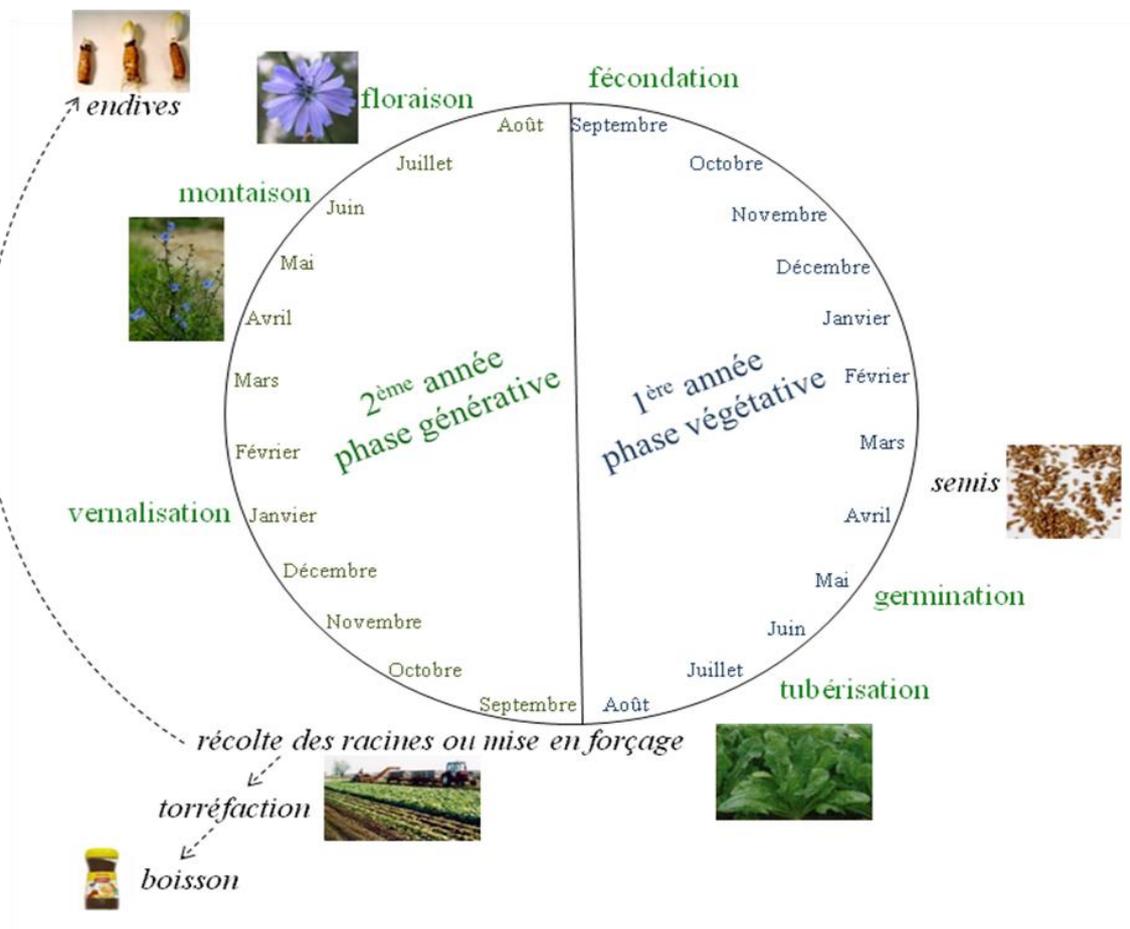


Figure 1. Le cycle biologique (en vert) et agronomique (en noir et en italique) de *Cichorium intybus* L.

En tant que plante d'intérêt agro-alimentaire, la chicorée est caractérisée par son amertume. A travers les siècles, elle a été utilisée comme plante médicinale. Ces deux aspects sont liés à son contenu biochimique et plus exactement aux molécules issues du métabolisme secondaire.

II. Polyphénols : caractéristiques et intérêts.

Le métabolisme primaire réunit les éléments essentiels à la vie de la plante (carbohydrates, protéines, acides et esters gras, acides nucléiques). En opposition à ce métabolisme commun à tout le règne végétal, le métabolisme secondaire symbolise la diversité des molécules retrouvées chez les plantes et regroupe un panel de molécules permettant à celles-ci de réagir dans un environnement donné (répulsion des herbivores, protection contre les UV, attraction des pollinisateurs, etc.). Parmi ces molécules issues du métabolisme secondaire, on retrouve les polyphénols qui constituent un groupe de plusieurs milliers de molécules caractérisées par des groupements hydroxyles (OH) associés à un ou plusieurs cycles aromatiques. En fonction du nombre et de l'agencement des cycles aromatiques, on peut distinguer quatre grands groupes (figure 2): les lignanes, les stilbènes, les flavonoïdes et les acides phénoliques. Les lignanes et certains flavonoïdes sont compartimentés dans la paroi cellulaire tandis que les autres composés phénoliques sont, en majorité, localisés dans les vacuoles (Yamaki, 1984). La polarité de ces molécules augmente en fonction de la position du groupe hydroxyle sur le cycle aromatique : la présence d'un -OH en position 4 confère plus de polarité qu'en position 3 et en position 2. Les groupes méthoxy (-O-CH₃) ou acryloyl $(-(O=)C-CH=CH_2)$ peuvent réduire davantage la polarité.

II.1. Les différentes classes de polyphénols

II.1.A. Lignanes et stilbènes

Du point de vue structure chimique, les stilbènes sont caractérisés par un squelette C₆-C₂-C₆ et les lignanes par un squelette (C₆-C₃)₂. Ils ne sont présents qu'en petite quantité dans notre alimentation, à l'exception des graines de lin qui représentent la source majeure de lignanes (Adlercreutz & Mazur, 1997; Vitrac *et al.*, 2002).

II.1.B. Flavonoïdes

Les flavonoïdes sont constitués de deux noyaux aromatiques reliés entre eux par trois atomes de carbone formant un hétérocycle oxygéné (squelette C₆-C₃-C₆). Selon le degré d'oxygénation et d'insaturation du segment C₃, six classes de flavonoïdes sont définies : flavonols, flavones, isoflavones, flavanones, flavanols et anthocyanes (Manach *et al.*, 2004).

Particulièrement accumulés sous l'action d'un stimulus lumineux, les flavanols sont retrouvés dans les tissus périphériques et aériens des plantes où ils jouent un rôle protecteur contre les ultra-violet (Price *et al.*, 1995). Ce sont les flavonoïdes les plus abondants dans notre alimentation (oignons, brocoli, myrtille, etc.) avec, en tête, la quercétine et le kaempférol. Par contre, les flavones sont moins courants : nous ingérons essentiellement des glycosides de lutéoline et d'apigénine, la source majeure étant le persil et le céleri. Les flavanones, quant à elles, existent en grande quantité dans les agrumes (hespéridine pour l'orange, ériodictyol pour le citron, naringénine pour le pamplemousse). Elles sont généralement glycosylées par un disaccharide pouvant procurer un goût amer (néohespéridose donnant la naringinine chez le pamplemousse) ou pas (rutinose). Les molécules composant le groupe des isoflavones ont une structure qui leur confère des similarités avec les œstrogènes et donc des propriétés pseudohormonales. Ces phyto-œstrogènes sont contenus quasiment exclusivement dans les légumineuses. Les flavanols regroupent des composés que l'on peut aussi diviser en deux classes : sous forme monomérique (catéchines : abricot, thé vert, chocolat) ou sous forme polymérisée (proanthocyanidines ou tannins condensés : dimères, oligomères ou polymères de catéchines) (Manach *et al.*, 2004). Enfin, les anthocyanes sont des pigments vacuolaires colorant les tissus épidermiques des fleurs et des fruits en rose, rouge, bleu ou violet (Mazza & Miniati, 1993). Ils sont stabilisés en se complexant avec d'autres flavonoïdes. Retrouvée par exemple dans le vin rouge, les aubergines ou encore l'oignon rouge, la cyanidine est l'anthocyane que l'on consomme le plus.

II.1.C. Acides phénoliques

Le groupe des acides phénoliques peut se départager en deux sous-groupes : les acides hydrobenzoïques sont des dérivés de l'acide benzoïque tandis que les acides hydroxycinnamiques proviennent de l'acide cinnamique. Les acides hydrobenzoïques sont retrouvés dans certains fruits rouges, les oignons, les radis ou le thé (acide gallique). Ils

peuvent rentrer dans la composition de structures complexes comme les tannins hydrolysables (Clifford & Scalbert, 2000). Cependant, du fait de leur faible répartition dans notre régime alimentaire, ils restent peu étudiés.

Les acides hydroxycinnamiques sont retrouvés plus couramment que les précédents. Les principaux acides hydroxycinnamiques sont les acides *p*-coumarique, caféique, férulique et sinapique. Ils sont rarement retrouvés sous forme libre (squelette C₆-C₃) mais sont sous forme glycosylée ou estérifiée avec l'acide quinique, tartarique ou shikimique. Ainsi, l'acide caféique, avec l'acide quinique forme l'acide chlorogénique ; avec l'acide tartarique, l'acide caféique forme l'acide caftarique ou l'acide chicorique (une ou deux molécules d'acide caféique, respectivement). Ces formes estérifiées sont solubles et retrouvées dans la vacuole tandis que d'autres hydroxycinnamates sont insolubles et liés aux composants de la paroi cellulaire comme la lignine ou la cellulose (Shahidi & Chandrasekara, 2010). L'acide caféique et ses dérivés représentent généralement la plus grande partie, voire la totalité des acides hydroxycinnamiques contenus dans les fruits. Par contre, c'est l'acide férulique qui est le plus abondant dans les graines de céréales (Sosulski *et al.*, 1982).

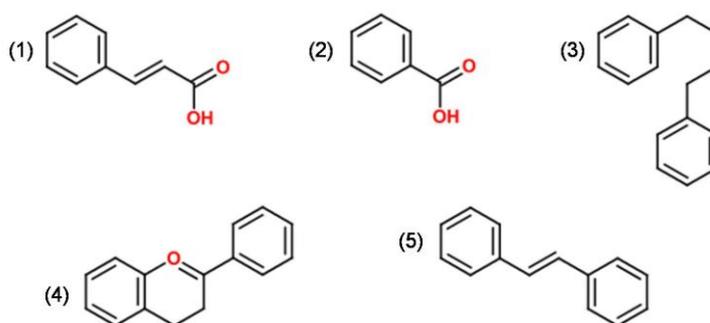


Figure 2. Structure de base des différents types de polyphénols : (1) acide hydroxycinnamique, (2) acide hydroxybenzoïque, (3) lignine, (4) flavonoïde, (5) stilbène.

II.2. Rôle en tant que métabolites secondaires

Chaque plante possède un niveau de base en polyphénols lui permettant de vivre dans son environnement. Par exemple, les composés colorés comme les anthocyanes pigmentent les pétales des fleurs pour attirer les pollinisateurs (Holton & Cornish, 1995). Cependant, il a été remarqué que l'accumulation de certains composés phénoliques est corrélée à des stress biotiques et abiotiques, suggérant leur participation dans les processus de tolérance et/ou de défense.

II.2.A. Exemple d'accumulation suite à un stress biotique

De nombreux polyphénols sont classés parmi les phytoaléxines, à savoir des molécules antimicrobiennes de faible poids moléculaire synthétisées par les plantes après une attaque par des pathogènes (Paxton, 1981). La teneur en phytoaléxines augmente de façon importante autour du site d'infection pour atteindre des concentrations toxiques pour le pathogène. Cette catégorie de composés regroupe plusieurs polyphénols dont certains flavonols comme la quercétine ou le kaempférol, des stilbènes et des isoflavones (Dixon, 2005). Les plantes naturellement résistantes aux pathogènes produiraient plus de composés phénoliques (Sztejnberg *et al.*, 1983) résultant de l'activité accrue d'enzymes catalysant les réactions de la voie métabolique correspondante. Cette hypothèse a été soutenue par les résultats de Darvill et Albersheim (Darvill & Albersheim, 1984) montrant une baisse de la résistance aux pathogènes lorsque l'enzyme clé du métabolisme des polyphénols est inhibée (PAL, Phénylalanine Ammonia Lyase). De manière similaire, l'astringence des tannins repousserait les herbivores. D'autres stress biotiques sont également la cause d'une augmentation en composés phénoliques : on remarque par exemple que les quantités d'acide chlorogénique - présent comme composés de défense ou précurseurs de la synthèse de lignines (Solecka, 1997) - augmentent considérablement suite à une blessure (McClure, 1960).

II.2.B. Exemple d'accumulation suite à un stress abiotique

Les ultra-violets, notamment les UV-B, stimulent la synthèse d'anthocyanes, de dérivés du kaempférol ou encore d'esters d'acides phénoliques. Ainsi, ces molécules protégeraient la plante des radiations UV pouvant atteindre le mésophile et affecter la photosynthèse (Stapleton, 1992; Chalker-Scott, 1999) mais aussi directement casser les brins d'ADN (Chalker-Scott, 1999).

Une carence en azote, phosphate et fer peut induire la synthèse de flavonoïdes, d'anthocyanes ou d'acides phénoliques, respectivement (Marschner, 1991; Wojtaszek *et al.*, 1993).

Une synthèse accrue d'acides phénoliques et d'anthocyanes est observée après une baisse de température, résultant d'une activité intense de l'enzyme PAL (Christie *et al.*, 1994). De même, une augmentation en lignine et subérine protégerait les végétaux des dommages

causés par le froid (Griffith *et al.*, 1985; Chalker-Scott & Fuchigami, 1989), (Griffith *et al.*, 1985).

II.3. Rôle antioxydant des polyphénols

Les polyphénols d'origine végétales consommés par l'Homme permettent de compléter l'action d'antioxydants endogènes comme les enzymes superoxyde dismutase, catalase, glutathione peroxidase et réductase, mais aussi d'économiser ou de régénérer d'autres molécules antioxydantes. Les composés phénoliques possèdent donc une activité antioxydante reconnue et d'ores et déjà utilisée dans l'agro-alimentaire ou la parapharmacie.

Leur capacité à céder un électron, permet de lutter contre les espèces actives de l'oxygène, les espèces réactives de l'azote (tableau 1) et de rétablir le potentiel antioxydant des vitamines A, E et C (Tabet *et al.*, 2002). L'apport d'hydrogène par une molécule antioxydante (AH) à un radical libre (R[•]) peut être résumé par cette équation :

$R^{\bullet} + AH \rightarrow RH + A^{\bullet}$. Dans ce cas, l'efficacité d'un antioxydant dépend du nombre de groupes hydroxyles disponibles et de leur position sur la molécule, de la solidité de la liaison A-H (plus la liaison est faible, plus le pouvoir antiradicalaire est fort) et ensuite de la capacité du radical phenoxyyl formé (A[•]) à rester stable pour ne pas initier d'autres radicaux libres (Robards *et al.*, 1999).

On attribue aussi aux antioxydants la capacité de piéger les métaux ioniques comme le fer, le cuivre, le zinc, et le magnésium (Wijeratne *et al.*, 2005). Ces métaux sont responsables de la production de radicaux hydroxyles très réactifs, aboutissant à des réactions d'oxydation en chaîne (réaction de Fenton ou d'Haber-Weiss, (Halliwell, 1978), (McCord & Day 1978)). L'exemple de la réaction initiatrice de l'oxydation du fer donnerait ainsi : $H_2O_2 + Fe^{2+} \rightarrow HO^{\bullet} + HO^- + Fe^{3+}$ (Benov, 2001). La chélation de ces métaux par les composés polyphénoliques se ferait alors par une complexion structure-dépendante entre les deux molécules (Van Acker *et al.*, 1996).

Espèces réactives de l'oxygène :	Espèces réactives de l'azote :
anion superoxide O_2^-	péroxydite $ONOO^-$
oxygène singulet 1O_2	monoxyde d'azote NO^\bullet
ion hypochlorite OCl^-	dioxyde d'azote NO_2
péroxyde d'hydrogène H_2O_2	trioxyde diazoté N_2O_3
ion hydroxyle OH^-	anion nitroxy NO^-
radical hydroxyle OH^\bullet	radical nitrosyl $ONOO^\bullet$

Tableau 1. Les espèces réactives de l'oxygène et de l'azote (Turrens, 2003), (Massion *et al.*, 2002).

Toutefois, à certaines concentrations, les polyphénols sont capables de s'auto-oxyder ou d'être oxydés, et ainsi initier d'autres réactions pro-oxydantes. Ce phénomène peut avoir lieu quand les composés phénoliques deviennent les substrats d'enzymes comme les polyphenol oxidases (PPO) responsables du brunissement des fruits. Les PPO oxyderaient les composés phénoliques en quinones qui, après polymérisation, formeraient des pigments bruns (Nicolas *et al.*, 1994). Selon les expériences de Kader *et al.* (Kader *et al.*, 1997), l'acide caféique, suivi de l'acide chlorogénique, semblent être d'excellents substrats pour les PPO.

II.4. Rôle préventif supposé des polyphénols chez l'Homme

Par son alimentation, l'Homme peut ingérer jusqu'à 1 gramme de polyphénols par jour, soit environ dix fois la quantité de vitamine C ou cent fois la quantité de vitamine E et de caroténoïdes, autres molécules antioxydantes (Manach *et al.*, 2004), (Scalbert & Williamson, 2000). Cet apport peut provenir de sources variées, essentiellement des fruits et légumes, mais aussi de produits dérivés des végétaux : thé, café, vin, céréales et chocolat. Concernant les acides hydroxycinnamiques, l'estimation haute peut atteindre jusqu'à 800 mg par jour (Clifford, 2000), (Radtke *et al.*, 1998), cette variation étant principalement due à la consommation en café, riche en acide chlorogénique.

Etant donné les propriétés antioxydantes des polyphénols et leur abondance dans notre régime alimentaire, le milieu des années 90 voit naître de nouveaux sujets de recherche concernant ces molécules jusqu'ici supplantées par d'autres composés antioxydants : les vitamines, les minéraux et les caroténoïdes (Scalbert *et al.*, 2005).

Ces recherches vont alors être dirigées vers le rôle préventif et supposé des polyphénols sur les maladies associées au stress oxydatif comme les cancers, les diabètes, ou encore les maladies cardiovasculaires et neurodégénératives, l'arthrose, la cataracte ou

l'ostéoporose. Effectivement, ces maladies sont liées à l'accumulation de dégradations oxydatives que peuvent subir l'ADN, les protéines ou les lipides d'une cellule au cours du temps. Ces dommages sont généralement associés aux espèces réactives de l'oxygène ou de l'azote (cf. Tableau 1). Scalbert *et al.* résument les rôles hypothétiques ou confirmés des polyphénols sur différentes maladies dégénératives dans l'article de synthèse « Dietary polyphenols and disease prevention », paru en 2005. Ainsi, dans le cas de l'athérosclérose (perte d'élasticité des artères due à la sclérose provoquée par l'accumulation de corps gras, notamment des LDL [Low Density Lipoprotein] associés au cholestérol dont il est le transporteur dans le plasma), la consommation de pro-anthocyanidines chez le lapin peut réduire la densité des LDL absorbés par les macrophages (indicateur du nombre de LDL oxydés) et la formation d'agrégats de LDL (Yamakoshi *et al.*, 1999). Une diminution de l'oxydation des LDL et de l' α -tocophérol (antioxydant contenu dans les LDL) grâce aux catéchines du thé vert a été constaté *in vitro* (Zhu *et al.*, 1999). Chez le rat, la consommation de vin rouge est corrélée à une inhibition de l'agrégation des plaquettes, suggérant un effet préventif des polyphénols en cas de phlébite (Russo *et al.*, 2001). Ce type d'étude suggère une réponse à la question du « paradoxe français », soit l'association paradoxale entre un régime alimentaire riche en cholestérol et acides gras saturés, et un taux de maladies coronariennes faible.

Parallèlement, plusieurs publications recensent l'effet anticancéreux des polyphénols : les rats ou souris qui ont ingéré ces molécules antioxydantes avant l'administration d'un agent carcinogène ont tendance à développer moins de tumeurs ou à ralentir la croissance de celles-ci (Yang *et al.*, 2001). Différentes hypothèses pour expliquer ces effets ont été suggérées, parmi lesquelles la modulation de l'expression des cytochromes P450, impliqués dans le métabolisme pro-carcinogène, par les polyphénols (Johnson *et al.*, 1994). Ils pourraient également former des molécules toxiques, stimulant alors le système de détoxification (Talalay *et al.*, 1988) tandis que d'autres publications discutent de leur capacité à inhiber la prolifération *in vitro* (Kuntz *et al.*, 1999) mais aussi *in vivo* (Ishikawa & Kitamura, 2000) en induisant l'apoptose de cellules tumorales.

Les maladies neurodégénératives, également résultantes du stress oxydatif, interpellent la communauté scientifique étant donné que 1% et 15% des plus de 65 ans sont atteints de la maladie de Parkinson et d'Alzheimer, respectivement (Cantuti-Castelvetri *et al.*, 2000). Dans ce contexte, un régime alimentaire riche en polyphénols chez le rat a démontré une amélioration des fonctions cognitives et de la transduction du signal neuronal (Joseph *et al.*, 1999). Il existe néanmoins peu d'informations concernant la proportion de polyphénols, ou de

leurs métabolites, capables de traverser la barrière hémato-encéphalique et d'atteindre le cerveau (Singh *et al.*, 2008).

De même, la recherche sur le diabète s'est penchée sur le cas des polyphénols. En effet, quelques acides phénoliques ont été reconnus comme ayant un rôle anti-hyperglycémique : l'acide caféique, l'acide cinnamique (Huang *et al.*, 2009), l'acide férulique (Ohnishi *et al.*, 2004) et l'acide chlorogénique (van Dam & Feskens, 2002), entre autres.

Cependant, la dualité des pouvoirs antioxydant et oxydant des polyphénols énoncés dans le paragraphe II.2 a également interpellé le milieu médical. Quelques études ont démontré que d'importantes quantités de composés phénoliques pouvaient avoir des effets pro-oxydants. Ainsi, en réduisant le fer (III) en fer (II), les polyphénols généreraient des radicaux hydroxyles (réaction de Fenton). Les isoflavones, appelés aussi phyto-œstrogènes, agiraient en tant qu'agonistes des œstrogènes et pourraient alors augmenter les risques de certains cancers (Ju *et al.*, 2001).

Néanmoins et de manière générale, beaucoup d'études *in vitro* présentent des résultats ne tenant pas compte de la bio-efficacité de ces molécules une fois ingérées. Celle-ci dépendrait de la structure chimique (par exemple estérification, polymérisation, glycosylation) dans laquelle les molécules se trouvent, de la vitesse d'absorption, des transformations engendrées par la flore bactérienne et de la vitesse d'élimination (Manach *et al.*, 2005). Par exemple, des études ont montré que l'absorption de l'acide caféique (absorption essentiellement au niveau du colon) serait moins efficace sous forme estérifiée (Olthof *et al.*, 2001). Ainsi, l'acide chlorogénique peut être absorbé directement au niveau de l'intestin grêle mais sera clivé au niveau du colon par des estérases pour donner de l'acide caféique (Niggeweg *et al.*, 2004).

II.5. Polyphénols et chicorée

Les polyphénols ont été étudiés et identifiés dans les variétés de chicorées « salades », leur activité antioxydante étant une valeur ajoutée d'un point de vue commercial. Le tableau 2 récapitule ces données. Les publications citent également de nombreuses molécules sous le nom de dérivés d'acide caféique ou dérivés d'esters caféoyl, demandant une identification ultérieure plus précise. Seuls Rees et Harbone (1985) ont recensé les molécules présentes dans les feuilles de la chicorée Magdebourg, cultivée pour ses racines. L'étude de trois parties de la feuille (apex, milieu, base) pendant plusieurs mois consécutifs, indique une variation

(présence/absence) de certaines molécules selon leur localisation tissulaire et la saison. De manière globale, on remarque l'importance des acides hydroxycinnamiques, et notamment des acides chicorique et chlorogénique, dans les différentes variétés étudiées, quelque soit leur pigmentation. Les deux acides phénoliques précédemment cités sont les molécules retrouvées en plus grande quantité dans les extraits, ou partageant la première place avec les cyanidines quand il s'agit des chicorées rouges (cf. références du tableau 2).

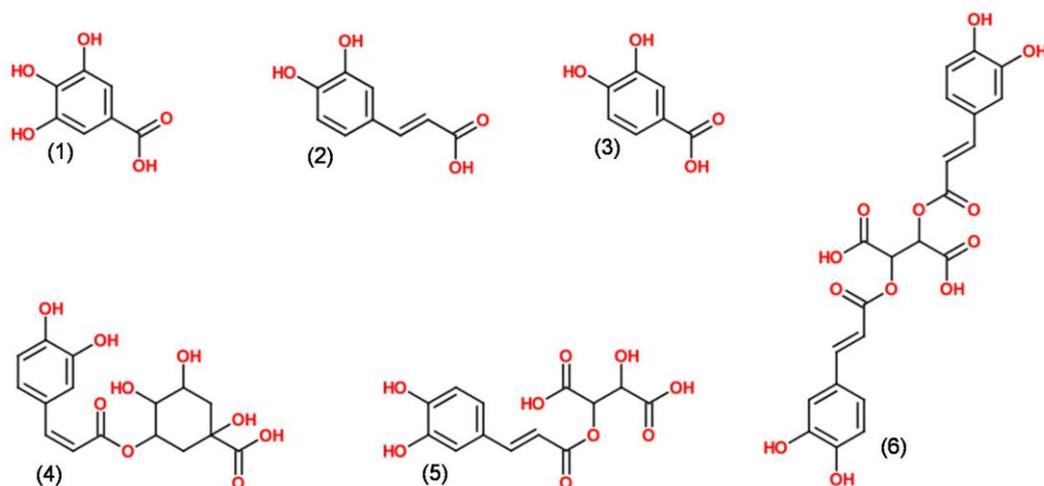


Figure 3. Les principaux acides phénoliques retrouvés chez la chicorée : (1) acide gallique, (2) acide caféique, (3) acide protocatéchique, (4) acide chlorogénique, (5) acide caftarique, (6) acide chicorique.

INTRODUCTION BIBLIOGRAPHIQUE

espèce	<i>Cichorium intybus</i>											<i>Cichorium endivia</i>			
	variété	sativum	foliosum				silvestre						escarole	crispum	latifolium
sous-variété	Magdebourg	Witloof	Spadona	Catalogna	Sauvage	Castelfranco	Verona	Chioggia	Treviso	Grumolo					
Couleur des feuilles	v	b	v	v	v	v/r	r	r	v	b/r	r	v	v	v	v
Acides phénols															
acide gallique						5	5	5	5	5	5	5			
acide caféique	1					5	5	5				5			
acide 5-O-caféoylquinique (acide chlorogénique)	1	4		4	2	5	5	5, 7, 4	5	5, 4	5	5	3	8	8
acide 3,5-Di-O-caféoylquinique (acide isochlorogénique)													3	8	
acide 3-O-caféoylquinique (acide néochlorogénique)													3		
acide 4-O-caféoylquinique (acide cryptochlorogénique)													3		
acide dicaféoyl tartarique (acide chicorique)	1	4	6	4	6, 2	5	5	5, 7, 4	5	5, 4	5	5	3		
acide monocaféoyl tartarique (acide caftarique)		4	6	4	6, 2								3		8
acide dicaféoyl quinique			6		6										
acide 3,4-Di-O-caféoylquinique														8	8
acide monocaféoyl <i>p</i> -hydroxycinnamique tartarique					2										
acide caféoyl feruloyl tartarique					2										
acide 5-O-feruloylquinique														8	8
acide protocatéchique						5	5	5	5	5	5	5			

INTRODUCTION BIBLIOGRAPHIQUE

Flavonoïdes															
kaempférol 3-O-glucuronide	1														
kaempférol 3-O-glucoside	1											3			
kaempférol malonyl glucoside			6		6										
lutéoline glucuronide			6		6		7								
lutéoline 7-O-glucuronide	1			4	2		7		4						
quercétine							7								
quercétine glucuronide			6	4	6		4								
quercétine 3-O-glucuronide	1				2				4						
quercétine 3-O-glucoside					2		7								
quercétine malonyl glucoside			6		6										
quercétine 3-O-galactoside	1														
apigénine glucuronide			6		6										
isorhamnetine 3-O-glucuronide	1														
crysoériol glucoside			6												
cyanidine 3-O-glucoside						5	5	5, 7, 4	5, 4	5					
cyanidine 3-O-malonyl glucoside							5	5, 7, 4	5, 4	5					
cyanidine 3-O-rutinoside							5								
delphinidine 3-O-malonyl glucoside								4	4						
malvidine							5	5	5	5					
malvidine 3-O-glucoside							5								
pelargonidine							5	5	5	5					
pelargonidine 3-O-glucoside							5								
péonidine							5	5	5	5					
péonidine 3-O-glucoside							5								

Tableau 2. Récapitulatif des molécules identifiées chez différentes espèces et variétés de chicorée, selon les articles de 1 - Rees, 1985 ; 2 - Mulinacci, 2001 ; 3 - Llorach, 2004 ; 4 - Innocenti, 2005 ; 5 - Rossetto, 2005 ; 6 - Heimler 2007 ; 7 – Lavelli 2008 ; 8 - Papetti, 2008.

b = blanche ; v = verte ; r = rouge ; v/r = verte avec des tâches rouges ; b/r = blanches avec des tâches rouges.

D'autres dérivés de l'acide cinnamique sont retrouvés chez la chicorée : des coumarines dont l'aesculétine, l'aesculine et la cichoriine. Celle-ci aurait un rôle de répulsif face aux herbivores (Rees & Harborne, 1985). D'autres publications sur la chicorée font état de la présence de molécules du métabolisme secondaire, peu antioxydantes mais répulsives : les sesquiterpènes lactones, conférant un goût amer aux racines et aux feuilles de la chicorée. Sont d'ores et déjà identifiées la lactucine, lactucopicrine, la 8-déoxylactucine et des dérivés 11 β ,13-dihydro (Foster *et al.*, 2006); (Hance *et al.*, 2007).

Etant donnée l'importance des polyphénols pour la plante elle-même et leurs effets potentiellement bénéfiques sur l'Homme, connaître les gènes responsables de la variation de leur concentration intéresse tant le secteur fondamental que le secteur appliqué.

III. Biosynthèse des polyphénols et gènes candidats

La voie de biosynthèse des polyphénols est appelée voie des phénylpropanoïdes. Elle débute avec un acide aminé, la phénylalanine. La première étape de cette voie est catalysée par l'enzyme phénylalanine ammonia-lyase (PAL). Au cours de cette réaction, la phénylalanine est convertie en acide cinnamique. Cette molécule aboutira ensuite sur les différents précurseurs de chacune des classes de polyphénols. Parmi eux, on trouve les acides hydroxycinnamiques comme l'acide coumarique, précurseur de la voie des flavonoïdes et les acides caféique, férulique et sinapique ainsi que leurs dérivés, précurseurs de la voie de biosynthèse des lignines (figure 4).

Etant donné l'importance des hydroxycinnamates chez la chicorée et particulièrement des acides caféique, chlorogénique, caftarique et chicorique (tableau 2), l'étude bibliographique présentée dans les paragraphes suivants se focalise sur les gènes candidats régulant la synthèse de ces molécules c'est-à-dire des enzymes et facteurs de transcription.

III.1. Biosynthèse de l'acide caféique : enzymes impliquées

La formation de l'acide caféique (figures 4 et 5), molécule dont dérivent plusieurs molécules identifiées chez la chicorée, est effectuée essentiellement par les enzymes phénylalanine ammonia-lyase (PAL), cinnamate-4-hydroxylase (C4H) et *p*-coumarate-3-hydroxylase (C3H).

Chez d'autres organismes (Kyndt *et al.*, 2002), (Watts *et al.*, 2006), l'enzyme tyrosine ammonia-lyase (TAL) peut remplacer la PAL et la C4H en produisant directement de l'acide *p*-coumarique à partir de la tyrosine (figure 4).

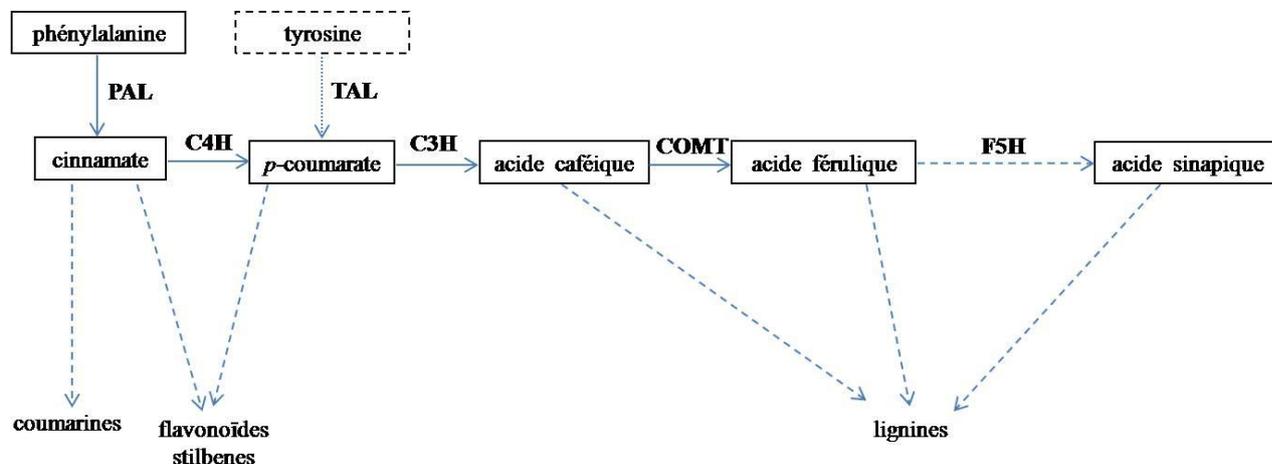


Figure 4. Synthèse des principaux hydroxycinnamates, précurseurs dans la voie des phénylpropanoïdes. PAL, phénylalanine ammonia-lyase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumarate-CoA-ligase; C3H, *p*-coumarate-3-hydroxylase; COMT, caféoyl-O-méthyltransférase; F5H, férulate-5-hydroxylase.

III.2. Biosynthèse des acides chlorogéniques : enzymes impliquées

On appelle acide chlorogénique (ACG) les esters d'acides caféique ou férulique avec l'acide quinique. Ce sont des précurseurs de la voie des lignines. Ils sont classés en trois catégories : les acides caféoylquiniques, les acides di-caféoylquiniques et les acides féruloylquiniques.

On a découvert récemment qu'à partir de l'enzyme C4H et la formation de l'acide *p*-coumarique, deux acyltransférases catalysent la formation et l'hydrolyse des acides chlorogéniques (Hoffmann *et al.*, 2004), (Niggeweg *et al.*, 2004). Il s'agit de l'hydroxycinnamoyl-CoA shikimate/quinique hydroxycinnamoyl transférase (HCT) et de l'hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transférase (HQT). Chez *Arabidopsis thaliana*, l'inactivation de l'expression du gène codant l'enzyme HCT génère une plante naine et altère la composition de la lignine, montrant l'implication de cette enzyme dans la biosynthèse des ACG (Hoffmann *et al.*, 2004). La surexpression du gène HQT chez la tomate a provoqué une augmentation de 85% de la quantité d'ACG dans les feuilles alors que son inactivation induit une baisse de 98% de la teneur en ACG (Niggeweg *et al.*, 2004).

Une troisième enzyme, déjà vue dans le paragraphe III.1 intervient dans le métabolisme des ACG : la C3H, appartenant à la famille des cytochromes P450 (Schoch *et*

al., 2001). Ainsi, l'hydroxylation en position 3 de l'ester formé par l'acide *p*-coumarique et l'acide quinique forme l'acide caféoylquinique, retrouvé chez la chicorée. De la même manière, C3H permet l'hydroxylation de l'acide *p*-coumaroyl shikimique, aboutissant alors à l'acide caféoyl shikimique. Les deux esters ainsi précédemment cités peuvent ensuite être hydrolysés par l'HCT ou l'HQT, pour former le caféoyl Co-A. Ce dernier sert de substrat à l'enzyme caféoyl-CoA 3-O-méthyltransférase (CCoAOMT) pour l'obtention du féruloyl CoA (figure 5), marquant le début de la voie de biosynthèse des lignines (Zhong *et al.*, 1998).

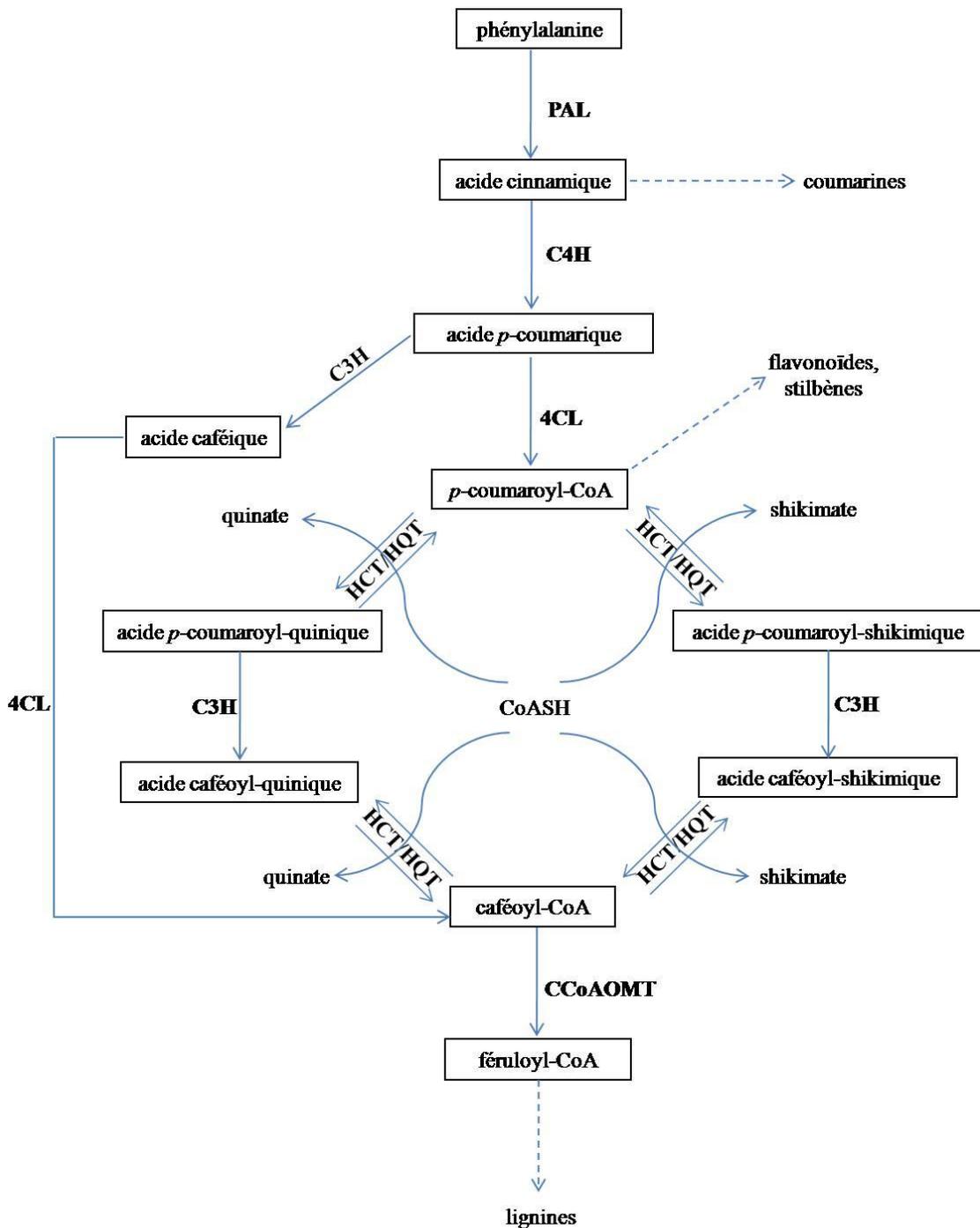


Figure 5. Détail de la voie de métabolisme des polyphénols chez *Arabidopsis* selon Hoffmann *et al.*, 2004 : biosynthèse de l'acide caféique et chlorogénique. PAL, phénylalanine ammonia-lyase ; C4H, cinnamate-4-hydroxylase ; 4CL, 4-coumarate-CoA-ligase ; C3H, *p*-coumarate-3-hydroxylase ; HCT, hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transférase ; HQT, hydroxycinnamoyl-CoA quinate hydroxycinnamoyl tranférase ; CCoAOMT, caféoyl-CoA 3-O-méthyltransférase.

III.3. Biosynthèse des acides chicorique et caftarique : enzymes impliquées

Ces deux acides hydroxycinnamiques, accumulés par la chicorée, sont retrouvés beaucoup moins couramment chez d'autres espèces que les molécules citées dans les paragraphes précédents. Ces molécules sont cependant accumulées chez d'autres membres de la famille des Astéracées : le pissenlit (*Taraxacum*) (Schütz *et al.*, 2005), la laitue (*Lactuca*) (Romani *et al.*, 2002) et l'échinacée (*Echinacea*) pour laquelle l'acide chicorique est le composé phénolique le plus important (Hudec *et al.*, 2007). A notre connaissance, aucune trace de ces deux molécules n'est pour l'instant retrouvée chez l'artichaut (Coinu *et al.*, 2007) et le tournesol (Weisz *et al.*, 2009). L'acide chicorique est également détecté chez d'autres genres tels que les Lamiacées, avec le basilic (Lee & Scagel, 2010), les Ericacées avec la myrtille (Chkhikvishvili & Kharebava, 2001) ainsi qu'une plante maritime, *Syringodium filiforme*, qui accumule en plus l'acide caftarique (Nuissier *et al.*, 2010). Toutefois, ne s'accumulant pas chez *Arabidopsis*, la voie de biosynthèse de l'acide chicorique et celle de l'acide caftarique restent encore relativement floues.

Les acides chicorique et caftarique sont des esters d'acide caféique et d'acide tartarique : deux molécules d'acide caféique et une molécule d'acide tartarique forment l'acide chicorique tandis que pour l'acide caftarique, une seule molécule d'acide caféique suffit. L'acide tartarique est aussi appelé thréarate. La forme L-(+) est la plus commune mais certaines légumineuses peuvent contenir la forme D-(-). Toutes les étapes de la biosynthèse du tartarate ne sont pas encore totalement élucidées et selon les espèces, trois voies différentes existent ou co-existent. Les deux premières voies sont issues de la décomposition de la vitamine C (DeBolt *et al.*, 2006), (Green & Fry, 2005) tandis que la troisième voie possible provient du D-gluconate.

Chez la plupart des plantes, l'acide ascorbique se dégrade comme indiqué dans la figure 6. Les étapes (1) à (6) seraient possibles sans intervention enzymatique bien que certaines réactions peuvent être accélérées grâce à certaines enzymes. Par exemple, les réactions (1) et (2) peuvent être catalysées par une oxydase et la réaction (6) par une oxalyl estérase. Après clivage entre les carbones 2 et 3, on obtient de l'acide oxalique et de l'acide tartarique (Loewus, 1999). La réaction (7) serait dépendante d'une oxalate oxydase tandis que

l'étape (8) reste mal connue. L'intervention d'une oxydase et d'une molécule intermédiaire, le L-thréo-tétruronate, est envisagée (Green & Fry, 2005).

Chez les Viticeae, une autre voie peut co-exister avec la première. L'acide ascorbique serait converti en acide 2-kéto l-idonique, puis réduit en acide l-idonique. Cette molécule serait alors oxydée pour former l'acide 5-kéto d-gluconique, réaction catalysée par l'enzyme idonate déhydrogénase. Suite au clivage entre les carbones 4 et 5, on obtient une molécule à 4 carbones, le L-thréo-tétruronate qui s'oxyderait pour former l'acide tartarique. On suppose qu'avec le fragment de deux carbones restant, il va se former un glycoaldéhyde (DeBolt *et al.*, 2006).

L'acide tartarique est ensuite utilisé dans la voie de métabolisme de l'acide glyoxylique (figure 7).

Les enzymes catalysant la formation des acides caftarique et chicorique à partir des molécules d'acide caféique et tartarique ne sont pas encore caractérisées. Il pourrait s'agir d'enzymes de type hydroxycinnamoyltransférase comme c'est le cas pour d'autres esters d'acide caféique et tartarique : les acides mono- et di-caffeoyl-meso-tartarique (Hohlfeld *et al.*, 1996).

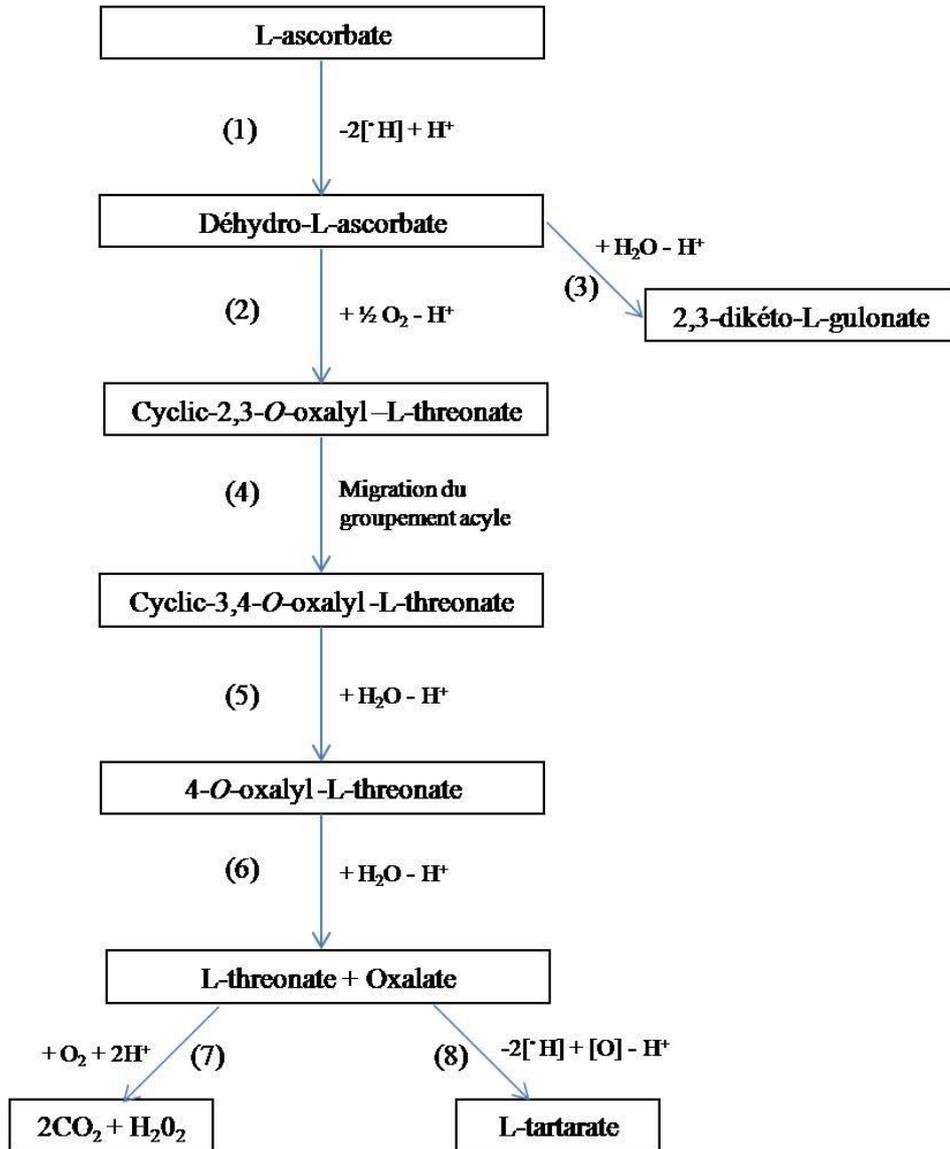


Figure 6. La voie de catabolisme de l'acide ascorbique selon Green *et al*, 2005. Les étapes (7) et (8) donnant l'acide tartarique d'une part et deux molécules de dioxyde de carbone et l'eau oxygénée d'autre part, sont des réactions enzyme-dépendantes.

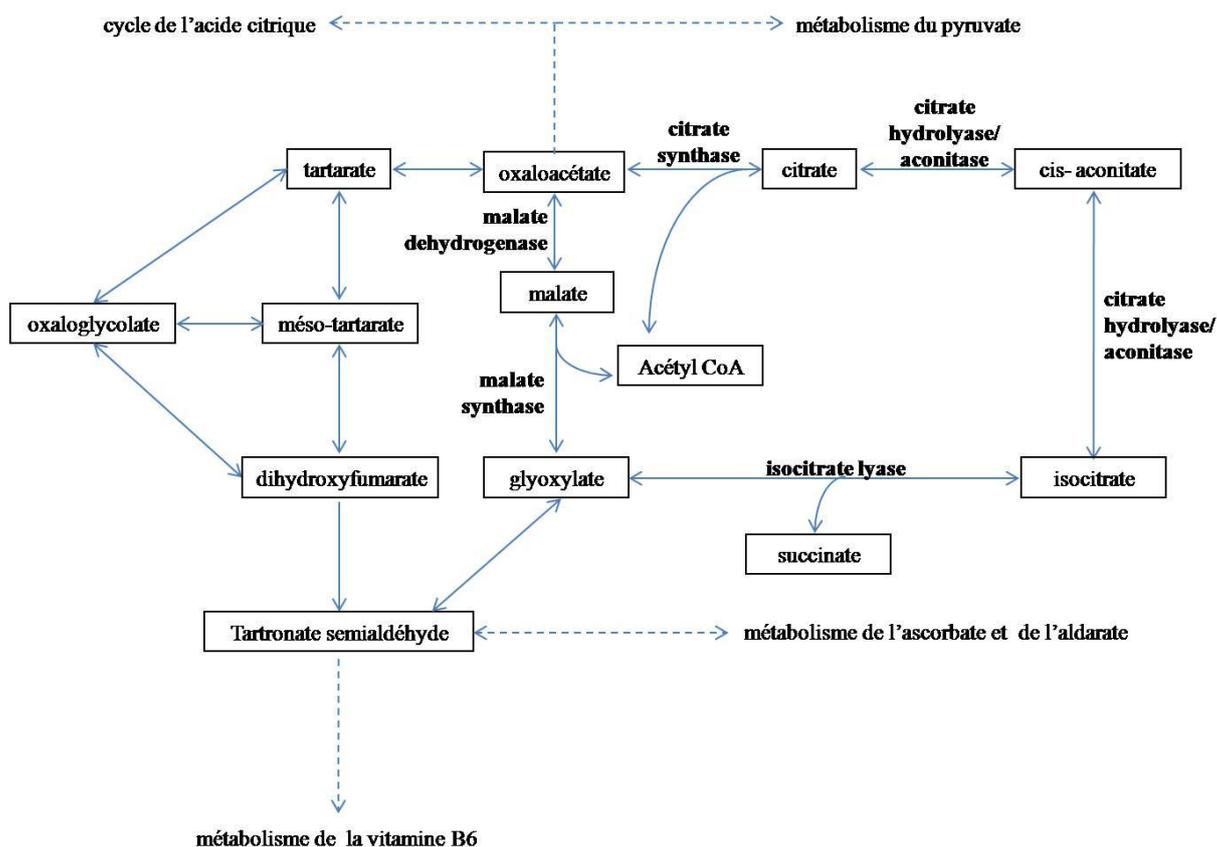


Figure 7. Détail de la voie de métabolisme du glyoxylate chez *Arabidopsis thaliana* (<http://www.genome.jp/kegg/pathway.html>).

III.4. Facteurs de transcription régulant la biosynthèse des polyphénols

Toute protéine nécessaire au contrôle de la transcription, menée par l'ARN polymérase, est appelée facteur de transcription. Certains facteurs de transcription peuvent reconnaître d'autres facteurs de transcription liés à l'ADN ; d'autres reconnaissent directement l'ARN polymérase tandis qu'un troisième type de facteurs de transcription agit sur des séquences *cis*-régulatrices au niveau des promoteurs de leurs gènes cibles (Maston *et al.*, 2006). Ils peuvent être ubiquitaires, c'est-à-dire communs à plusieurs gènes, ou spécifiques. Ils peuvent également être constitutifs ou inductibles, c'est-à-dire spécifiques à un tissu ou un moment donné du cycle cellulaire ou en réponse à un signal extérieur.

Les facteurs de transcription de type MYB peuvent directement se lier à l'ADN et agir en tant que répresseur ou activateur de l'expression des gènes. Ils reconnaissent de façon spécifique les séquences YAAC(G/T)G via un motif protéique de type hélice-tour-hélice (Biedenkapp *et al.*, 1988). Les protéines MYB se définissent par la présence de domaines

constitués d'environ 50 acides aminés. Ces domaines peuvent être au nombre d'un, deux ou trois : R1, R2, R3.

Le premier gène codant une protéine MYB a été découvert dans le domaine animal : il s'agit de l'oncogène *v-myb* issu du virus du MyéloBlastome aviaire, intégré dans le génome du poulet (Gonda *et al.*, 1982), (Klempnauer *et al.*, 1982) et donnant aux protéines MYB leur nom. Chez les végétaux, le premier MYB identifié est le gène *c1*, intervenant dans la biosynthèse des anthocyanes du grain de maïs (Paz-Ares *et al.*, 1987). Il s'agit d'un MYB avec deux domaines de liaison à l'ADN (R2R3) comme c'est le cas de la plupart des MYB végétaux. Ensuite, d'autres protéines MYB type R2R3 ont été identifiées, chez *Antirrhinum majus* ou Gueule de loup : AmMYB308 et AmMYB330 répriment la synthèse d'acides phénols et de lignines quand leur gène est surexprimé chez le tabac (Tamagnone *et al.*, 1998). En 2000, une plante transgénique (gène *PAP1* surexprimé) très pigmentée d'*Arabidopsis thaliana* a été isolée (Borevitz *et al.*, 2000). Cette surexpression a entraîné une augmentation de transcrits codant l'enzyme PAL, la chalcone synthase (CHS, enzyme du début de la voie de biosynthèse des flavonoïdes), ainsi que les enzymes DFR (dihydroflavonol réductase, voie des flavonoïdes, spécifique aux anthocyanes) et GST (Glutathione S-transferase, certaines étant impliquées dans le transport des anthocyanes vers la vacuole). Phénotypiquement, on note une accumulation anormale de lignines, d'esters d'acides hydroxycinnamiques et de flavonoïdes dont des anthocyanes, responsables de la couleur violette de ce mutant nommé *pap1-D* (production of anthocyanin pigment 1-Dominant). Le gène *PAP1*, surexprimé ici, a été déterminé comme codant un facteur de transcription MYB de type R2R3 capable d'accroître l'expression de diverses enzymes de la voie des phénylpropanoïdes, dans tous les organes végétatifs et tout au long de leur développement. Un EST (Expression Séquence Tag) très similaire à *PAP1*, codant également un facteur de transcription MYB responsable d'un phénotype semblable à celui de *pap1-D*, a été appelé *PAP2*. La même année, un gène codant pour un répresseur transcriptionnel du gène *C4H* chez *Arabidopsis* a été caractérisé : *AtMYB4* (Jin *et al.*, 2000).

Chez d'autres espèces, des facteurs de transcription MYB R2R3 semblent également réguler plusieurs branches de la voie de biosynthèse des polyphénols. Jusqu'à la publication de Luo en 2008 (Luo *et al.*, 2008), *AtMYB12* était considéré comme un facteur de transcription spécifique à l'activation de gènes impliqués dans la production de flavonols chez l'Arabette, mais l'expression d'*AtMYB12* chez la tomate est également associée à la voie de biosynthèse des acides cafféoylquiniques et augmente les transcrits des gènes PAL, 4CL, C3H, C4H, HCT, HQT, entre autres. La surexpression du gène *VvMYB5a* (vigne) chez

Nicotiana tabacum altère le métabolisme des lignines et provoque une accumulation d'anthocyanes, de flavonols et de tannins condensés (Deluc *et al.*, 2006). De manière analogue, la surexpression du gène *StMtf1^M* chez la pomme de terre est associée à l'activation du métabolisme des anthocyanes mais permet aussi d'augmenter d'un facteur cinq le taux de transcrits du gène *Hqt*, entraînant alors un taux quatre fois plus élevé en caffeoylquinat que la normale (Rommens *et al.*, 2008).

IV. Architecture génétique de caractères quantitatifs

Afin de compléter la liste des gènes codant des enzymes et des facteurs de transcription impliqués dans une voie métabolique d'intérêt, l'analyse QTL (Quantitative Trait Loci) représente une alternative prometteuse. En effet, ce type d'analyse offre la possibilité d'associer des zones chromosomiques aux caractères quantitatifs étudiés. La découverte des gènes déterminant le phénotype peut alors se faire ultérieurement, en se focalisant sur les régions mises en évidence, via la cartographie fine par exemple.

Dans le cas où des gènes candidats seraient déjà connus, les analyses QTL permettent de supposer l'implication de certains gènes candidats et de connaître l'intensité de leur effet éventuel mais aussi de mettre en exergue des gènes à effet mineur, souvent non détectés par des méthodes de génétique directe comme la mutagenèse.

IV.1. Origine de l'analyse QTL

De tout temps, les espèces d'intérêt économique ont été sélectionnées sur la base de leur phénotype. Ces croisements dirigés par l'Homme ont abouti à des génotypes plus résistants aux stress biotiques et abiotiques, plus riches en tel ou tel composé ou encore plus vigoureux et plus productifs. L'amélioration de ces caractères n'est possible que s'il existe une variabilité génétique. Cette variabilité est révélée lorsque la distribution des caractères mesurés est continue, indiquant que le phénotype « P » est sous l'influence de facteurs génétiques « G » (plusieurs gènes) et de facteurs environnementaux « E » : $P = G + E$. Ainsi, la génétique quantitative permet de mettre en relation phénotype et génotype (Fisher, 1918). En 1923, Sax découvre une association entre la pigmentation des téguments de la graine (caractère mendélien) et la taille du grain du haricot (caractère à variation continue). Cette corrélation a été attribuée à la liaison entre le facteur génétique qui contrôle la taille du grain

et le locus déterminant sa couleur. Ensuite, Smith (1937), Thoday (1961) puis Spickett et Thoday (1966) ont montré que les caractères mendéliens pouvaient être utilisés comme marqueurs de régions chromosomiques, pour identifier des loci affectant un caractère quantitatif. Cependant, cette approche a été freinée par la difficulté de trouver des marqueurs balisant l'ensemble du génome, les marqueurs employés jusqu'ici étant morphologiques, protéiques ou isoenzymatiques.

Avec le développement des techniques de biologie moléculaire au début des années 80, les facteurs génétiques ne sont plus exclusivement reliés au phénotype mais aussi à des marqueurs moléculaires localisés sur une carte génétique. Des loci correspondants aux caractères quantitatifs peuvent alors être localisés sur des groupes de liaisons denses en marqueurs moléculaires, correspondant au nombre de chromosomes de l'espèce étudiée. Outre la simple cartographie, l'analyse QTL, pour Quantitative Trait Loci (Geldermann, 1975), permet aussi de rendre compte de l'effet des loci (positif/négatif, pourcentage de variation phénotypique expliquée) intervenant dans l'expression d'un phénotype donné.

IV.2. Outils d'une analyse QTL

IV.2.A. Une descendance en ségrégation

Pour ce type d'analyse reposant sur l'association entre la variabilité génotypique et phénotypique, la descendance étudiée doit être polymorphe pour ces deux aspects. Il y a plusieurs façons d'obtenir une population en ségrégation.

Les espèces autogames étant souvent peu polymorphes, le choix des parents se fait de préférence sur des lignées pures sans ancêtre commun. Les espèces allogames, hétérozygotes pour de nombreux loci, sont riches en polymorphismes. Le choix des parents se dirige alors vers deux individus les plus contrastés possibles pour le caractère à analyser.

Les descendance les plus couramment utilisées sont:

- haploïdes doublés : plantes diploïdes après avoir subi un doublement chromosomique (Zang *et al.*, 2003);
- les descendance F_2 issues de l'autofécondation d'individus F_1 (Mano *et al.*, 2005);
- les populations « bulk F_3 » : autofécondation de chaque individu F_2 (Zhou *et al.*, 2003);

- les lignées recombinantes (en anglais RIL pour Recombinant Inbred Lines), obtenues par autofécondation d'individus F_1 , F_2 puis F_3 etc. jusqu'à stabilisation des lignées. Les individus à étudier sont homozygotes et polymorphes (Simon *et al.*, 2008);
- les populations issues de rétrocroisement (en anglais BIL pour Back-cross Inbred Lines). L'hybride F_1 est croisé avec un des parents : les descendants sont autofécondés jusqu'à l'obtention de lignées d'homozygotes (Dong *et al.*, 2008).
- les lignées d'introgession (en anglais IL pour Introgression Lines) : population de lignées quasi-isogéniques, c'est-à-dire issues de rétro-croisement entre une lignée receveuse (celle dont le génome se retrouve en proportion majoritaire et qui constitue le fond génétique des lignées) et une lignée donneuse (Rousseaux *et al.*, 2005).

Ces types de descendance, toujours issues de parents homozygotes, sont difficilement envisageables avec les espèces pérennes à cycle long comme les arbres tandis que l'obtention de lignées pures n'est pas possible avec les espèces auto-incompatibles comme c'est le cas, parmi d'autres, pour beaucoup de Brassicacées comme *Arabidopsis hallieri* et *lyrata*, mais aussi pour des Astéracées comme la chicorée, le tournesol, l'artichaut ou le séneçon (Takayama & Isogai, 2005).

Dans le cadre de mes travaux de thèse, deux parents hétérozygotes K59 et K28 ont été croisés afin d'obtenir une descendance F'_1 : il peut y avoir jusqu'à quatre allèles par locus (Hayashi & Awata, 2004). Le choix des parents pour cette descendance a été fait en fonction de leur potentiel à produire des embryons somatiques, sujet à l'origine du projet de cartographie chez *Cichorium intybus* L. (Cadalen *et al.*, 2010) (cf. Chapitre II, Données Génotypiques). Ainsi, l'apport manuel de pollen du parent K59, très embryogène, sur les stigmates du parent K28, mâle stérile et peu embryogène, a permis d'obtenir une descendance en ségrégation pour le caractère « capacité à former des embryons somatiques ».

IV.2.B. Les marqueurs moléculaires et la construction d'une carte génétique

La saturation d'une carte génétique par des marqueurs moléculaires est une étape clé précédant une analyse QTL. Plus une carte génétique sera dense en marqueurs, plus la localisation des QTL détectés et leur effet seront précis. Il existe de nombreux types de marqueurs moléculaires. Ils permettent de révéler soit du polymorphisme d'unité de répétition

(microsatellites), soit du polymorphisme de séquence caractérisé par des différences au niveau de sites de restriction, de conformation, de stabilité ou encore de sites d'hybridation d'amorces oligonucléotidiques. Ces marqueurs moléculaires, qu'ils soient dominants ou co-dominants, représentent le fondement d'une carte génétique. Les marqueurs y sont répartis en fonction de leur taux de recombinaison r . Plus la chance d'observer un crossing-over est importante, plus la distance génétique entre deux marqueurs est grande.

Pour la carte génétique correspondant au croisement K28 x K59, trois types de marqueurs ont été générés pour couvrir l'ensemble des neuf groupes de liaison de la carte génétique de la chicorée :

- des marqueurs révélant du polymorphisme de longueur de fragments d'amplification (AFLPTM : Amplification Fragment Length Polymorphism) : marqueurs dominants et révélés en masse (Vos *et al.*, 1995), figure 8.
- des marqueurs révélant du polymorphisme de conformation simple brin via la technique SSCP, Single Strand Conformation Polymorphism) (Orita *et al.*, 1989), figure 9.
- des marqueurs révélant du polymorphisme de répétition de séquences simples, di-, tri- ou tétranucléotidiques, appelés marqueurs SSR ou Simple Sequence Repeats (Tautz, 1989), figure 10.

Ces deux derniers types de marqueurs sont co-dominants et révélés de façon individuelle.

INTRODUCTION BIBLIOGRAPHIQUE

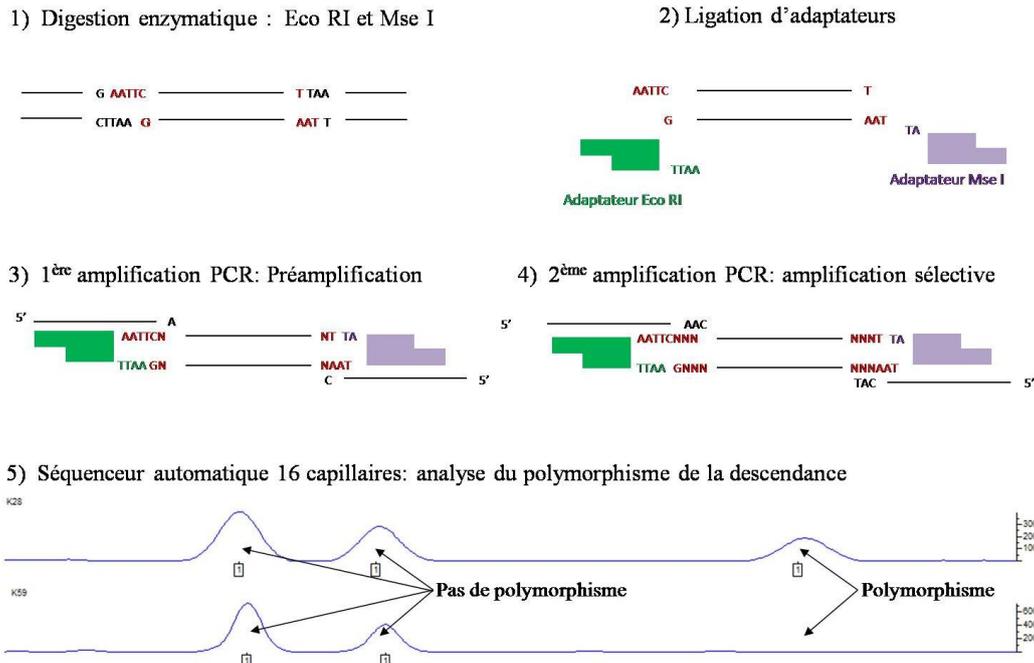


Figure 8. Illustration de la technique AFLP, polymorphisme de longueur de fragments d'amplification. L'ADN est digéré par des enzymes de restriction Eco RI et Mse I. Des adaptateurs de séquence connue sont ensuite ajoutés aux extrémités des bouts cohésifs. Ensuite, deux amplifications PCR sélectives se succèdent : à chaque fois, des amorces qui correspondent aux adaptateurs sont prolongées en 3' par une puis deux amorces arbitraires, respectivement. Ce type de polymorphisme est alors visualisé après migration des échantillons sur électrophorèse capillaire en condition dénaturante.

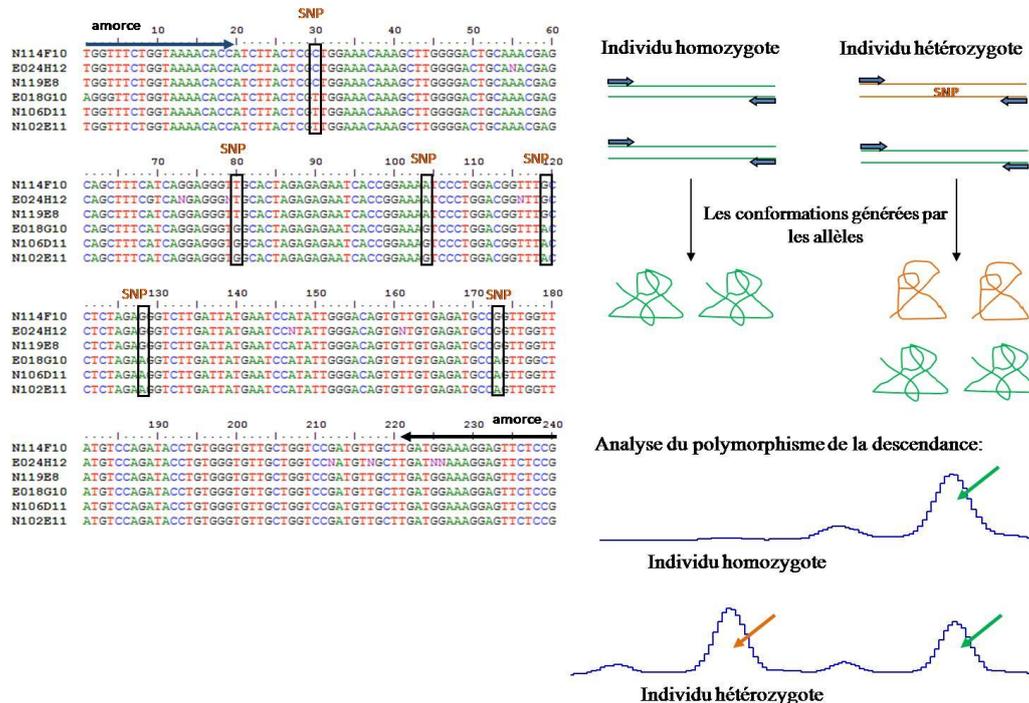
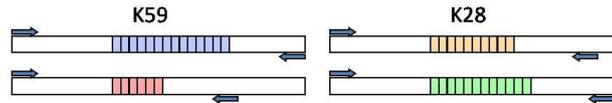


Figure 9. Illustration de la technique SSCP, polymorphisme de conformation. Les fragments d'ADN sont préalablement amplifiés avec des amorces spécifiques marquées en 5' par un fluorochrome. L'ADN double brin est dénaturé par chauffage puis refroidi rapidement dans la glace, ce qui empêche la réassociation des doubles brins d'ADN et autorise la formation de structures secondaires stables dues à la présence de SNP. La migration dans un gel de polyacrylamide non dénaturant permet la visualisation du polymorphisme généré par les différentes conformations obtenues.

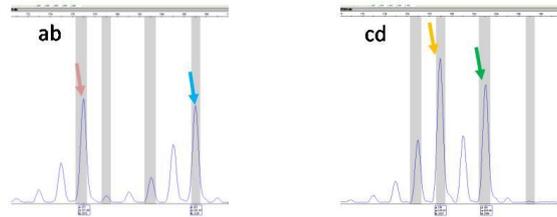
INTRODUCTION BIBLIOGRAPHIQUE

1) PCR: amplification de l'ADN des individus avec les amorces marquées



2) Séquenceur automatique 16 capillaires

3) Analyse du polymorphisme des parents: ab x cd.



4) Analyse du polymorphisme dans la descendance. Profils observés dans la descendance: ac, ad, bc et bd.

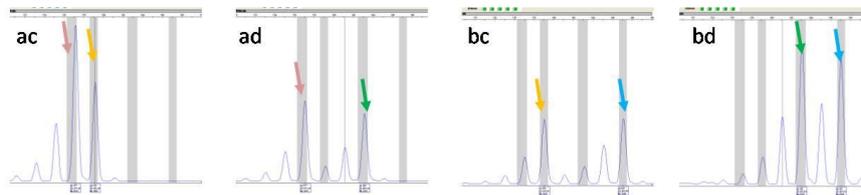


Figure 10. Illustration de la technique SSR, polymorphisme d'unité de répétition. Les microsatellites sont composés de répétitions de motifs mono-, di-, tri- ou tétranucléotidiques. Des amorces sont définies dans les régions flanquantes des microsatellites pour déceler un polymorphisme de longueur de ces séquences répétées après migration des échantillons sur électrophorèse capillaire en condition dénaturante.

IV.2.C. Le phénotypage

Les caractères liés à la biosynthèse des polyphénols sont exprimés sous la forme de caractères quantitatifs comme par exemple des concentrations, quantités ou pourcentages de composé ou d'un ensemble de molécules d'intérêt extrait à un instant t de la vie du végétal, dans des conditions spécifiques. Ces caractères quantitatifs, s'ils suivent une distribution gaussienne, impliquent plusieurs loci ou QTL.

IV.2.D. Cartographie de QTL

L'analyse QTL permet de mettre en relation le phénotypage et le génotypage grâce à des outils statistiques. Des loci impliqués dans la détermination du phénotype sont ainsi localisés sur la carte génétique construite préalablement. Il existe deux façons de situer ces QTL : soit en considérant les marqueurs un à un, soit en considérant l'intervalle entre deux marqueurs.

La première méthode consiste à rechercher une relation statistique entre le génotype de chaque marqueur et la valeur du caractère quantitatif étudié, par une analyse de variance. On teste donc un seul marqueur à la fois. Son avantage réside en sa simplicité, mais si la densité de marqueurs est faible, la localisation du QTL sur la carte sera d'autant moins précise. Cette méthode n'apporte pas d'indication quant à l'effet du QTL.

La seconde méthode est appelée cartographie d'intervalle, méthode initiée par Weller en 1987 et reprise par Lander et Botstein en 1989. On teste la présence d'un QTL entre deux marqueurs, reliés entre eux par un taux de recombinaison r , en calculant le logarithme décimal du rapport de vraisemblance pour tester l'hypothèse nulle « absence de QTL » contre l'hypothèse alternative de la « présence d'un QTL ». Ce rapport, appelé aussi LOD score (logarithm -base 10- of odds), est calculé selon la formule suivante.

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

- a = effet additif et d = degré de dominance ;
- $V(a_1, d_1)$ correspondant à la valeur de la fonction de vraisemblance sous l'hypothèse de présence d'un QTL ;
- $V(a_0, d_0)$ correspondant à la valeur de la fonction de vraisemblance d'absence d'un QTL, c'est-à-dire lorsque $a_0 = 0$ et $d_0 = 0$.

Ce calcul est appliqué pour chaque intervalle génétique entre deux marqueurs d'un groupe de liaison donné. Si le LOD score, pour un intervalle entre deux marqueurs donnés, est égal à 2, 3, ou 4, cela indique que la présence d'un QTL est 100, 1000 ou 10000 fois, respectivement, plus probable que son absence. On peut ainsi tracer une courbe avec en abscisse les marqueurs et en ordonnée le LOD score. Le maximum atteint par la courbe indique alors la localisation la plus probable du QTL, s'il dépasse un seuil de signification préfixé (détails dans le chapitre III, Analyse QTL).

Darvasi *et al.* ont démontré que la puissance de détection et la précision de la localisation d'un QTL sont augmentées si le nombre d'individus est plus important (Darvasi *et al.*, 1993).

En 1993, Zeng propose d'introduire des marqueurs adjacents comme co-variables pour réduire les effets des QTL avoisinant. Cette méthode, appelée Composite Interval Mapping (CIM) n'est pourtant pas suffisante pour l'étude de l'architecture génétique de caractères complexes dotés de plusieurs QTL. Le Multiple Interval Mapping (MIM) suggéré par Kao et Zeng en 1999 prend en compte les interactions entre les différents QTL et discute du phénomène d'épistasie (Zou & Zeng, 2008). Ces deux types d'approche sont plus généralement appelées MQM pour Multiple QTL Mapping.

Plusieurs analyses QTL ont été entreprises afin d'identifier les régions impliquées dans le métabolisme des composés phénoliques. Comme précisé dans le paragraphe II.2, ces molécules font partie du métabolisme secondaire et sont associées à de nombreux stress biotiques ou abiotiques, et donc particulièrement dépendantes de l'environnement « E », comme l'a souligné l'équipe de Rousseaux en 2005 (Rousseaux *et al.*, 2005). En effet, les QTL impliqués dans la biosynthèse de divers antioxydants (polyphénols, acide ascorbique) chez *Lycopersicon pennelli* se sont révélés peu stables d'une année sur l'autre : la plupart des loci détectés n'ont été observés qu'une fois sur trois saisons en champs. Lorsque la même expérience a été répétée en serre, la majorité des QTL n'a plus du tout été détectée.

L'identification de loci liés à la biosynthèse des hydroxycinnamiques restent peu courante (nombre limité de publications à ce sujet). C'est d'autant plus le cas pour les acides chicorique et caftarique accumulés par un nombre restreint d'espèces.

La laitue, pour laquelle des cartes génétiques ont été publiées (Landry *et al.*, 1987; Kesseli *et al.*, 1994; Syed *et al.*, 2006; Truco *et al.*, 2007), contient ces esters d'acide caféique mais aucune analyse QTL concernant ces molécules n'a été publiée jusqu'à présent. En 2001,

l'équipe de Baum a généré plus de 200 AFLP pour cinquante accessions d'*Echinacea*, utilisés comme empreinte ADN pour prédire les quantités d'acide chicorique (Baum *et al.*, 2001). Depuis 2008, une carte génétique d'échinacée, construite à partir de marqueurs AFLP, est disponible (Aziz & Sauve, 2008) mais aucune référence à une détection de QTL n'a été trouvée.

En 2006, une descendance F_1 de 96 artichauts (*Cynara cardunculus*, $2n = 2x = 34$) a permis l'obtention d'une carte génétique basée, entre autres, sur des AFLP et SSR (Lanteri *et al.*, 2006). Depuis, des séquences codant pour les enzymes PAL ont été isolées à partir d'amorces dégénérées (De Paolis *et al.*, 2008). Les gènes *HCT*, *HQT* et *C3H* ont été isolés de la même façon et cartographiés (Comino *et al.*, 2007; Comino *et al.*, 2009; Moglia *et al.*, 2009). En 2009, une seconde carte génétique a été publiée (Portis *et al.*, 2009).

INTRODUCTION BIBLIOGRAPHIQUE

Espèce	Partie de plante	Molécules étudiées	Population		Marqueur		Analyse QTL			Références
			descendance	taille	type	nombre	Type d'analyse	Nombre de QTL détectés	% de la variation expliquée	
Maïs	soie	chlorogénate	F ₂ x 3	450 437 380	SSR	84 à 88	CIM	4 à 6 selon la population	de 2,9 à 42,7	Szalma, 2002
Pomme de terre	tubercule	chlorogénate	DH	249	AFLP	180	IM	1	21	Werij, 2007
Riz	grain	- polyphénols - flavonoïdes - capacité antioxydante	DH	117	RFLP SSR	243	CIM	3 3 3	de 6,07 à 16,91 de 6,94 à 12,71 de 8,83 à 13,47	Jin, 2009
Riz	paille	acide férulique	RIL	165	RFLP	289	CIM	3	de 5,5 à 16,9	Dong, 2005
Maïs	chaume	- esters d'acide p-coumarique - acide férulique et esters	IL RIL	25 131	SSR	108	ANOVA	4 6	de 7,5 à 10,3 de 6,8 à 12,1	Fontaine, 2003
Maïs	chaume	acide p-coumarique et férulique (et dérivés)	RIL	242	SSR	249	CIM	7 21	de 5,3 à 13,3 de 5,8 à 21,8	Barrière, 2008
Lys	tépale	anthocyane	F ₁	96	RAPD ISSR	212	-	2	de 16,9 à 63,8	Abe, 2002
Framboisier	fruit	anthocyane	F ₁	188	AFLP SSR		IM	15	de 10 à 60	Kassim, 2009
Tomate	fruit	ascorbate	IL BC RIL	75 130 144	AFLP RFLP SSR Marq. PCR	217	IM	5 à 12 selon la population	de 2,14 à 22	Stevens, 2007
Soja	graine	tocophérol	RIL	144	SSR	463	ANOVA	21	de 2,4 à 16,7	Li, 2010
Arabette	feuille et graine	glucosinolates	RIL	162	AFLP SSR	162	IM, CIM	20	de 28 à 75	Kliebenstein, 2001
Tomate	feuille	- capacité antioxydante - polyphénols - flavonoïdes	IL	52	SSR	223	ANOVA	35 32 42	-	Frery, 2010

Tableau 3. Exemples de recherche de QTL impliqués dans la biosynthèse de métabolisme secondaire. “-“: information manquante.

Abbreviations: DH : Haploïde Doublé ; IL : Inbred Line ; RIL : Recombinant Inbred Line ; BC : Back Cross.

SSR : Simple Sequence Repeats; AFLP : Amplification Fragment Length Polymorphism ; RFLP : Restriction Fragment Length Polymorphism ; RAPD : Random-Amplified Polymorphic DNA; ISSR : Inter Simple Sequence Repeats; PCR : Polymerase Chain Reaction.

IM: Interval mapping; CIM: Composite Interval Mapping; ANOVA: ANALYSE OF VARIANCE.

Chez le tournesol, une carte génétique a été établie et des analyses QTL ont déjà été entreprises mais aucun de ces travaux ne concerne directement les voies de biosynthèse de molécules antioxydantes (Ronicke *et al.*, 2005; Dechaine *et al.*, 2009; Ebrahimi *et al.*, 2009).

Le tableau 3 donne un aperçu de ce qui a été réalisé chez différentes espèces, le type de descendance étudié, le type et le nombre de marqueurs utilisés pour la carte génétique et le nombre de QTL détectés, relatifs à la biosynthèse de différentes molécules issues du métabolisme secondaire. Ces éléments seront repris pour commenter les résultats présentés dans les chapitres suivants.

IV.3. En complément de l'analyse QTL : cartographie de gènes candidats

L'objectif de la détection de QTL est la mise en évidence de régions chromosomiques impliquées dans les variations phénotypiques observées. Il est donc très intéressant de trouver les gènes situés dans ces régions, c'est-à-dire les gènes dont le polymorphisme moléculaire serait lié à la variation phénotypique : les gènes candidats. Cette approche a été utilisée avec succès en génétique humaine et animale (Tabor *et al.*, 2002) ainsi que chez les végétaux depuis une vingtaine d'années (Pflieger *et al.*, 2001). Concrètement, les gènes candidats polymorphes sont à localiser sur la carte génétique comme n'importe quel autre marqueur moléculaire pour savoir s'ils co-localisent ou non avec les QTL. Dans notre cas, l'approche gènes candidats s'appuie sur les gènes correspondant aux facteurs de transcription et enzymes cités dans le paragraphe III, impliqués dans la biosynthèse des polyphénols. Le polymorphisme a été révélé par la technique SSCP et par la technique HRM (High Resolution Melting). Cette dernière approche, rapide, économique et sensible a permis de collecter des informations co-dominantes. Le principe de cette technique est basé sur une PCR avec incorporation d'un fluorochrome saturant complètement le double brin, suivi d'une courbe de fusion (figure 11). Deux génotypes qui diffèrent par la mutation d'une seule base présentent des courbes de fusion différentes, ce qui permet de mettre en évidence le polymorphisme.

V. Contexte de la thèse de doctorat

V.1. Partenariat industriel

Depuis une vingtaine d'années, les industriels produisant (Ets Florimond-Desprez) ou utilisant (Ets Leroux) la chicorée industrielle ont tissé d'étroites relations avec le laboratoire Stress Abiotiques et Différenciation des Végétaux cultivés.

Depuis près de deux siècles, la société Florimond Desprez Veuve et Fils se consacre à la sélection et l'amélioration génétique, pour l'obtention de variétés de culture à destination des agriculteurs et des industries agroalimentaires, français et européens. Les différents programmes de sélection concernent la betterave sucrière, la chicorée industrielle, le blé tendre et dur, l'orge, le colza, le pois, le lupin, la betterave fourragère et la luzerne.

La société Leroux déshydrate, torréfie puis transforme des racines de chicorée à Orchies dans le Nord de la France, depuis 1858. Spécialisé dans la production et la commercialisation de produits à base de chicorée, Leroux est aujourd'hui le leader en français et mondial sur le marché de la chicorée pure, aromatisée ou mixée à des céréales, avec 100 000 racines traitées par an. La filière Recherche & Développement vise à exploiter les propriétés technologiques et fonctionnelles de la chicorée notamment en vue d'applications à visée santé.

Au regard des éventuels bénéfices de santé accordés aux composés polyphénoliques, une étude du contrôle génétique de la production et de l'accumulation des molécules antioxydantes chez la chicorée est ainsi été initiée sous la forme d'une thèse de doctorat, co-financée par la région Nord-Pas-de-Calais et les industriels Leroux et Florimond Desprez.

V.2. Contexte scientifique et objectifs

Mes travaux portent sur les feuilles de chicorée uniquement bien que cette variété de chicorée soit cultivée pour ses racines. Plusieurs raisons justifient ce choix. Tout d'abord, le prélèvement foliaire n'est pas destructeur pour la plante contrairement au prélèvement des racines et peut être ainsi renouvelé pour multiplier les données. De plus, le développement foliaire est beaucoup plus rapide que le développement racinaire.

La thèse peut être décomposée en trois parties.

La première partie correspond à la mise en place du matériel végétal et d'un protocole d'extraction adapté au type végétal, aux molécules décelées et au nombre d'échantillons prévus. Comme la chicorée est une plante strictement allogame, chaque génotype est précieux et le clonage se fait par bouturage ou par embryogenèse somatique. Un des premiers objectifs de la thèse a donc été d'obtenir ce matériel végétal puis de savoir si cette population F1' présentait une telle variabilité.

Pour la deuxième partie, le génotypage des individus bouturés précédemment a été complété afin de construire une carte génétique spécialement dédiée à ce projet. Afin de compléter ce travail, une recherche de gènes candidats a été réalisée et ceux-ci ont été intégrés à la carte génétique.

La troisième partie consiste à relier le phénotypage (première partie) au génotypage (deuxième partie) pour la détection de QTL impliqués dans la biosynthèse de molécules antioxydantes.

CHAPITRE I

Mise en place d'une méthode
d'extraction des hydroxycinnamates et
collection des données phénotypiques

Ce premier chapitre évoque l'obtention du matériel végétal, la mise au point d'un protocole d'extraction des molécules les plus abondantes dans les feuilles de chicorée, et l'obtention des données phénotypiques correspondant aux teneurs des composés identifiés (acides caftarique, chlorogénique et chicorique) ainsi qu'à l'activité antiradicalaire (AAR) des extraits. Ces étapes ont permis de constater que la descendance, déjà utilisée au laboratoire SADV pour d'autres sujets de recherche, présentait une variabilité phénotypique pour l'AAR et pour les trois hydroxycinnamates cités précédemment. Elle s'est ainsi révélée tout à fait exploitable pour la recherche de QTL associés à ces phénotypes.

I. Historique de l'utilisation de la chicorée au sein du laboratoire SADV

Dès le début des années 1970, le laboratoire SADV (alors intitulé laboratoire de physiologie végétale) s'intéresse à *Cichorium intybus* L. et plus exactement à sa capacité à bourgeonner à partir des tissus racinaires. Au milieu des années 1980, les premiers liens entre le laboratoire et les établissements privés de la région (Ets Florimond-Desprez) se tissent par le biais de contrats de recherche, notamment sur le sujet de la stérilité mâle cytoplasmique. A la fin des années 1980, le phénomène d'embryogenèse somatique est observé sur l'hybride « 474 », provenant du croisement entre *Cichorium intybus* (variété Koospol, chicorée « industrielle ») et *Cichorium endivia* (variété Catalane, chicorée « scarole »). L'embryogenèse somatique directe a été mise en valeur d'abord sur les styles (Dubois *et al.*, 1988), puis les anthères (Guedira *et al.*, 1989), les racines (Guedira *et al.*, 1990) et enfin sur les feuilles (Dubois *et al.*, 1991). Plusieurs disciplines ou techniques biologiques se sont développées autour de cette thématique : cytologie (Blervacq *et al.*, 1995), physiologie (Vuylsteker *et al.*, 1997), biologie moléculaire (Galland *et al.*, 2001), donnant naissance ensuite à des travaux de doctorat en protéomique, transcriptomique ou encore en génétique.

Au fur et à mesure des années, d'autres sujets ayant pour modèle la chicorée se sont ajoutés : l'auto-incompatibilité (Gonthier *et al.*, 2010), l'amertume (Hance *et al.*, 2007) et enfin les antioxydants.

II. Matériels et méthodes

II.1. Obtention de la population à phénotyper

Dans le but de créer une descendance en ségrégation pour une étude génétique (analyses QTL) sur l'embryogenèse somatique, des individus phénotypiquement très différents ont été recherchés. C'est dans la population Koospol que deux individus au comportement très contrasté pour ce caractère ont été trouvés. Effectivement, l'individu K59 est très embryogène alors que l'individu K28 l'est très peu (capacité à produire des embryons somatiques en condition de culture *in vitro*). Le premier croisement a été réalisé en 2001 : le pollen de K59 (hermaphrodite) a été déposé manuellement sur les stigmates de K28 (mâle stérile). Ce croisement a donné une descendance F1' de 177 individus appelée sous-descendance F1'a. En 2004, un deuxième croisement à partir des mêmes parents a généré une nouvelle sous-descendance de 214 individus, nommée sous-descendance F1'b.

Une telle descendance, dont les individus sont en partie génotypés (cf. Chapitre II), représentait une opportunité à saisir pour l'étude génétique d'autres caractères quantitatifs. Dans le cadre du projet Oxychic créé en 2007 et issu de la collaboration entre le laboratoire SADV et les sociétés Florimond-Desprez et Leroux, il a été décidé de profiter de cette même descendance pour des études génétiques visant à identifier les loci impliqués, *via* une analyse QTL, dans la biosynthèse des principales molécules présentes dans les feuilles de chicorée, reconnues pour leur propriétés antioxydantes en tant que métabolites secondaires. Etant donné que le principe d'une analyse QTL s'appuie sur une variabilité phénotypique à associer à la variabilité génotypique, la première étape était de vérifier si cette population présentait une variabilité pour les différentes molécules détectées et l'AAR.

II.2. Obtention du matériel végétal

II.2.A. Obtention des plantes mères (octobre 2006)

Les descendants de K59 et K28 ont été bouturés pour constituer deux lots : un premier lot dédié au projet Oxychic et un second lot pour des recherches parallèles sur l'auto-

incompatibilité (figure 1). Le bouturage s'est effectué comme suit : des bourgeons axillaires sont prélevés en prenant un fragment de racine au niveau du collet. Le fragment de racine est trempé dans une poudre contenant des hormones de bouturage (Chryzotop vert composé de 0,25% d'acide β -indole butyrique, KB Substral, Allemagne) et planté dans un terreau de rempotage (composé de 45% de tourbe de sphaigne brune, 45% de tourbe de sphaigne blonde et 10% d'argile) mélangé à 1,5 kg d'engrais par mètre cube. Ces boutures sont placées en mini-serres à ouvrir au fur et à mesure du bon développement des plantules.

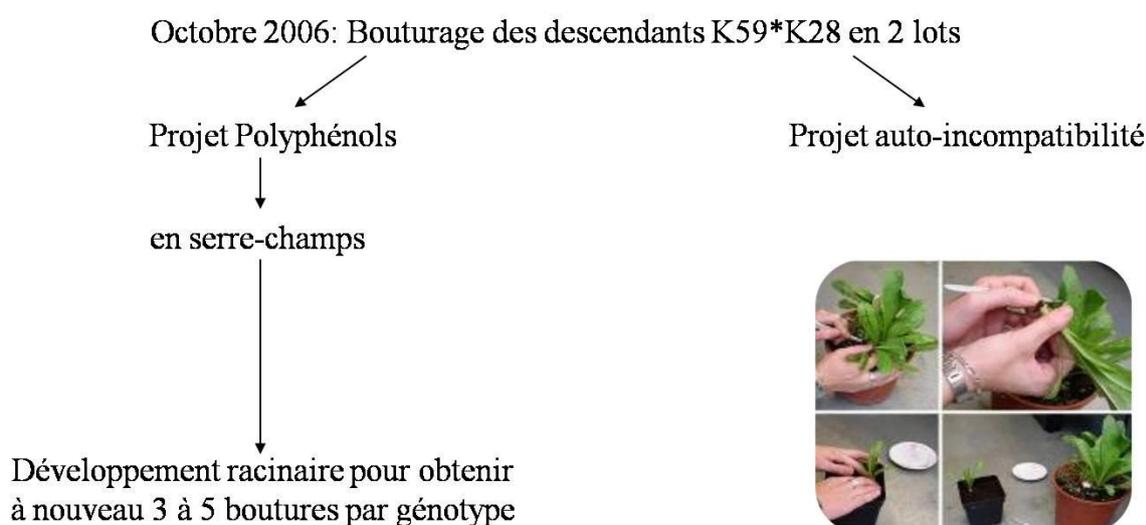


Figure 1. Origine du matériel végétal et illustration du bouturage des bourgeons axillaires à partir des plantes mères.

Les plantes destinées aux recherches sur l'auto-incompatibilité ont été disposées en fosse jusqu'à juin 2007 afin d'induire la montaison et la floraison. Les chicorées du projet Oxychic ont été placées en serre-champs (sans lumière artificielle et contrôle de la température par ventilation). Ces plantes serviront à un prochain bouturage dans le but d'obtenir cinq clones par génotype, c'est-à-dire les répétitions biologiques nécessaires à nos analyses. Pour cela, les chicorées doivent se développer suffisamment au niveau racinaire afin de produire des bourgeons axillaires.

Durant cette période de développement racinaire, les chicorées ont été déplacées à l'extérieur suite à des problèmes d'homogénéité de l'arrosage automatique (par le plafond) et de ventilation (mauvais développement des plantes près des ventilateurs). Afin d'éviter une hétérogénéité liée aux conditions de culture, les plantes sont restées à l'extérieur du mois de juin au mois de septembre 2007. Cependant, l'excès de pluie n'a pas favorisé le bon développement des plantes. Elles ont donc ensuite été à nouveau placées en serre-champs, le

plus loin possible de la ventilation pour éviter à nouveau des conditions de culture hétérogènes et avec un ajout d'engrais (engrais azote, phosphate, potassium, magnésium, soufre et oligo-éléments, "Nitrophoska bleu spécial", Compo GmbH & Co. KG). Dans ces conditions, la période de développement racinaire a duré un peu plus d'une année. La figure 2 illustre la lenteur du développement racinaire par rapport au développement foliaire : cinq mois après l'étape du bouturage, le diamètre de la racine n'excède pas 3 cm. La figure 3 est une photo prise à la même époque que la photo de la figure 2. Elle démontre l'influence de la position des plantes par rapport au ventilateur sur le développement foliaire.



Figure 2. Taille d'une racine de chicorée environ cinq mois après le bouturage à partir de bourgeons axillaires.



Figure 3. Effet de la ventilation sur le développement des plantes. A droite une plante placée près des ventilateurs ; à gauche une plante placée loin des ventilateurs.

II.2.B. Obtention des clones

II.2.B.a. 1^{er} bouturage

Pour procéder à une analyse QTL, il était nécessaire d'avoir plusieurs répétitions biologiques donc plusieurs clones par individu. A partir des plantes du projet Oxychic, une à six boutures par génotype sont réalisées en janvier 2008 comme expliqué dans le paragraphe II.2.A. Seules 30% de ces boutures ont survécu et ont été repotées début mars.

II.2.B.b. 2^{ème} bouturage

Suite à une perte importante des boutures précédentes, une deuxième session de bouturage est effectuée en mars 2008. Les parents ainsi que 218 descendants ont été bouturés. Au moins trois boutures par génotype pour les descendants et au moins cinq boutures pour les parents K59 et K28 ont été mises en place.

II.3. Constitution des blocs de clones

Après repotage (mai 2008), les descendants présentant moins de trois clones ont été conservés en « collection ». Les descendants ayant entre trois et cinq clones seront utilisés pour acquérir les données phénotypiques concernant l'AAR et les concentrations des différentes molécules identifiées ultérieurement. Afin de réduire la variance environnementale, les plantes ont été réparties en cinq blocs randomisés (de I à V). Ainsi, les blocs I, II et III sont de taille égale avec 207 plantes. Par contre, les blocs IV et V possèdent moins de plantes, 186 et 147 respectivement. Dans chaque bloc, les parents sont présents quatre fois. L'ordre des plantes (parents compris) est aléatoire. Ces cinq blocs sont disposés en serre-champs jusqu'à la date de prélèvement (juillet 2008). L'arrosage se fait automatiquement par le plafond, une fois par jour et manuellement si besoin est.



Figure 4. Disposition des cinq blocs (répétitions biologiques de la descendance et des parents) dans la serre-champs.

II.4. Protocole final : extraction et analyses des cinq blocs (954 chicorées)

II.4.A. Prélèvement et extraction

Lors du prélèvement, pour chaque plante, deux feuilles d'âge similaire sont prélevées (feuille α et feuille β). Ces plantes sont au stade végétatif et n'ont pas subi de vernalisation. Six disques sont perforés à l'aide d'un tube eppendorf sur chaque feuille. Différentes informations ont été relevées lors du prélèvement telles que la taille des feuilles (longueur et largeur) pouvant donner une idée de l'âge des feuilles ou de l'état général de la plante, la présence d'oïdium (nom usuel donné à certains champignons de la famille des Erysiphacées) susceptible d'augmenter la variabilité environnementale, la surface des feuilles (degré de « boursoufflures », pouvant influencer sur la quantité de matériel foliaire emporté par le tube eppendorf) et les positions de chaque plante dans les blocs et par rapport aux ventilateurs, au système de refroidissement (filet d'eau froide coulant le long d'un panneau) ou aux vitres de la serre-champs.

Ces rondelles sont mises à macérer à température ambiante à l'obscurité pendant deux semaines dans un mélange éthanol-acide acétique-eau (80/5/15 ; v/v/v ; VWR, West Chester, PA, USA). Les six rondelles de chicorée sont ensuite extraites du liquide de macération et séchées à l'étuve pour effectuer une mesure de la matière sèche (correction des données finales en divisant par le poids sec résiduel). Cinq cent microlitres de chloroforme (Merck, Darmstadt, Allemagne) puis 100 μ L d'eau ultrapure sont ajoutés au liquide de macération. Après agitation et centrifugation (2 minutes, 10000 rpm, 4°C), 150 μ L de la phase aqueuse supérieure sont transférés en microplaque (AB gene, Thermo scientific, UK). Les extraits sont ensuite séchés au speed-vac (RC1022 et piège à froid RCT90, Jouan, Thermo Electron Corporation, Courtaboeuf, France) et les résidus secs sont remis en solution dans 100 μ L de méthanol à 50% préparée dans de l'eau ultrapure (VWR, West Chester, PA, USA). Les 1908 extraits ainsi obtenus (descendance et parents x 5 blocs x 2 feuilles) sont conservés en microplaques recouvertes d'un film thermocollant (Thermo scientific, UK) et stockés à -20°C en attendant l'analyse par chromatographie liquide haute performance (CLHP).

II.4.B. Analyse des acides hydroxycinnamiques par CLHP

Les échantillons sont analysés grâce à l'appareil de CLHP LC-20AD de Shimadzu System (Kyoto, Japon). Cet appareil est équipé d'un détecteur UV-visible de type barrette de diodes SPD-M20A (Shimadzu System). Deux programmes différents ont été utilisés.

II.4.B.a. Programme long

Six microlitres d'échantillon sont injectés sur une colonne de type LiChrospher RP18 (phase inverse : silice greffée par des chaînes linéaires de 18 atomes de carbone) de 150 x 4,6 mm, 5 μ m (Phenomenex, Etat-Unis), précédée d'une pré-colonne de 4 mm x 3 mm. L'élution est réalisée en appliquant un gradient linéaire à la colonne chauffée à 25°C délivré avec un flux de 1 mL/minute. L'éluant A est composé d'eau et l'éluant B d'acétonitrile additionné de 2% d'acide acétique. Les solvants sont de qualité « CLHP » et proviennent de chez VWR (West Chester, PA, USA). Le gradient mis en place par Sébastien Grec est décrit dans le tableau 1. La durée totale d'une analyse est de 36 min et la colonne est équilibrée 5 minutes entre chaque injection.

II.4.B.b. Programme court

Les molécules sont séparées sur une colonne de type LiChrospher RP18-4 μ , 75 mm x 4 mm (Phenomenex, Etat-Unis). Six microlitres d'échantillon sont élués par un gradient binaire. Le solvant A est composé d'eau et d'acide acétique 1%. Le solvant B est un mélange d'acétonitrile et d'acide acétique 2% (colonne à 55°C ; flux de 1,8 mL/minute). Le gradient mis au point par Philippe Hance (projet Oxychic) est détaillé dans le tableau 1. La colonne est équilibrée 2,8 minutes entre chaque analyse.

Programme long (36 minutes)		Programme court (8 minutes)	
Temps (min)	solvant A (%)	Temps (min)	solvant A (%)
0	100	0	94
2	92	1.4	90
5	87	4.2	84.3
15	85	4.4	81
16	83.5	4.8	0
18	81.5	5.2	94
22	80		
26	0		
29	0		
31	100		

Tableau 1. Les gradients linéaires CLHP : programme long et programme court.

Les composés sont identifiés et quantifiés sur la base de leur temps de rétention et en référence à des solutions standards. Afin de confirmer l'identification de certains composés ou d'identifier certains composés inconnus, des analyses ont été effectuées ponctuellement par chromatographie liquide couplée à la spectrométrie de masse (CL/MS).

II.4.C. Analyse par CL/MS

Afin d'identifier les molécules présentes dans nos échantillons ou confirmer leur présence, différents standards chimiques ont été analysés par CL/MS : des acides phénoliques comme l'acide caféique, chicorique et chlorogénique ; des flavonoïdes comme le kaempférol et la quercétine ; l'acide ascorbique. Ces standards ont été dilués à la concentration de 0,1 mg/mL et 10 μ L sont injectés sur un système de CLHP Hewlett Packard HP 1100 liquid chromatography (Agilent, Palo Alto, USA). Les molécules sont séparées sur la même colonne que précédemment, à 24°C selon le gradient décrit dans le tableau 2.

Programme CL/SM	
Temps (min)	solvant A (%)
0	93
11	78
20	25
21	0
25	0
26	93
40	93

Tableau 2. Le gradient de la CLHP couplée au spectromètre de masse. Solvant A : eau et 1% acide acétique. Solvant B: acétonitrile et 2% acide acétique. Temps total d'analyse : 40 minutes. Flux : 1 mL par minute.

La CLHP est couplée au spectromètre de masse Quattro II (Waters Micromass, Manchester, Royaume-Uni). La source d'ionisation est de type électrospray. La source d'ionisation du spectromètre de masse, maintenue à 120°C, a été utilisée en mode négatif, excepté pour l'acide chlorogénique. La tension du capillaire est fixée à 3,5 kV et la tension du cône à 30 V. L'azote sert de gaz nébulisateur et de gaz de séchage. Les spectres de masse sont acquis sur une gamme de m/z de 240-700. Les spectres sont ensuite traités avec l'aide du logiciel MassLynx 4.0.

Les molécules inconnues, c'est-à-dire sans standard auquel les associer, sont fragmentées une seconde fois via une analyse en spectrométrie de masse en tandem. L'ion précurseur est fragmenté dans la cellule de collision à une énergie de 15 eV grâce à un gaz inerte (argon) avec lequel les ions entrent en collision.

II.4.D. Analyse de la capacité antiradicalaire globale

Le DPPH (1,1-diphényl-2-picryl-hydrazyl) est une molécule radicalaire libre stable. Un antioxydant va être capable de donner un électron et donc supprimer ce radical libre. La perte de ce radical entraîne un changement du spectre d'absorption du DPPH qui peut être suivi au spectrophotomètre via la couleur violette typique du DPPH qui se décolore au cours de la réaction.

La mesure de l'activité antiradicalaire par le test DPPH est une adaptation de la méthode décrite par Brand-Williams (1995). Le protocole a été miniaturisé au format microplaque étant donné le volume final de nos échantillons. Pour ce test, les échantillons

sont au minimum dilués trois fois (sinon saturation) et trois répétitions techniques par échantillon sont analysées. Le radical libre DPPH méthanolique à 100 µM est ensuite ajouté (Sigma-Aldrich, St-Louis, MO). La lecture de l'absorbance se fait à l'aide d'un lecteur de microplaques à 517 nm (Multiskan Spectrum, Thermo Electron Corporation, Courtaboeuf, France), après un temps de réaction exact de 30 minutes à l'obscurité. Le contrôle correspond à la moyenne de trois puits contenant uniquement du DPPH 100 µM. Le blanc correspond à la moyenne des absorbances de trois puits contenant seulement du méthanol. L'activité antiradicalaire est calculée comme suit :

$$\text{AAR}\% = [(\text{absorbance contrôle} - \text{absorbance échantillon}) / \text{absorbance contrôle}] \times 100$$

II.5. Traitement statistique

Les analyses de variance ont été effectuées grâce à la procédure de modèle linéaire généralisé (GLM) du logiciel Systat 12 version 12.00.08 (SPSS Inc., Chicago, USA) ou du logiciel SAS (Procédure GLM, SAS institute 1990), selon le modèle :

$$Y = \text{moyenne} + \text{bloc} + \text{génotype} + \text{erreur},$$

où « Y » est le caractère analysé, « moyenne » est une constante, tandis que la composante « bloc » rend compte des différences micro-environnementales entre les cinq blocs de l'expérimentation et « génotype » correspond à l'effet du fond génétique. Chaque effet, « bloc » et « génotype », a été traité comme fixe. Est aussi incluse dans cette équation la part d'erreur non expliquée par le modèle.

Les différentes données analysées correspondent (i) aux aires (non transformées) des pics des molécules enregistrées par la CLHP, (ii) à la proportion (comprise entre 0 et 1) des aires des pics de chaque molécule par rapport à l'ensemble du chromatogramme CLHP, proportion transformée en arcsinus de la racine carrée, (iii) aux pourcentages d'AAR obtenus d'après les tests DPPH, pourcentages transformés en arcsinus de la racine carrée. Les abréviations utilisées sont :

- ACAFT pour l'aire des pics correspondant à l'acide caftarique ;
- ACHLO pour l'aire des pics correspondant à l'acide chlorogénique ;
- ACHIC pour l'aire des pics correspondant à l'acide chicorique ;
- ATOT pour la somme des aires des pics correspondant à l'acide caftarique, chlorogénique et chicorique ;

- PCAFT pour la proportion de l'aire du pic correspondant à l'acide caftarique par rapport à l'aire totale des pics du chromatogramme ;
- PCHLO pour la proportion de l'aire du pic correspondant à l'acide chlorogénique par rapport à l'aire totale des pics du chromatogramme ;
- PCHIC pour la proportion de l'aire du pic correspondant à l'acide chicorique par rapport à l'aire totale des pics du chromatogramme ;
- PTOT pour la somme des proportions des aires des pics des acides caftarique, chlorogénique et chicorique, par rapport à l'aire totale des pics du chromatogramme ;
- AAR pour l'activité antiradicalaire.

Les héritabilités ont été calculées à partir des carrés moyens selon une formule adaptée de Lynch & Walsh (1998) : $H_L^2 = (CM_{\text{individu}} - CM_{\text{erreur}}) / (CM_{\text{individu}})$.

Les moyennes ajustées (LSmeans) par génotype ont été calculées en fonction du modèle statistique défini et des effets de chaque composante. Ces moyennes ajustées sont les valeurs à utiliser pour les analyses QTL futures.

III. Résultats

III.1. Mise en place d'un protocole pour le phénotypage

III.1.A. Préambule au protocole final

Avant d'obtenir le matériel végétal nécessaire à la détection de QTL impliqués dans la biosynthèse de molécules antioxydantes, nous avons choisi de mettre en place un protocole de prélèvement et d'extraction rapide et simple, adapté au nombre prévu de plantes. Un premier protocole avait été établi lors de mon stage de Master 2. Plusieurs paramètres avaient été étudiés comme le mode de broyage, le choix des solvants d'extraction et d'élimination de la chlorophylle, la durée de macération et la constitution de l'échantillon final.

Les feuilles complètes étaient prélevées, lavées, pesées (500 mg) puis broyées. Le type de broyage a été choisi en fonction de la rapidité d'exécution mais également selon le meilleur taux de reproductibilité. Le broyage manuel à l'aide d'un piston, dans un tube eppendorf, de fragments foliaires s'était révélé être le plus fastidieux (dix à quinze minutes pour un tube) et le plus susceptible d'induire de la variabilité. Le broyage automatique, quant à lui, présentait l'avantage d'être très rapide (environ 3 minutes pour 48 tubes) mais présentait des défauts de

reproductibilité : il suffisait que la bille se coince entre les feuilles lors de l'agitation des tubes eppendorf par la machine pour que le rendement du broyage varie d'un tube à l'autre, comme le montre la figure 5A. Une extraction à l'éthanol 80% a été répétée trois fois, dans trois tubes différents contenant le même matériel végétal, pour vérifier la reproductibilité de l'expérience. La figure 5A montre un écart type de 68,42% entre les différentes répétitions biologiques, expliquant le rejet de la méthode par broyage automatique. A l'inverse, le broyage à l'azote liquide a démontré peu de variabilité : l'écart type est de 3,9%. Pour un nombre d'échantillon restreint, c'est une méthode de broyage robuste.

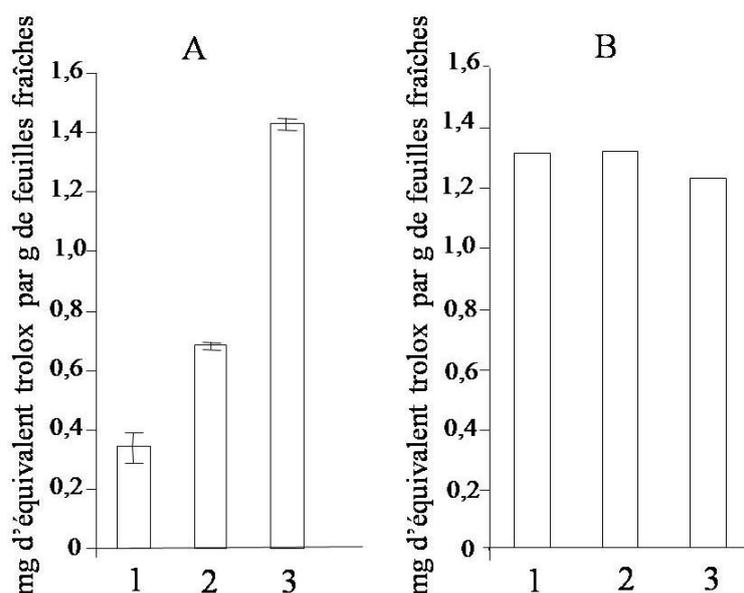


Figure 5. Evaluation de la variabilité du broyage A. automatique B. à l'azote liquide. Dosage de l'activité antiradicalaire par le test DPPH. Extraction à l'éthanol 80%. En ordonnée : mg d'équivalent trolox par gramme de feuilles fraîches. En abscisse, les répétitions biologiques. Les barres d'erreur correspondent aux répétitions techniques du test (invisibles quand trop faibles).

Ensuite, différents mélanges de solvants de macération ont été testés : éthanol/eau (4/1 : v/v), éthanol/acétate/eau (19/1/5 ; v/v/v), éther/éthanol/eau (10/9/6 ; v/v/v), acétone/éthanol/eau (7/7/3 ; v/v/v). Les deux derniers mélanges avaient déjà montré leur efficacité au sein de notre laboratoire pour l'extraction de molécules antioxydantes de la racine de chicorée et l'aubépine, respectivement. Les deux premiers mélanges sont couramment utilisés pour la chicorée dans la littérature, avec quelques variations dans les proportions (Innocenti *et al.*, 2005; Rossetto *et al.*, 2005; Heimler *et al.*, 2007; Milala *et al.*, 2009). Les mélanges éthanol/eau ou éthanol/acétate/eau se sont révélés jusqu'à quatre fois plus efficaces que les mélanges à base d'éther ou d'acétone, après dosage des échantillons selon trois tests chimiques classiquement utilisés : DPPH (pouvoir antiradicalaire global),

Pridham (dosage des ortho-diphénols) et Folin-Ciocalteu (dosage des polyphénols totaux), figure 6. La macération durait 16 heures, à 22°C, sous agitation, à l'obscurité. Pour chaque échantillon, le matériel biologique est le même.

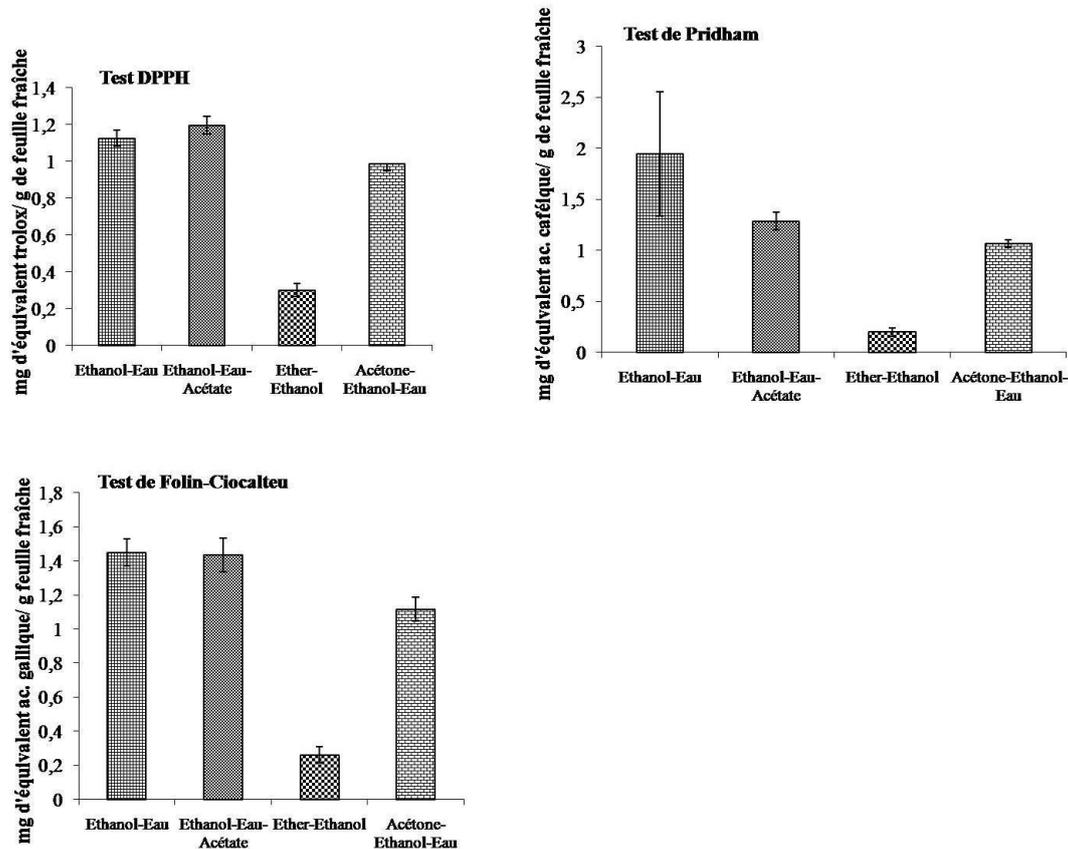


Figure 6. Comparaison de quatre solvants de macération pour l'extraction de polyphénols à partir de feuilles de chicorée. Résultats obtenus à partir du test DPPH, du dosage de Pridham et du dosage de Folin-Ciocalteu. Les barres d'erreur correspondent aux répétitions des extractions (3 fois) et aux répétitions techniques des tests (3 fois).

Pour ne pas endommager la colonne CLHP et ne pas interférer avec les tests colorimétriques, il était souhaitable d'éliminer la chlorophylle présente dans nos échantillons. Les quatre solvants testés étaient l'héxane (Innocenti *et al.*, 2005), le chloroforme (Mulinacci *et al.*, 2001), le cyclohexane, et le dichlorométhane. Le chloroforme est apparu comme étant le plus adapté à notre situation : les chlorophylles sont éliminées tout en conservant au maximum les molécules polyphénoliques et le pouvoir antioxydant (tests DPPH et Folin-Ciocalteu).

Après séchage au speed-vac, l'échantillon final est resolubilisé et conservé dans de l'eau et du méthanol (1/1 ; v/v).

III.1.B. Mise en place du protocole final

En partant du protocole précédent, nous avons souhaité miniaturiser davantage et simplifier les étapes de prélèvement et d'extraction et ainsi adapter le procédé à un nombre important de plantes à phénotyper dans un temps le plus court possible.

III.1.B.a. Prélèvement

La première étape nous permettant de gagner du temps est la pesée des feuilles. L'utilisation du tube eppendorf comme une perforeuse, permet le prélèvement de rondelles de feuille de 0,636 cm² (donc 3,82 cm² pour six rondelles). Un prélèvement de 6 rondelles a permis de passer de 500 mg de feuille fraîche à 70 mg environ, tout en conservant un profil correct à la CLHP et une réponse suffisante aux tests colorimétriques cités dans le paragraphe III.1. Les premiers essais étaient effectués en perforant trois feuilles d'une même plante. Des tests statistiques (ANOVA) ont mis en évidence l'absence de différence significative (valeur de $p > 0,05$) entre ces trois feuilles concernant les concentrations en molécules d'intérêt et l'activité antiradicalaire (tableau 3). Les manipulations suivantes ont alors été effectuées à partir de deux feuilles par plante, d'âge présumé similaire (même niveau d'implantation par rapport au centre).

Caractère	Source	ddl	Carré Moyen	F	p
acide chicorique	Feuille	286	0.00008943	0,25	0,999 ^{ns}
	Erreur	6	0.00035126		
acide chlorogénique	Feuille	226	0.00016579	1,67	0,270 ^{ns}
	Erreur	6	0.00009947		
AAR	Feuille	290	2.95263995	0,40	0,976 ^{ns}
	Erreur	6	7.31578447		

Tableau 3. Résultat de la procédure GLM sur les données obtenues pour 96 individus x 3 feuilles. Les caractères testés sont la concentration en acide chicorique, la concentration en acide chlorogénique, l'AAR. *** = $p < 0,001$ = différence significative entre les trois feuilles d'un même individu ; ns = différence non significative entre les trois feuilles d'un même individu. F de Snedecor = variance inter-groupes / variance intra-groupes, plus F est important, plus l'effet significatif est fort ; ddl = degré de liberté.

III.1.B.b. Extraction

Le broyage à l'azote liquide est la méthode classique utilisée pour de nombreux végétaux et pour différentes parties de ces végétaux (Kliebenstein *et al.*, 2001; Coinu *et al.*,

2007; Stevens *et al.*, 2007; Bunea *et al.*, 2008). Etant difficile à mettre en place sur le terrain, dans une serre ou tout autre lieu non équipé du matériel adéquat, d'autres méthodes ont été essayées comme le broyage manuel de rondelles fraîches ou sèches à l'aide d'un piston directement dans le tube eppendorf. Cette méthode restant fastidieuse et donc source de variabilités techniques, une méthode plus passive a été envisagée. Il s'agit de laisser macérer les rondelles de chicorée sans détériorer les tissus dans un mélange de solvants choisi au préalable (éthanol-acétate-eau, 19/1/5 ; v/v/v). Cette technique, analogue aux techniques de fixation en cytologie, inactive les protéines, stoppant ainsi toute réaction enzymatique. Comparée à la méthode de référence et montrant un résultat similaire (cf. annexe 1, A « fresh » protocol for the analysis of hydroxycinnamic acids in leaf tissue of chicory, soumis le 10/09/2010 à Journal of Agricultural and Food Chemistry), l'étape du broyage a été remplacée par une extraction par macération.

Cette méthode a été testée à deux températures différentes correspondant à la température ambiante (22°C) et à celle de la chambre froide (4°C). Les profils CLHP obtenus après extraction des molécules à ces deux températures se superposent (cf. annexe 1). Afin de connaître la durée minimale nécessaire à l'extraction des molécules d'intérêt et le temps maximal pour éviter la dégradation des molécules, plusieurs durées de macération ont été testées : 24, 48, 72 et 96 heures. Aucune différence qualitative n'a été observée (figure 7). Au regard du nombre de plantes à analyser, des temps plus long de macération ont été comparés (une et deux semaine(s)). Encore une fois, les profils chromatographiques étaient similaires.

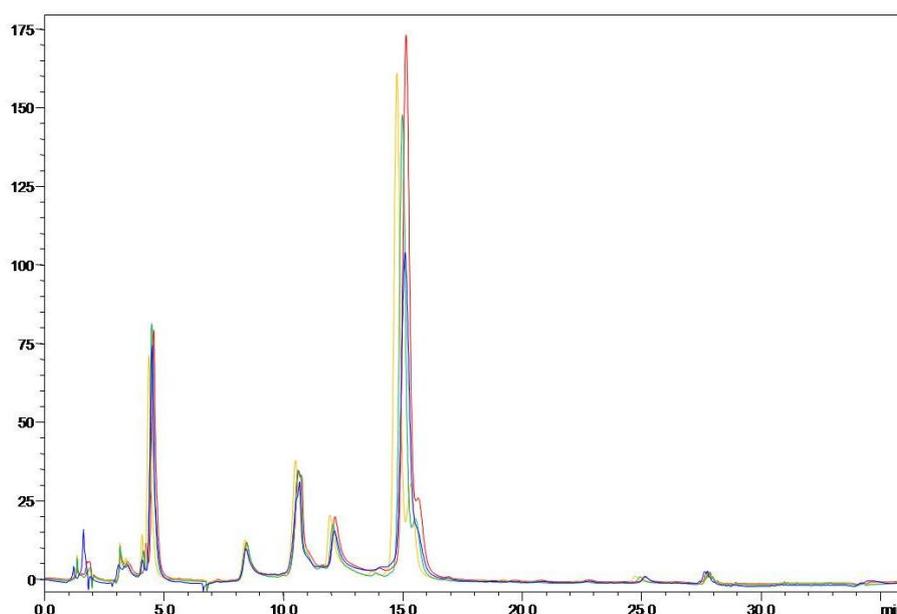


Figure 7. Superposition des quatre profils chromatographiques obtenus après avoir fait macérer six rondelles de feuille de chicorée dans un mélange d'éthanol, acétate, eau (76%/4%/20%) pendant 24h (profil bleu), 48h (profil rouge), 72h (profil vert), 96h (profil orange).

III.1.B.c. Après l'extraction

Après la phase d'extraction, les rondelles sont retirées et mises à sécher naturellement ou à l'étuve. Ces rondelles sont pesées pour rapporter au poids sec résiduel les données obtenues après phénotypage, l'épaisseur des rondelles prélevées pouvant varier de manière significative d'une feuille à l'autre bien que la surface de prélèvement reste la même (cf. paragraphe III.2.). En effet, la perte de masse liée à la macération est apparue proportionnelle au poids frais (perte supplémentaire de $2,9\pm 1\%$ par rapport au poids sec sans macération).

Suite aux expériences préliminaires décrites dans le paragraphe III.1.A, le chloroforme est le solvant choisi pour éliminer les chlorophylles. Il est mis en contact avec la phase aqueuse pendant une minute. La phase aqueuse est alors prélevée et déposée en microplaque, permettant un gain de place par rapport à un stockage en tube eppendorf. Le contenu des microplaques est séché puis resolubilisé dans 50% de méthanol, seule une forte proportion d'eau (réduite ici au maximum) pouvant dissoudre les résidus secs provenant de la phase aqueuse. La conservation à -20°C et dans une telle proportion de méthanol permet de garder les échantillons intacts pendant une longue période comme le montre des dosages répétés plusieurs mois après l'analyse initiale (cf. annexe 1).

III.1.B.d. Implication des manipulateurs

Afin de tester la robustesse de ce protocole d'extraction, quatre étudiants de l'Université de Lille 1 n'ayant aucun lien avec ce travail ont suivi ce protocole (cf. annexe 1). Cette expérience n'a montré aucune différence significative (valeur de $p > 0.05$) entre les échantillons issus d'une même feuille de chicorée manipulée par 4 personnes différentes et inexpérimentées.

III.1.C. Adaptation des outils de phénotypages

Les premiers tests d'outils allant servir au phénotypage de la descendance K28 x K59 ont été effectués à partir d'un programme CLHP de 36 minutes pour la détection des molécules d'intérêt et d'un test DPPH en cuvettes pour la mesure du pouvoir antiradicalaire global.

Ces deux étapes ont ensuite été adaptées dans la volonté d'une miniaturisation globale et d'une réduction de temps d'exécution du prélèvement/extraction et du phénotypage. Ainsi, nous sommes passés d'un gradient linéaire de 36 minutes sur une colonne de 150 x 4,6 mm, diamètre 5 μm , à un nouveau gradient linéaire de 8 minutes sur une colonne de 75 x 4 mm, diamètre 4 μm , en gardant un profil résolutif (cf. détails annexe 1), divisant ainsi les temps d'analyse CLHP par 4,5. Le test DPPH a été transféré en microplaque ce qui permet un gain de temps non négligeable au niveau de la manipulation du test lui-même mais aussi de la lecture des résultats, grâce à un lecteur de microplaque et non plus cuvette par cuvette (cf. II. Matériels et Méthodes).

III.1.D. Identification des molécules d'intérêt

Dans un premier temps, deux pics observés dans les profils CLHP ont été alloués aux acides chlorogénique et chicorique en comparant leur temps de rétention et leur spectre d'absorption UV avec ceux des standards correspondants. Leur identité a pu être confirmée par spectrométrie de masse. Les masses moléculaires $[\text{M}+\text{H}]^+$ de ces deux dérivés d'acide caféique sont de m/z 355 et m/z 473 pour l'acide chlorogénique et l'acide chicorique, respectivement (figure 5).

Un troisième pic non identifié était également présent sur les profils LC. L'ion moléculaire de cette troisième molécule a une masse de m/z 311. La fragmentation de cet ion négatif par SM/SM et la comparaison des différentes masses obtenues par rapport à celles déjà rapportées dans la littérature (Mulinacci *et al.*, 2001; Papetti *et al.*, 2008) ont permis son identification. Ainsi, les analyses de spectrométrie en tandem ont révélé la présence de trois fragments majoritaires. Les fragments à m/z 149 et 135 correspondent à l'acide caféique et sa décarboxylation. Le troisième signal détecté à m/z 179 a été attribué à l'acide tartarique. La reconstitution de ce spectre de masse a permis d'identifier l'acide caftarique (figure 8). L'absence d'un ion m/z 623, correspondant au dimère d'acide caftarique, typiquement trouvé avec la forme *trans* du l'acide caftarique, suggère qu'il s'agit de l'isomère *cis*. Un chromatogramme type est présenté dans l'annexe 1 (Papetti *et al.*, 2008).

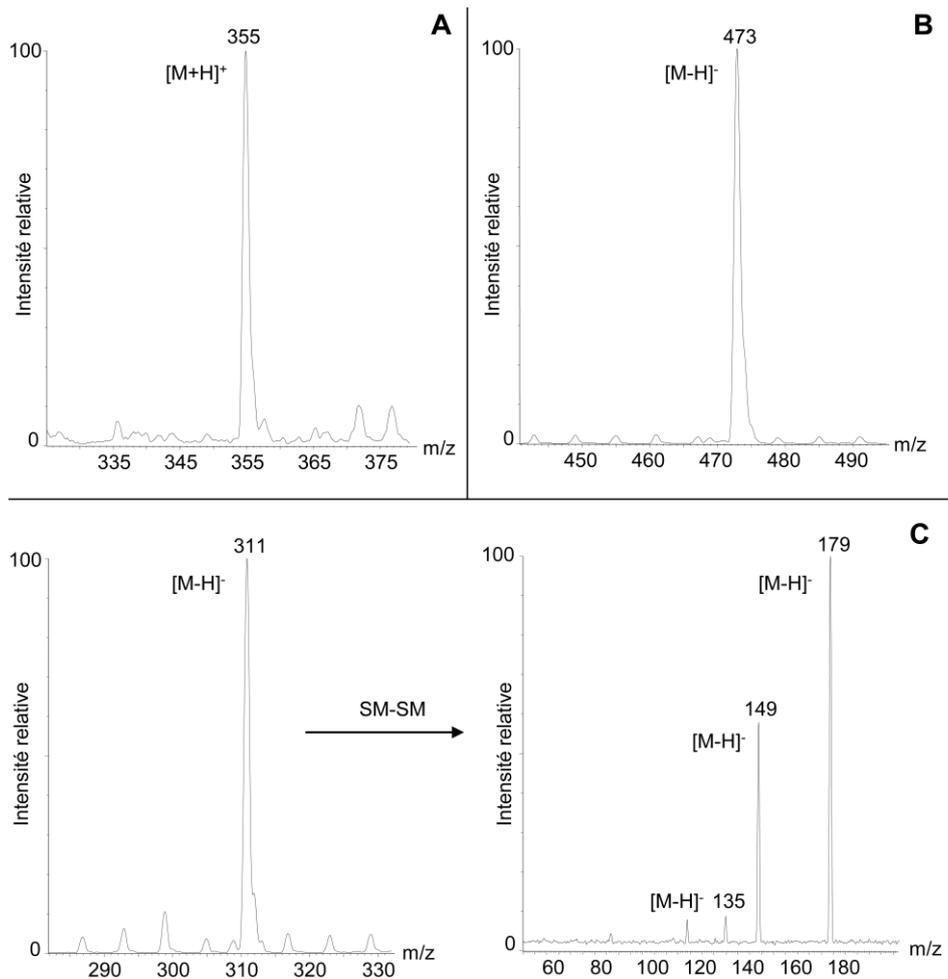


Figure 8. (A) Spectre de masse de l'acide chlorogénique obtenu après ionisation positive. (B) Spectre de masse de l'acide chicorique obtenu après ionisation négative. (C) Spectre de masse de l'acide caftarique obtenu après ionisation négative par CL-SM et SM-SM.

III.2. Essais préliminaires sur les parents et la descendance

En attendant la mise en place des cinq blocs de chicorées pour le prélèvement final, c'est-à-dire celui qui servira à la détection de QTL, des essais préliminaires ont été effectués sur les parents et sur une partie des descendants disponibles à ce moment (pas de clones pour les descendants).

Ces deux lots de plantes ont été prélevés à des moments différents (96 puis 192 chicorées, cf. figure 9). Le premier essai a été effectué sur des plantes encore au stade végétatif, sortant d'une période de vernalisation artificielle en fosse pouvant déclencher la montaison et la floraison. Le deuxième essai a été fait sur des chicorées au stade végétatif n'ayant pas subi de vernalisation. Les échantillons obtenus par le protocole miniaturisé ont été évalués pour leur pouvoir antioxydant (DPPH) et leur contenu en molécules d'intérêt

analysées par CLHP avec le programme 36 minutes dans un premier temps puis 8 minutes plus tard, quand celui-ci a été mis au point.

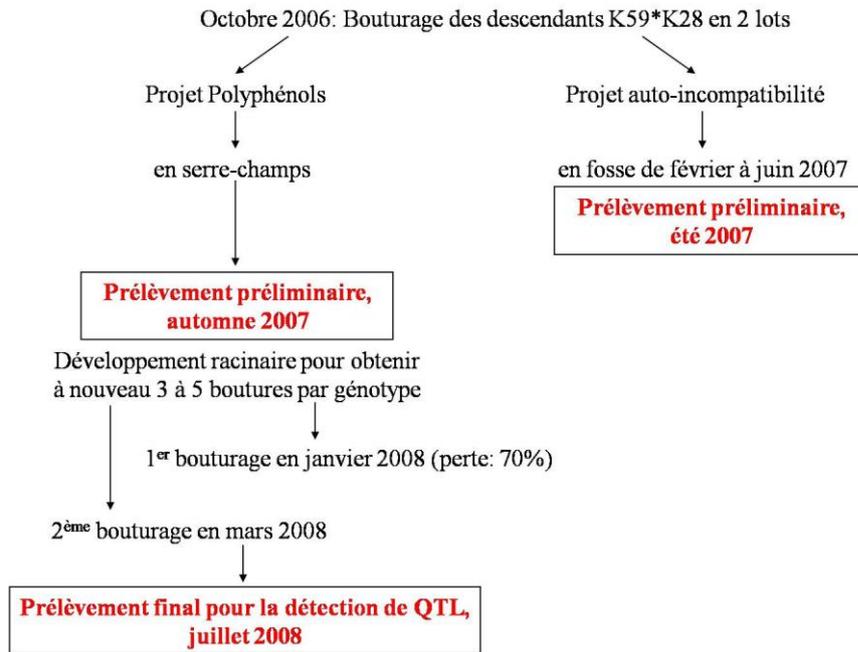


Figure 9. Schéma illustrant l'obtention du matériel végétal.

D'une manière globale, ces premiers essais nous ont permis de constater que les parents ne sont pas significativement différents (valeur de $p > 0,005$) pour les concentrations en acides chlorogénique et chicorique (l'identification de l'acide caftarique a été faite ultérieurement). Par contre, une différence significative a été trouvée pour l'AAR entre les deux parents (valeur de $p < 0,005$, parent K28 montrant une valeur AAR plus forte).

Pour les descendants, on observe une distribution normale des données relevées. Les analyses de variance ont indiqué une différence significative ($p < 0,001$) entre les génotypes étudiés, pour l'activité antiradicalaire, les concentrations en acides chicorique et chlorogénique, et le poids résiduel des rondelles sèches (tableau 4). On peut noter un très fort effet du poids résiduel ($F = 31,84$), indiquant qu'il est impératif de prendre en compte ce facteur. Ces résultats préliminaires suggèrent, d'une part, une variabilité phénotypique des descendants transgressant celle des parents et d'autre part, la possibilité de moyenner les données obtenues pour les feuilles d'un même individu.

	Caractère	Source	ddl	Carré Moyen	F	p
1 ^{er} essai préliminaire	acide chicorique	Individu	86	0.00037420	9,29	***
		Erreur	146	0.00004030		
	acide chlorogénique	Individu	96	0.00020084	4,68	***
		Erreur	196	0.00004288		
	AAR	Individu	96	7.51901447	8,43	***
		Erreur	200	0.89167451		
poids résiduel	Individu	96	5.67236900	31,84	***	
	Erreur	200	0.17817500			
2 ^{ème} essai préliminaire	acide chicorique	Individu	191	0.00245508	8,17	***
		Erreur	274	0.00030057		
	acide chlorogénique	Individu	189	0.00010493	5,36	***
		Erreur	268	0.00001959		
	AAR	Individu	191	19.36097899	6,53	***
		Erreur	274	2.96674751		

Tableau 4. Résultat de la procédure GLM sur les données obtenues après un essai préliminaire sur 96 individus et un second essai sur 192 individus. Les caractères testés sont la concentration en acide chicorique, la concentration en acide chlorogénique, l'AAR et le poids résiduel. *** = $p < 0,001$ = différence significative entre les individus ; ns = différence non significative. F de Snedecor = variance inter-groupes / variance intra-groupes, plus F est important, plus l'effet significatif est fort ; ddl = degré de liberté.

De fortes corrélations ont été constatées (figure 10), entre la concentration en acide chicorique dans les extraits et l'AAR (r entre 0,77 et 0,92) ainsi qu'entre la concentration en acide chlorogénique dans les extraits et l'AAR (r entre 0,76 et 0,87).

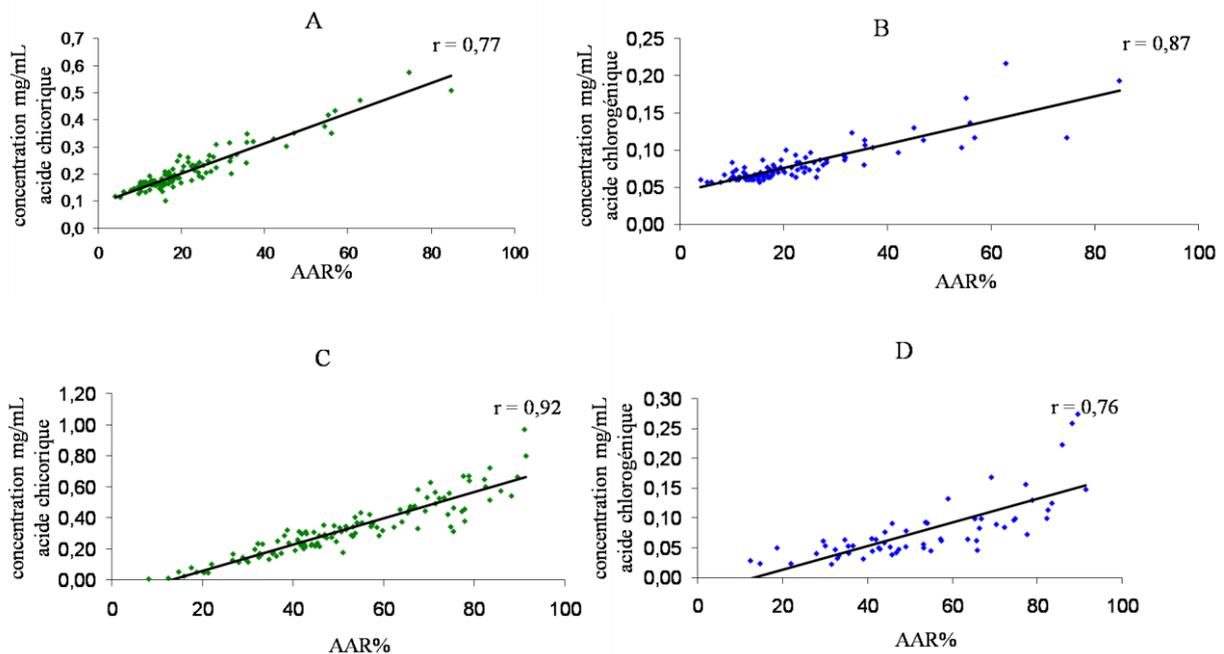


Figure 10. A et B : résultats des corrélations entre l'AAR et (A) la concentration en acide chicorique (mg/mL) et (B) en acide chlorogénique (mg/mL) pour le premier essai préliminaire avec 96 géotypes. C et D : résultats des corrélations entre l'AAR et (C) la concentration en acide chicorique (mg/mL) et (D) en acide chlorogénique (mg/mL) pour le second essai préliminaire, avec 192 géotypes.

III.3. Phénotypage pour la détection de QTL

III.3.A. Données CLHP

III.3.A.a. Obtention des données CLHP

Pour la détection de QTL impliqués dans la biosynthèse des molécules antioxydantes précédemment identifiées, 1908 échantillons ont été analysés : 207 génotypes ayant entre trois et cinq clones \times deux prélèvements par plante. Les profils chromatographiques obtenus à 320 nm avec le programme court d'élution ont permis la quantification des trois acides hydroxycinnamiques identifiés précédemment via la spectrométrie de masse : l'acide caftarique, l'acide chlorogénique et l'acide chicorique. Chacun des profils est analysé manuellement un à un : élimination des profils de mauvaise qualité, ajustement de la ligne de base et correction de l'intégration des trois pics d'intérêt (pic par pic).

Certains profils CLHP des plantes du bloc I se sont révélés différents des autres. En effet, environ la première moitié des chromatogrammes correspondant au bloc I présente une très bonne résolution du pic d'acide chicorique, suivi d'un pic inconnu. Les chromatogrammes des autres échantillons (deuxième moitié du bloc I + blocs II, III, IV et V) présentent un pic large d'acide chicorique, semblant co-éluer avec le pic inconnu (figure 11).

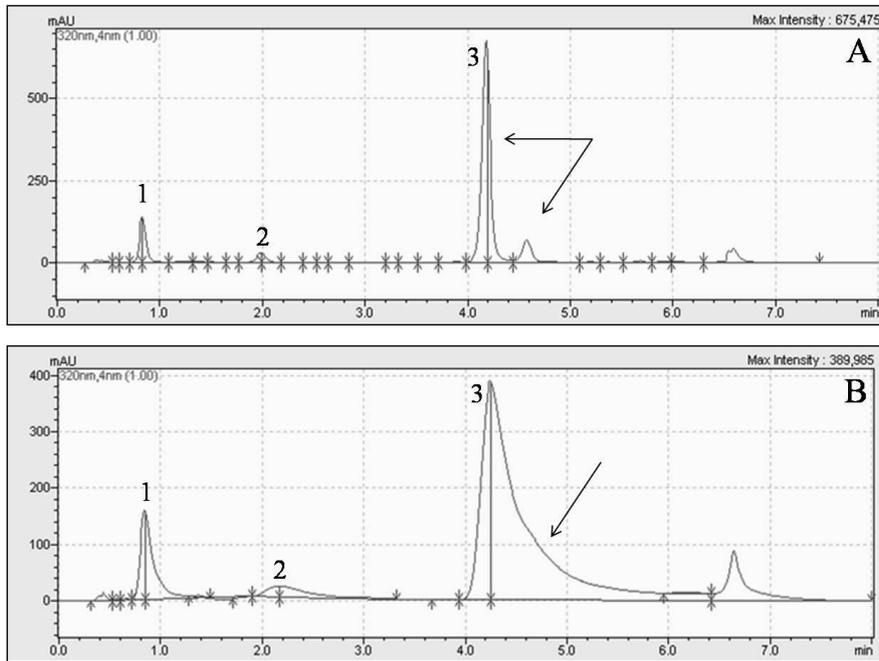


Figure 11. A) Profil CLHP typique de la moitié des plantes du bloc I. B) Profil CLHP typique de la seconde moitié des plantes du bloc I, et des plantes des blocs II, III, IV et V. 1 : pic d'acide caftarique ; 2 : pic d'acide chlorogénique ; 3 : pic d'acide chicorique. Notons la différence d'échelle et qu'il ne s'agit pas du même individu.

Ainsi, l'aire des pics a été volontairement coupée au milieu, c'est-à-dire au sommet du pic. C'est l'aire à gauche de cette limite qui a été prise en compte (figure 12). L'aire (unité arbitraire) de chacun de ces demi-pics ainsi que la proportion des demi-aires (comprise entre 0 et 1) par rapport au reste du profil chromatographique constituent les premières données phénotypiques. Ces données ont été corrigées par le poids sec résiduel obtenu pour les six rondelles de chaque échantillon. Les deux prélèvements par plante ont été moyennés (feuille α et feuille β).

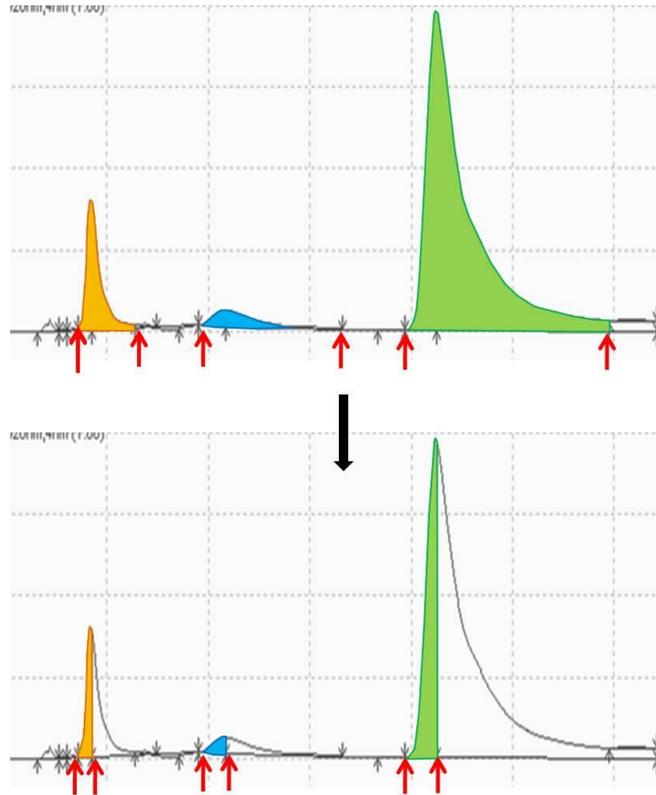


Figure 12. Retraitement des profils CLHP (320 nm). En orange : pic de l'acide caftarique ; en bleu : pic de l'acide chlorogénique ; en vert : pic de l'acide chicorique. Les flèches rouges correspondent aux limites de l'intégration, modifiées manuellement.

III.3.A.b. Analyses statistiques des données CLHP

Pour les trois molécules d'intérêt, il y a deux catégories de données phénotypiques résultant de l'analyse CLHP : l'aire absolue des pics et la proportion de l'aire des pics par rapport à l'ensemble des aires des pics du profil.

Les données aberrantes ont été éliminées : données supérieures et inférieures à la moyenne plus ou moins deux fois l'écart type. Les données PCAFT, PCHLO et PCHICO ont été transformées en $\arcsin\sqrt{\cdot}$ car il s'agit de valeurs comprises entre 0 et 1 (Alonso-Blanco *et al.*, 2003). Les données ACAFT, ACHLO et ACHIC sont conservées telles quelles et n'ont pas été transformées en \log_{10} car ce type de transformation n'a pas amélioré la normalité de la distribution et l'homoscédasticité des résidus par rapport aux valeurs estimées par le modèle (figure 13). Les données ACAFT, ACHLO et ACHIC et les données PCAFT, PCHLO et PCHIC ont été additionnées sous le nom de ATOT et PTOT, respectivement, pour savoir s'il n'y a pas un effet de la somme des trois principaux hydroxycinnamates. Pour chaque lot de données, l'obtention d'une distribution normale et continue suggère que ceux-ci sont sous la dépendance de plusieurs gènes.

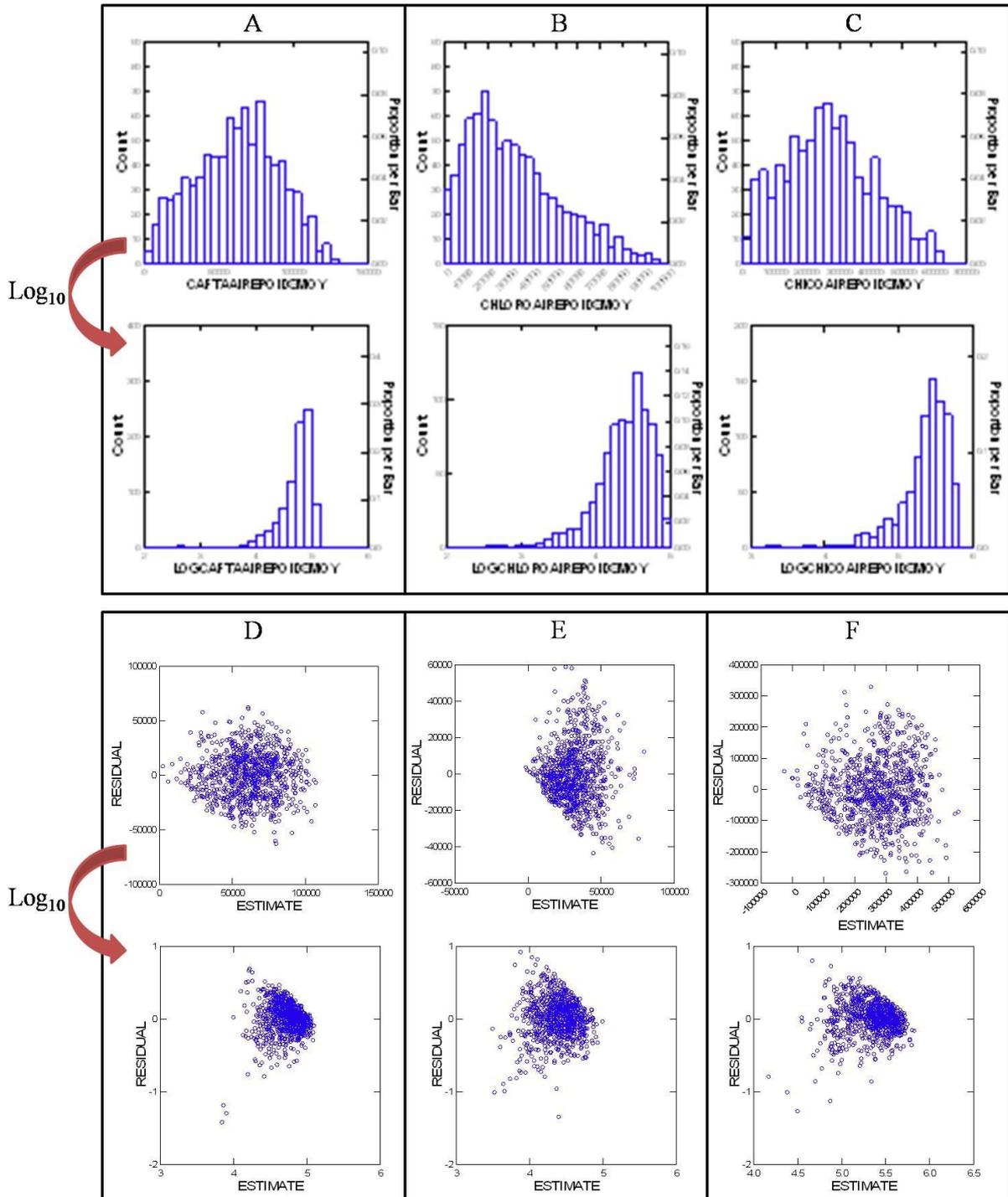


Figure 13. A, B, C : distribution des données CLHP correspondant à l'aire de chacun des pics avant et après transformation en Log_{10} . A : données ACAFT. B : données ACHLO. C : données ACHIC. D, E, F : répartition des résidus du modèle par rapport aux valeurs prédites par le modèle. D : données ACAFT. E : données ACHLO. F : données ACHIC.

La procédure GLM effectuée par le logiciel Systat donne les résultats exposés dans le tableau 5. On peut noter que pour la majorité des données, la valeur de p est suffisamment faible ($<0,001$) pour que l'on puisse considérer qu'il y ait un effet individu. Autrement dit, les valeurs moyennes de certains individus, en particulier les individus à valeurs extrêmes, sont

significativement différentes. A contrario, les valeurs de p de PCHIC et de PTOT sont trop élevées ($p > 0,05$) pour que l'on puisse considérer un effet significatif des individus.

La part de variance expliquée par l'effet individu est déterminée en calculant la valeur de F associée à cet effet. On remarque que pour le caractère « individu », la valeur F reste faible alors qu'elle est très importante pour la co-variable « bloc ». Ceci montre l'importance de l'environnement dans ce type d'expérimentation. Pour la détection de QTL, les données « individu » seront ajustées en prenant en compte les effets de chaque composante du modèle statistique (cf. paragraphe II.5).

L'héritabilité au sens large est utilisée pour estimer la part du génotype (G) qui contribue à l'expression du phénotype (P), ce dernier étant le résultat du génotype et de l'environnement (E) pour un caractère donné soit $P = G + E$. L'héritabilité h_L^2 est toujours comprise entre 0 et 1. L'héritabilité la plus faible (0,296) est recensée pour l'aire des pics des trois molécules (ATOT) et la plus élevée (0,531) pour la proportion en acide chlorogénique (PCHLO) (tableau 3).

Caractère	Source	ddl	Carré Moyen	F	p	H _L ²
ACAFT	Bloc	4	2,83E+10	49,643	***	
	Individu	199	9,67E+08	1,697	***	0,411
	Erreur	655	5,70E+08			
ACHLO	Bloc	4	8,96E+09	25,707	***	
	Individu	199	5,34E+08	1,533	***	0,348
	Erreur	647	3,49E+08			
ACHIC	Bloc	4	1,14E+12	86,342	***	
	Individu	199	2,00E+10	1,512	***	0,339
	Erreur	657	1,33E+10			
ATOT	Bloc	4	1,65E+12	74,693	***	
	Individu	199	3,14E+10	1,421	***	0,296
	Erreur	669	2,21E+10			
PCAFT	Bloc	4	0,099482	109,700	***	
	Individu	199	0,001786	1,969	***	0,492
	Erreur	672	0,000907			
PCHLO	Bloc	4	0,060025	63,155	***	
	Individu	199	0,002025	2,130	***	0,531
	Erreur	670	0,00095			
PCHIC	Bloc	4	0,826579	596,394	***	
	Individu	199	0,001549	1,118	0,1574 ^{ns}	x
	Erreur	670	0,001386			
PTOT	Bloc	4	9454,719	782,137	***	
	Individu	199	11,972497	0,990	0,525337 ^{ns}	x
	Erreur	672	12,088318			

Tableau 5. Résultat de la procédure GLM sur les données obtenues après analyse des profils CLHP. Plus la valeur de p est basse, plus l'hypothèse H_0 « les moyennes obtenues pour chaque individu ou bloc ne diffèrent pas l'une de l'autre » est rejetée. *** = $p < 0,001$; ns = différence non significative. F de Snedecor = variance inter-groupes / variance intra-groupes, plus F est important, plus l'effet significatif est fort ; ddl = degré de liberté ; H_L² = héritabilité.

Pour les parents, on note une différence significative pour certains caractères : ACHLO ($p < 0,05$), PCHLO ($p < 0,001$) et PCAFT ($p < 0,05$). C'est, dans chaque cas, le parent K59 qui montre les valeurs les plus fortes. Pour le reste des caractères, la valeur de p est supérieure à 0,05.

Ensuite, pour l'analyse QTL, les données par individu sont moyennées et ajustées en tenant compte de l'effet bloc (données « LSmeans » du logiciel Systat). Cette étape n'est pas effectuée pour les données PCHIC et PTOT puisqu'aucune différence significative n'a été détectée. La figure 14 montre la distribution des différents phénotypes. On remarque un phénomène de transgression : les valeurs obtenues pour la descendance ne sont pas comprises entre celles des parents mais de part et d'autre (valeurs inférieures ou supérieures).

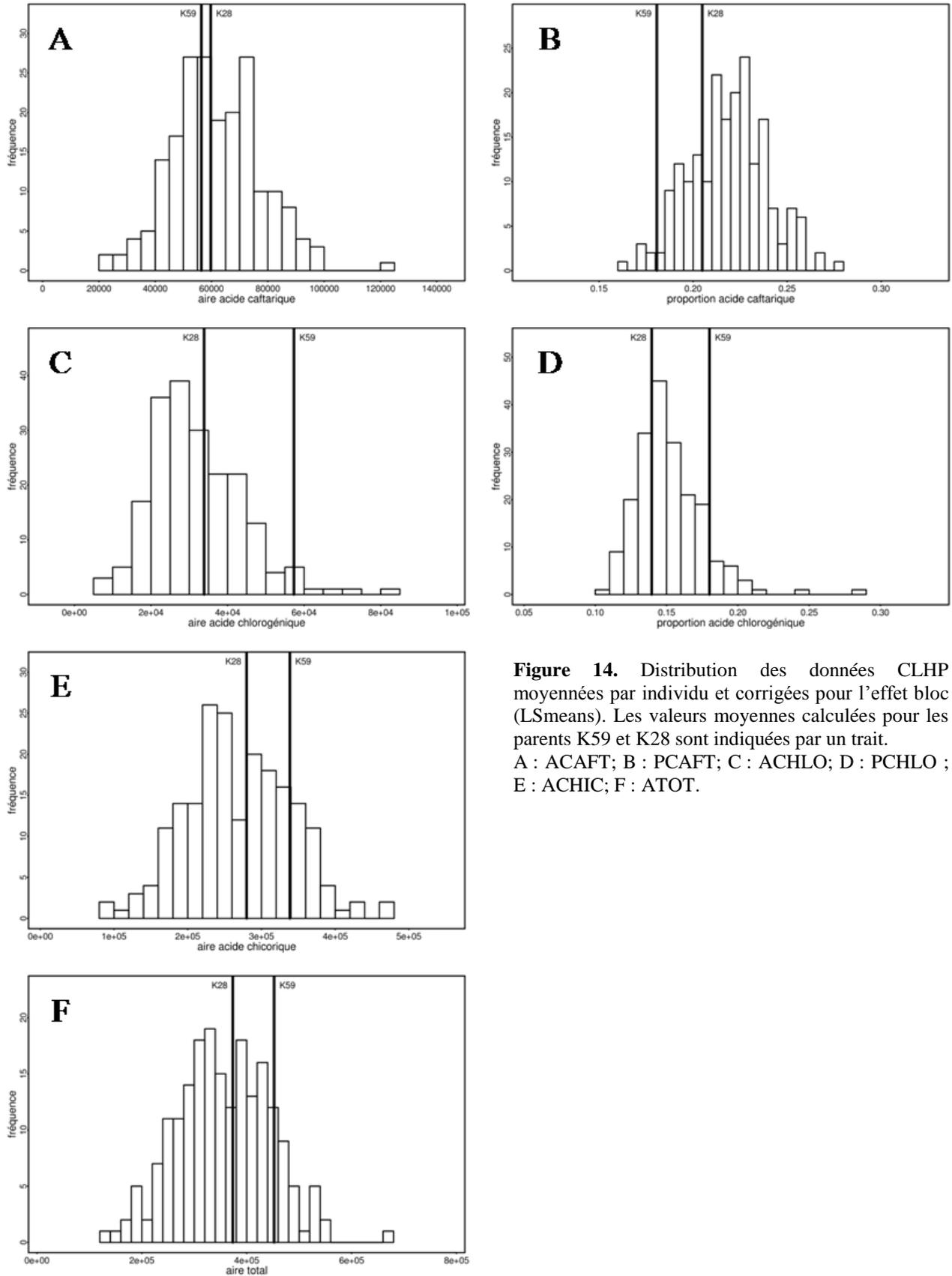


Figure 14. Distribution des données CLHP moyennées par individu et corrigées pour l'effet bloc (LSmeans). Les valeurs moyennes calculées pour les parents K59 et K28 sont indiquées par un trait.
 A : ACAFT; B : PCAFT; C : ACHLO; D : PCHLO ; E : ACHIC; F : ATOT.

III.3.B. Données activité antiradicalaire (AAR) et analyses statistiques

Les données de l'AAR ont été obtenues comme expliqué dans la partie Matériels et Méthodes. Ces données ont été corrigées par le poids résiduel des six rondelles de chicorées prélevées et transformées en $\arcsin\sqrt{\cdot}$. La distribution de ces données, normale et continue, suggère que ce caractère serait également sous la dépendance de plusieurs gènes. Le tableau 6 informe des résultats de l'analyse de variance ainsi que de l'héritabilité.

Caractère	Source	ddl	Carré Moyen	F	p	H _L ²
AAR	Bloc	4	0,017159	15,882	***	
	Individu	199	0,002304	2,133	***	0,531
	Erreur	676	0,00108			

Tableau 6. Résultat de la procédure GLM sur les données AAR. *** = $p < 0,001$; ns = différence non significative ; ddl = degré de liberté ; H_L² = héritabilité.

La figure 15 représente la distribution des valeurs moyennes pour l'AAR, corrigées en fonction de l'effet bloc dont la valeur de F est très supérieure à celle des individus. Les parents ne sont pas significativement différents ($p > 0,05$), contrairement à ce qui a été observé lors des essais préliminaires. A l'image des valeurs CLHP, on observe des descendants transgressifs : les valeurs obtenues pour les descendants excèdent les valeurs parentales de manière positive ou négative.

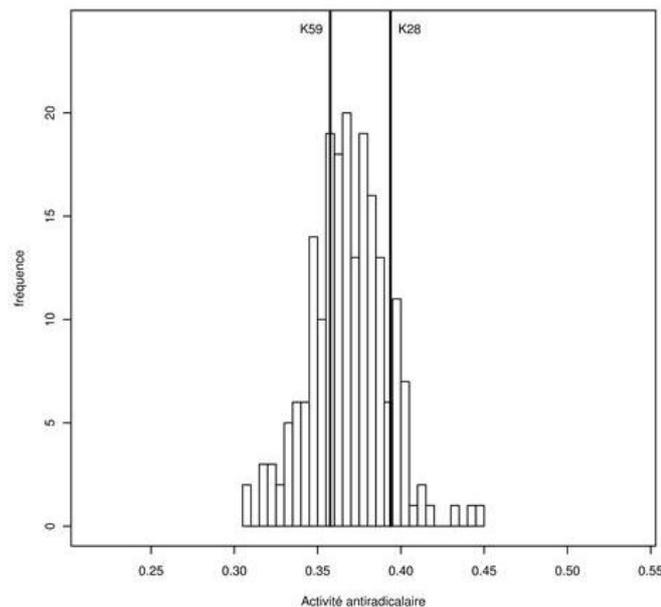


Figure 15. Distribution des données AAR moyennées par individu et corrigées pour l'effet bloc. Les valeurs moyennes calculées pour les parents K59 et K28 sont indiquées par un trait.

III.3.C. Corrélation entre les différentes données

Des corrélations de Pearson ont été effectuées entre les différents lots de données (tableau 7). Certaines se sont révélées non significatives (ns) comme celles entre AAR et ACHLO ou AAR et PTOT. On retrouve une assez bonne corrélation entre les aires des différentes molécules (r entre 0,55 et 0,72). ATOT est très influencé (90%) par la quantité d'acide chicorique puisque l'on retrouve une corrélation de $r = 0,95$ entre ces deux données. ACAFT explique environ 12% de la variation de l'activité antiradicalaire ($r = 0,35$) alors que ACHIC en explique $\sim 17\%$ ($r = 0,41$). Par contre la quantité d'acide chlorogénique ne semble pas influencer l'activité antiradicalaire puisque la corrélation n'est pas significative. Les proportions des différents hydroxycinnamates sont corrélées faiblement et négativement entre elles. PCHIC explique $\sim 8,5\%$ de la variation de l'activité antiradicalaire ($r = 0,29$) alors que c'est une corrélation négative que l'on peut observer entre AAR et les données PCAFT et PCHLO. Ces informations sont discutées dans le paragraphe IV.3.C.a de ce chapitre.

	ACAFT	ACHLO	ACHIC	ATOT	PCAFT	PCHLO	PCHIC	PTOT
ACHLO	+0,55***							
ACHIC	+0,72***	+0,64***						
ATOT	+0,79***	+0,71***	+0,95***					
PCAFT	+0,3***	n.i.	n.i.	n.i.				
PCHLO	n.i.	+0,66***	n.i.	n.i.	-0,16*			
PCHIC	n.i.	n.i.	+0,56***	n.i.	-0,23**	-0,23**		
PTOT	n.i.	n.i.	n.i.	+0,43***	n.i.	n.i.	n.i.	
AAR	+0,35***	ns	+0,41***	+0,39***	-0,16*	-0,2**	+0,29**	ns

Tableau 7. Corrélation entre les différents sets de données CLHP et AAR. Les symboles*, ** et *** indiquent si les données sont significativement corrélées à $p < 0,05$, $p < 0,01$ et $p < 0,001$, respectivement ; ns = corrélation non significative ; n.i = non informatif.

IV. Discussion

IV.1. Protocole final

Les différentes étapes du protocole utilisé pour le phénotypage sont illustrées dans la figure 16. Les résultats détaillés dans le paragraphe III.1.B et l'annexe 1, permettent de moduler le protocole en fonction des conditions de prélèvement, du nombre de plantes à analyser et du nombre de personnes impliquées tout en évitant la dégradation rapide des molécules, relatée antérieurement par Nüsslein (Nüsslein *et al.*, 2000). Ce type de protocole à petite échelle ne correspond pas à ce que l'on retrouve habituellement dans la littérature concernant le prélèvement et l'extraction des parties aériennes de chicorée, de laitue ou d'autres légumes-feuilles. La quantité de matière fraîche prélevée est généralement de l'ordre de plusieurs grammes, le volume de l'échantillon final correspond en général à 1 ou 2 mL tandis que les différentes étapes d'extraction et les programmes CLHP utilisés sont particulièrement chronophages. Le tableau 8 donne un aperçu des pratiques employées. Le travail du Kliebenstein *et al.* (2001) sur *Arabidopsis* indique une échelle de poids semblable à la nôtre ainsi que l'utilisation de rondelles de feuilles.

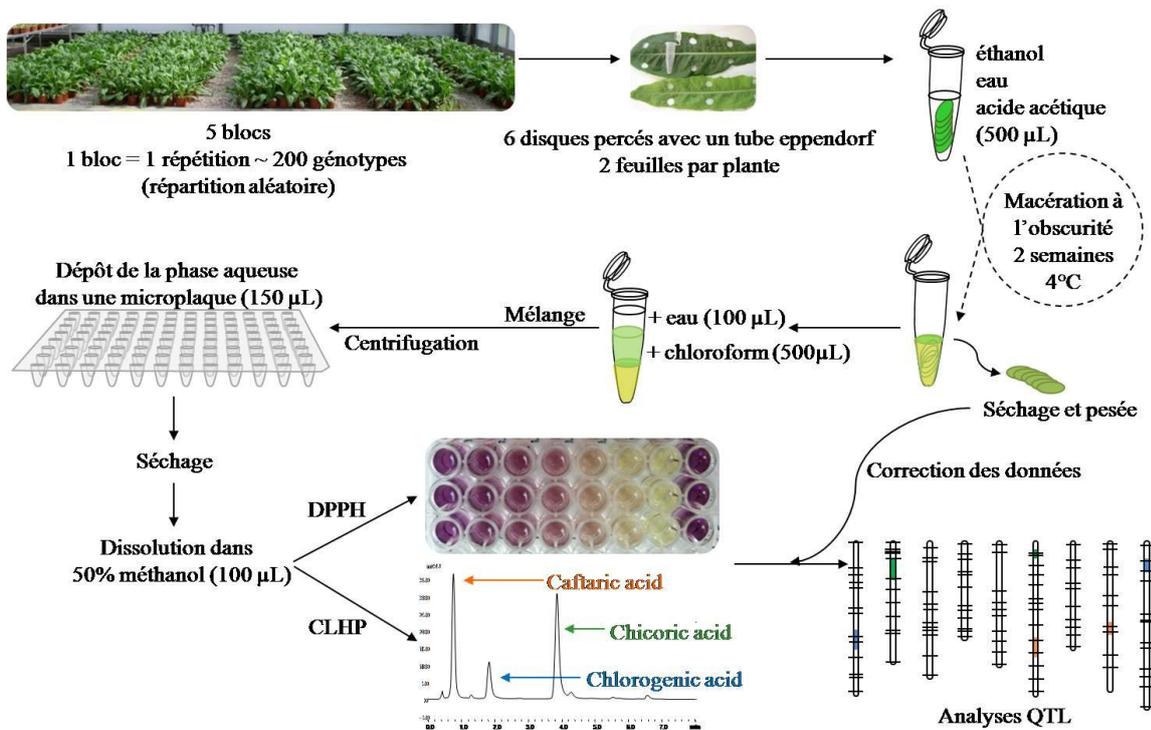


Figure 16. Protocole mis en place pour le phénotypage

CHAPITRE I

Domaine de recherche	Plante	Partie de la plante	Molécules d'intérêts	Quantité de matière (grammes)	Etapes du processus d'extraction	Volume échantillon final (mL)	Durée analyse CLHP (minutes)	Auteur
Caractérisation biochimique	Chicorée	feuille	polyphénols pouvoir antioxydant	1300	-extraction dans 10,4 L, 4 fois pendant 5 jours -distillation -concentration -lyophilisation	?	45	Milala, 2009
Caractérisation biochimique	Chicorée	feuille	polyphénols	20	-lavage -découpage -extraction dans 12,5 mL de méthanol et acide formique pendant 1h dans un bain glacé -extraction recommencé 3 fois pour les résidus insolubles -filtration	50	90	Papetti, 2008
Caractérisation biochimique	Chicorée	feuille	polyphénols	500 (mf)	-broyage à l'azote liquide -extraction dans 60 mL d'eau et d'éthanol, 3 fois -extraction à l'hexane -concentration -dissolution dans de l'eau et de l'éthanol	1-2	57	Innocenti, 2005
Caractérisation biochimique	Laitue Roquette Chicorée Plantain	feuille	polyphénols pouvoir antioxydant	2 (ms)	-broyage à l'azote liquide -extraction dans 30 mL d'éthanol 70% toute une nuit -concentration -rinçage à l'acétonitrile, méthanol et eau -extraction à l'hexane	2	30	Heimler, 2007
Caractérisation biochimique	Laitue Chicorée	feuille	polyphénols pouvoir antioxydant	1000 (mf)	-extraction à chaud (méthanol ou eau) pendant 1h -filtration -séchage -ajout d'eau -lyophilisation	?	60	Llorach, 2004

CHAPITRE I

Domaine de recherche	Plante	Partie de la plante	Molécules d'intérêts	Quantité de matière (grammes)	Etapes du processus d'extraction	Volume échantillon final (mL)	Durée analyse CLHP (minutes)	Auteur
Caractérisation biochimique	Epinard	feuille	polyphénols pouvoir antioxydant	1 (ms)	-lavage -lyophilisation -extraction dans 50 mL de méthanol, eau et hydrochlorate, pendant 1h -filtration -récupération du surnageant et à nouveau extraction de la phase inférieure, 3 fois -concentration -hydrolyse dans un bain-marie à 35°C -extraction 4-5 fois avec du diéthyl ether -séchage	10	25	Bunea, 2008
Caractérisation biochimique	Artichaut	feuille	polyphénols pouvoir antioxydant	10-30 (ms)	-lyophilisation -extraction dans 300 mL d'éthanol 7% -rinçage à l'hexane -séchage -fractionnement à l'acétate d'éthyl -séchage -dissolution -lyophilisation pour conservation	5	50	Coinu, 2007
Caractérisation biochimique	Amaranthe	feuille	pouvoir antioxydant	~330 (mf)	-lavage -découpage -lyophilisation -broyage -extraction à l'eau pendant 1h -filtration	?	x	Amin, 2006
Détection de QTL	Tomate	fruit	acide ascorbique	5 (ms)	-broyage à l'azote liquide -extraction dans 50 mL	?	x	Stevens, 2007
Détection de QTL	Arabette	feuille	glucosinolate	0,01 (ms)	-lyophilisation des feuilles -broyage automatique -extraction dans 410 µL de méthanol et acétate -60 minutes d'incubation -filtration -ajout de 300 µL d'eau -élimination de l'eau et séchage -répéter les deux dernières étapes	0,3	29	Kliebenstein, 2001et 2005

Tableau 8. Exemple de protocole de prélèvement-extraction et de programme CLHP. Abréviations : ms = masse sèche, mf = masse fraîche.

Cependant, les protocoles habituellement utilisés pour la chicorée ont une ambition différente de la nôtre : leur but est généralement l'identification de molécules polyphénoliques de la façon la plus exhaustive possible, éventuellement pour ensuite comparer plusieurs variétés de chicorées (Innocenti *et al.*, 2005; Rossetto *et al.*, 2005) entre elles ou plusieurs légumes-feuilles entre eux (Llorach *et al.*, 2004; Heimler *et al.*, 2007). Ces investigations n'exigent donc pas un nombre d'échantillons aussi important que le demandent des études génétiques mais nécessitent en revanche une extraction la plus complète possible (extraction répétée plusieurs fois, fractionnement, etc.).

Dans le cadre de caractérisations biochimiques, les analyses CLHP peuvent être complétées par un ou plusieurs tests chimiques déterminant la capacité antioxydante. Il existe de nombreux tests de ce genre, le terme d'antioxydation étant associé à plusieurs phénomènes. On peut ainsi retrouver des tests mesurant la capacité à piéger les radicaux peroxy (Rossetto *et al.*, 2005), ou les radicaux libres avec les tests DPPH et ABTS (2,2'-azinobis-(3-ethylbenzothiazoline-6-sulfonic acid)) ou encore la capacité à réduire le fer (FRAP, ferric reducing antioxidant power) (Llorach *et al.*, 2004). Ces expériences peuvent aussi être accompagnées de tests ciblant une classe de molécules comme c'est le cas pour le test de Pridham, qui dose les ortho-diphénols. Milala et ses collaborateurs (2009) ont utilisé les tests DPPH et Folin-Ciocalteu (phénols totaux) tandis que Heimler *et al.* (2007) ont rajouté à ces deux tests un troisième, permettant de juger de la capacité à piéger les ions Fe^{2+} .

Dans le contexte d'une recherche de loci associés à la synthèse des molécules antioxydantes ou de métabolites secondaires, les outils déployés varient d'une étude à l'autre. L'équipe de Rousseaux (2005, étude sur la tomate) a utilisé les tests TRAP (total radical-trapping potential) et Folin-Ciocalteu en complément des analyses CLHP. En 2009, Jin *et al.* ont publié la détection de neuf QTL basés sur les tests ABTS, Folin-Ciocalteu et un dosage colorimétrique des flavonoïdes, sans utiliser de CLHP (étude sur les grains de riz). Il en est de même pour Fontaine *et al.* (2003, étude sur les chaumes de maïs) qui suggèrent dix QTL impliqués dans la biosynthèse des esters d'acides p-coumarique et férulique, en s'appuyant exclusivement sur des tests biochimiques. A l'inverse, Dong et ses collaborateurs (2005, étude sur des pailles de riz) n'ont pas eu recours aux tests biochimiques mais uniquement aux données CLHP et CPG (chromatographie en phase gazeuse), à l'instar de Szalma en 2002 (étude sur les soies du maïs), qui a utilisé la CLHP uniquement (cf. tableau 3 du chapitre analyse bibliographique). Les analyses CLHP sont plus précises que les tests biochimiques si l'on veut mettre en évidence des loci impliqués dans la synthèse de molécules particulières. Cependant, certains tests peuvent leur être complémentaires voire redondants (Folin-

Ciocalteu, dosage des phénols totaux) tandis que d'autres rendent compte d'une bioactivité telle que la capacité antiradicalaire (DPPH, ABTS et TRAP) ou la capacité à réduire certains métaux (FRAP). Ce choix dépend donc de ce que l'on souhaite démontrer. Dans notre cas, les tests mesurant la capacité à piéger les ions métalliques auraient ainsi pu compléter notre analyse phénotypique.

Le protocole « fresh » (cf. annexe 1), simple à mettre en place et économique, pourrait intéresser le domaine industriel : il permettrait d'obtenir une réponse rapide concernant les principaux acides phénoliques, molécules présentes dans les feuilles de tous les cultigrupes de chicorées (cf. tableau 2 du chapitre analyse bibliographique). Il pourrait également être miniaturisé davantage, robotisé et adapté à d'autres légumes-feuilles comme la laitue ou l'épinard, par exemple. Dans le cas précis de la chicorée industrielle, les feuilles sont coupées et laissées dans le champ, lors de la récolte des racines. Ces feuilles, servant d'engrais vert mais aussi sujet au pourrissement (Llorach *et al.*, 2004), peuvent être à l'origine de problèmes phytohygiéniques (*Botrytis cinerea*, *Sclerotinia sclerotiorum*) et environnementaux (dégagement de méthane). Une des alternatives à cette situation serait de valoriser ces déchets de feuilles en tant que source de molécules antioxydantes.

IV.2. Identification des molécules majoritaires

Trois pics ont été identifiés comme étant l'acide caftarique, l'acide chlorogénique et l'acide chicorique, dans l'ordre d'élution. Leur identification a d'abord été effectuée en comparant leur temps de rétention et leur spectre d'absorption UV à celui de standards puis confirmée par CL/SM. L'acide caftarique a été identifié grâce à la masse des ions obtenus après fragmentation en SM/SM, en comparaison à ce qui est reporté dans les publications. Pour compléter ces données et conclure définitivement sur l'identification de ce pic, l'analyse d'un standard d'acide caftarique en CL/MS serait nécessaire. Le spectrogramme obtenu après fragmentation pourrait être comparé à celui obtenu dans ce travail. Les ions dont la masse m/z est inférieure à 135 et l'absence de l'ion parent sont dus à une force de fragmentation trop importante (figure 8).

Ces résultats révélant l'acide chicorique comme le composé phénolique le plus abondant, concordent avec les résultats déjà publiés sur *Cichorium intybus* L. (Mulinacci *et al.*, 2001; Innocenti *et al.*, 2005; Heimler *et al.*, 2007) ou *C. endivia* (Llorach *et al.*, 2004). Au

cours des différentes expériences précédant le prélèvement final, nous avons observé des variations au niveau de la concentration absolue des molécules mais l'acide chicorique reste toujours le composé majeur, quelque soit l'individu analysé, l'âge de la plante ou la saison de prélèvement.

Si l'on compare nos données à celles de Rees et Harbone (1985), qui sont les seuls - à notre connaissance - à avoir étudié les feuilles de chicorées industrielles, la présence constante de l'acide chicorique est également relevée, non seulement dans les feuilles mais aussi dans les racines. Par contre, l'acide caféique n'a pas été détecté dans nos échantillons. En se reportant au tableau 2 de l'analyse bibliographique, la présence de l'acide caféique est également constatée chez les chicorées « salade » comme Chioggia, Grumolo, Castelfranco et Verona (Rossetto *et al.*, 2005). Néanmoins, d'autres cultigrupes n'ont révélé que des esters d'acide caféique (estérification avec l'acide tartarique ou quinique), à l'instar de notre bilan (Rees & Harborne, 1985; Mulinacci *et al.*, 2001; Llorach *et al.*, 2004; Innocenti *et al.*, 2005; Heimler *et al.*, 2007; Papetti *et al.*, 2008).

Dans la littérature, d'autres acides phénoliques sont mentionnés : les acides gallique et protocatéchique ainsi que d'autres dérivés d'acide caféoylquinique ou tartarique sont retrouvés couramment dans plusieurs variétés de chicorées (cf. tableau 2 du chapitre analyse bibliographique). Concernant les flavonoïdes, la caractérisation chimique des feuilles de chicorées "salade" sans pigmentation rouge (Mulinacci *et al.*, 2001; Innocenti *et al.*, 2005; Heimler *et al.*, 2007) et de chicorées industrielles (Rees & Harborne, 1985), indique la présence de kaempférol, de quercétine, de lutéoline, d'apigénine, et d'isorhamnétine tous couramment liés à une molécule de glucose ou à un glucuronide. Les standards de kaempférol, quercétine, ascorbate, acide coumarique et acide protocatéchique ont été analysés par CL/SM mais aucune des ces molécules n'a été retrouvée dans l'échantillon de feuille de chicorée analysé en spectrométrie de masse. Les analyses devraient être complétées par plusieurs échantillons de plantes cultivées dans différentes conditions et à différents stades développementaux afin d'élargir la panoplie de molécules détectées et confirmer ou non les résultats mentionnés.

IV.3. Phénotypage pour la détection de QTL

IV.3.A. Obtention des données CLHP

En fonction des analyses CLHP et CL/SM, nous avons choisi de concentrer notre recherche sur les trois principaux polyphénols retrouvés dans les feuilles de chicorée industrielle. Ce choix n'est pas irréversible puisque chaque profil chromatographique est enregistré à toutes les longueurs d'ondes ce qui permet une exploitation ultérieure des données. Ainsi, les sesquiterpènes lactones (détectés aux alentours de 210 nm) et certains flavonoïdes (~ 350 nm), composés présents dans les feuilles de chicorée industrielle (Rees & Harborne, 1985) pourraient être les sujets d'une étude analogue et complémentaire à celle de cette thèse.

Les données CLHP préliminaires étaient rapportées en concentration (mg/mL) selon une courbe étalon obtenue à partir de standards des trois molécules validées en CL/SM. Cependant, lorsque les concentrations retrouvées dans nos échantillons sont faibles, cette méthode se révèle imparfaite, certains pics n'étant pas détectés si le chromatogramme est légèrement décalé. A l'instar de Clé *et al.* (2008), les valeurs des aires des pics correspondant aux molécules étudiées ont alors été relevées (unité arbitraire). Les quantités de molécules obtenues varient selon l'âge de la plante et les conditions de culture, les valeurs recensées lors des différentes expériences reflètent les contenus en métabolites secondaires à l'instant t du prélèvement.

Concernant les profils chromatographiques obtenus après le prélèvement des différents blocs de chicorées (détection de QTL), la différence entre les échantillons de la moitié du bloc I et les échantillons des autres blocs, reste difficilement explicable. En effet, cette différence ne correspond pas à un bloc en entier, à un événement précis ou à un jour en particulier (les prélèvements ont duré deux semaines).

Le choix d'enregistrer les aires des demi-pics (première moitié) se justifie par le fait (i) que le commencement du pic soit facilement définissable, contrairement à la fin du pic, (ii) que la largeur de la fin du pic semble correspondre à une saturation de la colonne et/ou à une co-élution avec une autre molécule, pouvant alors fausser les valeurs. Cette molécule co-éluant avec l'acide chicorique pourrait correspondre à l'acide méso-di-*O*-caféoyltartarique,

comme le suggèrent les chromatogrammes présentés par Llorach et ses collaborateurs (Llorach *et al.*, 2008). Dans le cas d'une dégradation, l'absence d'un témoin interne est regrettable car, ajouté dès les premières étapes du protocole de prélèvement/extraction, il aurait pu être un bon indicateur d'une éventuelle altération des molécules.

IV.3.B. Essais analyses statistiques

Constatant une différence dans les profils CLHP pour une partie des plantes du bloc I, les tests statistiques ont été recommencés en enlevant ces données. Cela n'a pas engendré une augmentation de la valeur de F ou une hausse de l'héritabilité. D'autre part, différentes informations ont été notées lors du prélèvement telles que l'état général de la plante, la taille des feuilles, la présence d'oïdium etc. (cf. paragraphe II.3). Ajoutées comme co-variables, elles n'ont pas eu d'effet sur les résultats de l'analyse statistique. Les analyses de variance ont donc été effectuées avec les données totales, c'est-à-dire les cinq blocs, imputées des valeurs aberrantes (moyenne \pm deux fois l'écart type).

Dans le tableau 5, on note qu'il n'y a pas de différence significative entre les individus pour le caractère PCHIC. En observant les profils chromatographiques, on constate que l'aire de l'acide chicorique représente en moyenne ~ 40 % de l'aire totale, ~ 5 % pour l'acide chlorogénique et ~ 10 % pour l'acide caftarique. Une variation de l'aire de l'acide chicorique fait donc varier l'aire totale de façon importante. En conséquence, la proportion de l'aire de l'acide chicorique par rapport au total change relativement peu. L'absence de différence significative au sein de la population pour la variable PTOT découle de l'explication précédente : l'acide chicorique étant la molécule la plus abondante, la valeur de p des données PTOT est logiquement associée à celle de PCHIC.

IV.3.C. Résultats des analyses statistiques

IV.3.C.a. Corrélations

A partir des profils CLHP, deux lots de données ont été recensés : l'aire des (demi-) pics et la proportion de l'aire des (demi-)pics par rapport à l'ensemble du chromatogramme. Tandis que les aires correspondant au pic des molécules reflètent la quantité de molécules présentes dans les feuilles au moment du prélèvement, la proportion renvoie plutôt à la co-

régulation des différentes molécules : existe-t-il un phénomène de balancier entre les différentes molécules étudiées, autrement dit, est-ce que les proportions des molécules augmentent ensemble ou la baisse de l'une engendre-t-elle la hausse de l'autre?

D'un côté, les corrélations positives entre les aires des différentes molécules indiquent qu'un individu au phénotype « fort » posséderait une concentration élevée des trois molécules d'intérêt ; d'autre part, les corrélations négatives (bien que faible) entre les proportions des différentes molécules suggèreraient que la baisse de synthèse d'une molécule profiterait à la synthèse d'une autre et vice versa (figure 17) et qu'il existerait donc peut-être une compétition entre ces voies de biosynthèse. La figure 18 schématise la compétition supposée entre chaque molécule d'intérêt par rapport à leurs précurseurs.

CHAPITRE I

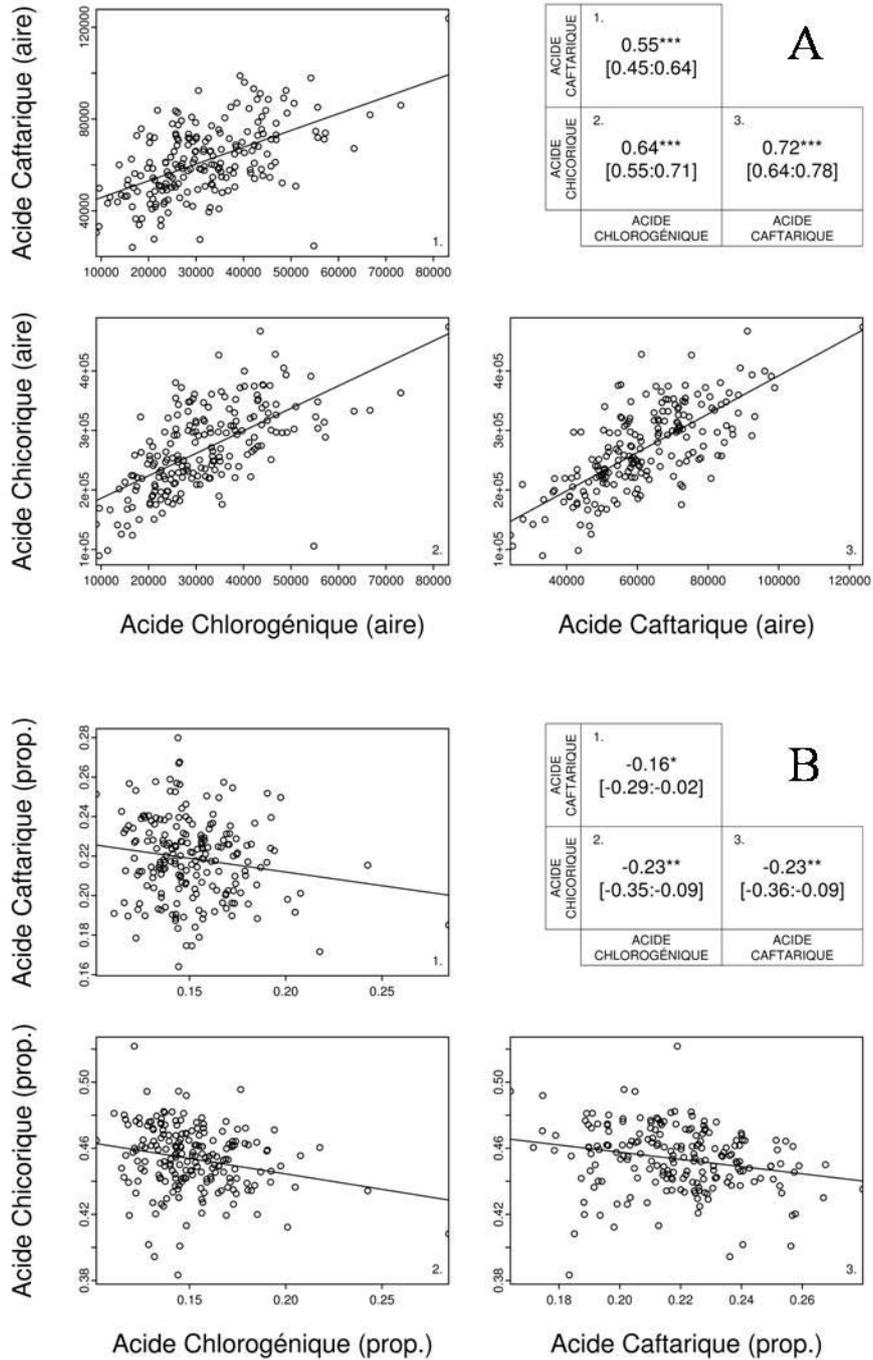


Figure 17. Corrélation de Pearson entre A : les aires de chaque molécule d'intérêt entre elles ; B : la proportion de chaque molécule d'intérêt entre elle. Les symboles*, ** et *** indiquent si les données sont significativement corrélées à $p < 0,05$, $p < 0,01$ et $p < 0,001$, respectivement. Les valeurs entre crochets correspondent à l'intervalle de confiance.

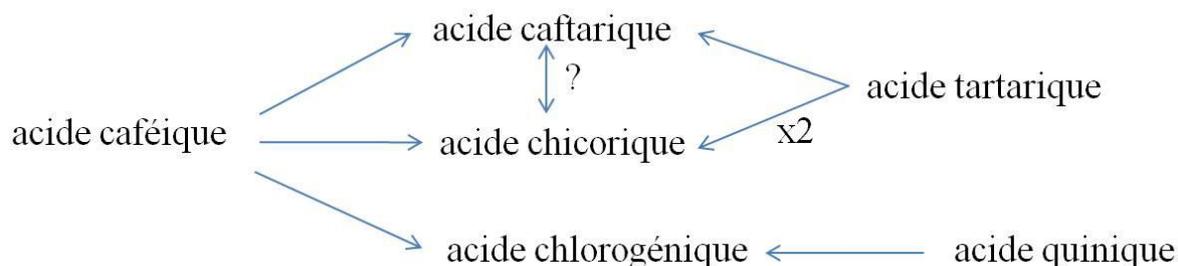


Figure 18. Schéma de la compétition supposée entre l'acide caftarique, chicoriquique et chlorogénique par rapport à leurs précurseurs.

On remarque aussi que seule ACHLO n'est pas corrélée à l'AAR (tableau 5), ce qui semble contradictoire avec les informations issues des essais préliminaires (paragraphe III.2). De manière analogue, l'équipe de Rousseaux (2005) avait signalé une corrélation quasi-nulle (et surprenante) entre le taux d'acide ascorbique et la variation de la capacité antioxydante dans leurs échantillons de tomate. Cependant, lorsqu'ils ont à nouveau mené l'expérience un an plus tard, la concentration en acide ascorbique a expliqué 51% de la variation en activité antioxydante contre 1% l'année précédente, suggérant l'importance de l'effet environnement. Ainsi, dans un environnement donné, le pouvoir antioxydant n'est pas toujours expliqué par un même composé, mais plutôt par un mélange variable de molécules, chaque composé étant plus ou moins spécifiquement synthétisé suite à un stress particulier (cf. analyse bibliographique, paragraphe II.2). Dans notre cas, d'autres molécules présentes dans les feuilles de chicorées doivent compléter l'activité antiradicalaire. Il peut s'agir d'autres acides phénols encore non identifiés dans nos échantillons, de flavonoïdes ou de vitamines bien que l'acide caféique, le kaempférol, la quercétine et l'acide ascorbique n'aient pas été détectés dans l'échantillon analysé en CL/SM. Néanmoins, dans une publication de 2004, Llorach remarque qu'une activité antioxydante est décelée en l'absence de composés phénoliques. Cette activité pourrait éventuellement être due à des sesquiterpènes lactones comme la lactucine et ses dérivés dont l'activité anti-radicalaire -bien que faible- ait été démontrée chez *Lactuca scariolia* (Kim, 2001).

IV.3.C.b. Intérêt du croisement

A l'image de Davey *et al.* (2006), nous avons choisi un croisement déjà existant au sein de notre laboratoire, pour une étude des régions chromosomiques impliquées dans la biosynthèse de trois hydroxycinnamates et de l'activité antiradicalaire. Pour sept lots de

données sur neuf, une différence significative entre les individus est obtenue ($p < 0,001$), montrant que cette population ségrége aussi pour d'autres caractères quantitatifs que ceux choisis initialement (embryogenèse somatique), autorisant alors cette nouvelle étude de QTL.

IV.3.C.c. Transgression

On observe de la transgression pour chaque donnée (figures 14 et 15) : les valeurs obtenues pour la descendance ne sont pas seulement comprises entre celles des parents mais aussi de part et d'autre (valeurs inférieures ou supérieures à celles de K59 et K28). Après analyse de plus de cent études discutant de la variation phénotypique dans des populations végétales en ségrégation (Rieseberg *et al.*, 1999), le phénomène de transgression ne semble pas exceptionnel mais plutôt commun : 58% des traits étudiés présentent une ségrégation transgressive. Parmi les différentes hypothèses suggérées dans la littérature, le même auteur souligne trois possibilités pour expliquer ce fait. Parmi elles, l'hypothèse la plus courante est la complémentarité des allèles des deux parents : l'addition des allèles parentaux provoquerait de nouvelles combinaisons influençant de façon négative ou positive le phénotype. Dans ce cas, on peut aussi noter l'expression d'allèles rares récessifs, phénomène visible surtout pour des croisements entre populations sauvages et cependant peu retrouvé dans la littérature (Rick & Smith, 1953). Une autre hypothèse pouvant expliquer la transgression réside dans l'existence de super-dominance, aussi appelée hétérosis : la combinaison allèle récessif/allèle dominant donne une valeur phénotypique plus importante que celle obtenue par deux allèles dominants. La troisième suggestion de Rieseberg s'appuie sur le phénomène d'épistasie : la combinaison des allèles parentaux peut générer de nouvelles interactions épistatiques tandis que d'autres peuvent disparaître.

L'expression d'allèles rares récessifs est un phénomène peu rapporté qui ne s'adapte pas à notre étude (plantes cultivées et auto-incompatibles). Le phénomène de super-dominance est possible dans notre descendance issue du croisement entre deux parents hétérozygotes pour de nombreux loci. Les deux autres hypothèses, souvent mises en avant dans la littérature, peuvent également s'appliquer à notre cas de figure.

IV.4. Héritabilité au sens large et effet de l'environnement

Les plus fortes héritabilités ont été obtenues pour l'AAR et la proportion en acide chlorogénique (0,531) tandis que l'héritabilité la plus faible correspond à l'aire des pics des 3 molécules (0,296). Ces valeurs sont relativement faibles par rapport à celles retrouvées pour d'autres caractères : avec quasiment la même population, les analyses sur l'embryogenèse somatique donnaient une l'héritabilité, en général, supérieure à 0,65 (Thèse Aline Clabaut, 2009). Concernant l'héritabilité de caractères liés au métabolisme secondaire, Kliebenstein et ses collaborateurs (2001) notent une héritabilité de 0,81 pour les concentrations en glucosinolates aliphatiques totaux des feuilles d'*Arabidopsis* mais de seulement 0,40 pour les graines. Une héritabilité de 0,91 a été trouvée pour la concentration en acide *p*-coumarique dans les chaumes de maïs alors que pour les dérivés de l'acide férulique, aucune variation phénotypique n'a été trouvée, donc l'héritabilité n'a pas pu être calculée (Barrière *et al.*, 2008). En 2007, Werij *et al.* détectent des QTL impliqués dans la biosynthèse de tyrosine et d'acide chlorogénique dans les tubercules de pomme de terre. L'héritabilité correspondant au taux de tyrosine est de 0,88 tandis que l'héritabilité liée à l'acide chlorogénique n'est pas précisée. Les travaux sur les taux de différents tocophérols dans la graine de soja relatent une héritabilité variant entre 0,22 et 0,40 (Li *et al.*, 2010), valeurs correspondant à peu près aux nôtres. Les autres recherches de QTL portant sur des molécules antioxydantes, citées dans le tableau 3 de l'analyse bibliographique, n'indiquent pas l'héritabilité trouvée pour les caractères phénotypés ce qui rend assez difficile toute comparaison avec nos résultats (Abe *et al.*, 2002; Szalma *et al.*, 2002; Fontaine *et al.*, 2003; Dong *et al.*, 2005; Rousseaux *et al.*, 2005; Davey *et al.*, 2006; Stevens *et al.*, 2007; Jin *et al.*, 2009; Frary *et al.*, 2010). De manière générale, l'héritabilité varie bien sûr en fonction du caractère étudié mais aussi en fonction du dispositif expérimental, du modèle végétal et de la descendance.

Le design expérimental et le nombre de répétitions changent selon chaque publication précédemment citée. Les travaux de Rousseaux (2005), Fontaine (2003) et Stevens (2007) ont opté pour un dispositif en trois blocs (répartition aléatoire) avec des répétitions sur deux ou trois années. Contrairement à nos travaux, aucun effet bloc n'a été relevé alors qu'un effet année a été rapporté, sauf pour l'équipe de Stevens dont les QTL détectés sont restés stables d'une année sur l'autre et dans différents environnements (champs et serre). La disposition des génotypes en bloc et une analyse en plusieurs années ou dans différents lieux ne sont

cependant pas systématiques (Abe *et al.*, 2002; Szalma *et al.*, 2002; Dong *et al.*, 2005; Werij *et al.*, 2007; Frary *et al.*, 2010). Néanmoins, étant donné la particularité du caractère étudié, lié de manière relativement forte à l'environnement (Rousseaux *et al.*, 2005), le nombre de blocs et les répétitions dans le temps et l'espace, dans la mesure du possible, semblent être une précaution à ne pas négliger pour assurer la stabilité des données.

V. Conclusion

Un protocole rapide et simple à mettre en œuvre a été établi pour extraire les molécules polyphénoliques des feuilles de *Cichorium intybus* L. Trois acides hydroxycinnamiques ont été identifiés en CLHP et validés en LC/SM et SM/SM : l'acide caftarique, l'acide chlorogénique et l'acide chicorique. Cette caractérisation biochimique pourra être complétée ultérieurement, notamment concernant les flavonoïdes ou les sesquiterpènes lactones, connus pour être présents dans les feuilles de chicorée. Une descendance de chicorée F1' (croisement K59xK28) a ensuite été clonée en trois à cinq exemplaires. Ces plantes ont été analysées pour leur contenu en hydroxycinnamates et leur activité antiradicalaire. Les descendants de ce croisement présentent de la transgression par rapport aux parents. Les tests statistiques ont montré que cette descendance présentait de la variabilité phénotypique pour sept caractères sur neuf (ACAFT, ACHLO, ACHIC, ATOT, PCAFT, PCHLO, AAR). L'héritabilité des caractères mesurés est néanmoins faible, soulignant l'importance de l'effet environnemental dans ce type d'étude. Cependant, les résultats obtenus vont permettre d'associer ces variations phénotypiques à une variation génotypique, dans le cadre d'une analyse QTL visant à détecter les régions chromosomiques responsables de la variation des sept caractères quantitatifs cités précédemment.

CHAPITRE II

Mise en place d'une carte génétique
pour la détection de QTL et
cartographie de gènes candidats

Dans le chapitre précédent, nous avons mis en place un protocole pour l'extraction des principaux hydroxycinnamates contenus dans les feuilles de chicorée. Un test antiradicalaire (DPPH) et des analyses CLHP ont permis d'enregistrer des valeurs phénotypiques pour les caractères quantitatifs AAR, ACAFT, ACHLO, ACHIC, PCAFT, PCHLO, PCHIC et PTOT. Des analyses de variance ont établi de façon significative que ces caractères présentaient de la variabilité sauf PCHIC et PTOT. L'importance de l'effet environnemental a été révélé par ces mêmes analyses et confirmé par le calcul de l'héritabilité des caractères mesurés (entre 0,296 et 0,531) qui reste faible.

Dans le cadre d'une analyse QTL, la variabilité phénotypique observée doit être mise en relation avec la variabilité génotypique afin d'identifier des loci responsables des caractères quantitatifs cités dans le paragraphe précédent. Les individus de la descendance étudiée, dont une partie n'avait pas été génotypée antérieurement à ce travail, ont été génotypés à leur tour via des marqueurs présentant du polymorphisme AFLP, SSCP et SSR. Ces marqueurs ont pu alors être placés sur une carte génétique dédiée à ce projet et présentée dans ce chapitre sous le nom de carte « oxychic ». D'autre part, en synthétisant ce qui a été trouvé chez d'autres espèces, une liste de gènes impliqués dans la voie de biosynthèse des composés phénoliques a été établie. La recherche de séquences correspondantes à ces gènes chez la chicorée et ensuite la recherche de polymorphisme SSCP et HRM au sein de la descendance K28 x K59 a permis la localisation de gènes candidats sur la carte « oxychic ».

I. Historique de la carte génétique de la chicorée ($2n=2x=18$) au laboratoire SADV

En 2001, le laboratoire SADV de l'Université des Sciences et Technologies de Lille1 s'est associé aux semenciers Ets Hoquet et Ets Florimond Desprez, et la société FINALER (groupe Leroux) sous la forme d'un Groupement d'Intérêt Scientifique pour la construction d'une carte génétique de la chicorée, espèce dont la taille du génome se situerait entre 1300 (Pierre Devaux, donnée personnelle) et 1400 Mbp (Koopman, 2002). Ce projet, baptisé Cartochic, a permis l'obtention d'une carte consensus (881,7 cM, 472 marqueurs) basée sur trois descendance (Cadalen *et al.*, 2010). La première (F_1 ' , 177 individus) est issue du croisement entre deux parents hétérozygotes, K28 et K59, préalablement choisis en fonction de leur capacité ou incapacité à produire des embryons somatiques (annexe 2, Laboratoire SADV, USTL, France). Le parent K28 a été utilisé comme parent femelle car il est mâle

stérile. Le second croisement est appelé Rubis118. Il s'agit d'une descendance de 96 individus de type F₂, provenant d'un clone mâle stérile MS8 et d'un cultivar de chicorée « industrielle » dénommé Cassel (Florimond-Desprez, Cappelle-en-Pévèle). La troisième descendance nommée BR (F₂, 145 individus) résulte de l'autofécondation d'un hybride F₁ lui-même issu du croisement entre deux lignées parentales de chicorée « endive » (Hoquet Endives, Raillencourt), l'une à feuilles rouges et l'autre à feuilles blanches (Cadalen *et al.*, 2010).

Notons aussi que deux autres cartes ont été préalablement publiées chez *Cichorium intybus* L., toutes deux basées sur des marqueurs essentiellement dominants. La première (De Simone *et al.*, 1997) a été construite à partir de marqueurs 283 AFLP, 72 SAMPL et 16 RAPD et la seconde (Van Stallen *et al.*, 2003) à partir de 129 marqueurs RAPD (Random-Amplified Polymorphic DNA), utilisée ensuite pour la détection de QTL (Van Stallen *et al.*, 2005). Les marqueurs n'étant pas les mêmes, aucun lien ne peut être établi entre ces deux cartes.

D'autres cartes consensus sont également disponibles chez d'autres Astéracées comme la laitue (Jeuken *et al.*, 2001) et le tournesol (Yu, 2003 #377).

Deux types de marqueurs moléculaires sont utilisés pour la construction de la carte consensus : il s'agit de marqueurs dominants et codominants. Ces derniers peuvent présenter simultanément les allèles des deux parents.

Le premier type de marqueurs codominants utilisé correspond aux SSR (Simple Sequence Repeats) ou microsatellites. Il s'agit de répétitions de motifs di-, tri- ou tétranucléotidiques, révélant un polymorphisme de nombre d'unités de répétitions. Les marqueurs SSR ayant servis à la construction de la carte génétique proviennent de trois banques enrichies en motifs microsatellites: Agrogène (AGR, France), Ecogenics (SW, Suisse) et notre laboratoire SADV (EU, France).

Pour le second type de marqueurs moléculaires codominants utilisés, la visualisation du polymorphisme s'est faite grâce à la technique SSCP (Single Strand Conformation Polymorphism). L'ADN double brin est dénaturé à 95°C puis refroidi immédiatement afin que les brins complémentaires ne se réassocient pas. L'ADN simple brin va alors former des structures tridimensionnelles spécifiques obtenues pour des conditions de température donnée. Des différences dans la séquence entre deux génotypes (mutation ponctuelle, petites insertions/délétions) sont à l'origine de ce polymorphisme de conformation. Cette technique a été utilisée pour cartographier des gènes isolés au cours de travaux sur l'embryogenèse somatique, des EST (Expressed Sequence Tags) générés par le laboratoire SADV et par

d'autres laboratoires travaillant sur la chicorée. Depuis 2006, le site internet intitulé « the compositae genome project » de l'université Davis en Californie regroupe les données EST publiques de *Cichorium intybus* (38323 EST) mais aussi de *Cichorium endivia* (30171 EST) et d'autres Astéracées (<http://cgpdb.ucdavis.edu/>).

Pour compléter la carte génétique, des marqueurs dits dominants ont été ajoutés. Ils ont été révélés par la technique AFLP (Amplification Fragment Length Polymorphism). Cette technique met en exergue du polymorphisme de sites de restriction et d'hybridation d'amorces arbitraires: l'ADN est digéré par des enzymes de restriction puis des adaptateurs de séquence connue sont ajoutés aux extrémités des bouts cohésifs. Ensuite, deux amplifications PCR sélectives se succèdent : à chaque fois, des amorces qui correspondent aux adaptateurs sont prolongées en 3' par un puis deux nucléotides arbitraires, respectivement. Cette technique permet une densification rapide d'une carte génétique: les couples d'amorce révèlent du polymorphisme simultanément à plusieurs endroits du génome et non de façon individuelle. La lecture de ce polymorphisme a été effectuée sous la forme présence/absence, lecture donnant moins d'informations que celles allouées aux marqueurs co-dominants. Notons qu'il est néanmoins possible de lire les AFLP comme des marqueurs co-dominants à l'aide de logiciels capables de reconnaître les marqueurs AFLP liés (Vuylsteke *et al.*, 1999).

Cent onze des 201 individus utilisés pour l'identification de régions chromosomiques impliquées dans la biosynthèse de molécules détectées (cf. chapitre I) ont été génotypés lors des travaux de thèse d'Aline Clabaut (Analyse génétique de l'embryogenèse somatique chez la chicorée : cartographie des QTL et gènes candidats, juin 2009) et pour la carte de référence K28xK59 (annexe 2). Pour l'obtention d'une carte génétique spécifique à mes travaux de thèse, l'autre partie des individus (les 90 restants) a été génotypée avec 127 marqueurs moléculaires (SSR, SSCP et AFLP). Ils ont été choisis par rapport à la carte de référence et aux travaux d'Aline Clabaut, pour recouvrir les neufs groupes de liaison de façon à avoir un marqueur tous les 10 cM environ (figure 1). Les gènes candidats ont ensuite complété cette carte (polymorphisme détecté par SSCP ou HRM).

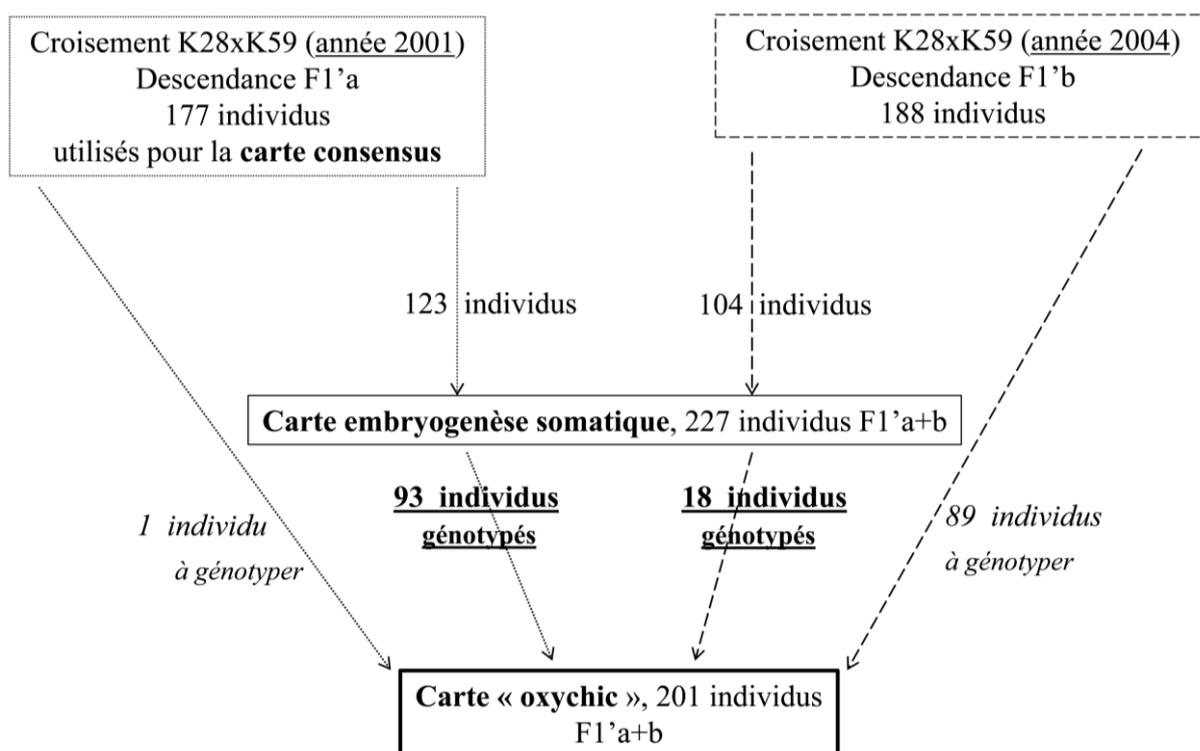


Figure 1. Provenance des individus phénotypés : 111 individus ont été génotypés préalablement pour la carte embryogénèse somatique. 94 individus (47%) proviennent d'un croisement entre K59 et K28 datant de 2001 alors que 107 individus (53%) proviennent du croisement de 2004.

II. Matériels et méthodes, carte « oxychic »

II.1. L'extraction d'ADN

L'extraction de l'ADN a été effectuée à l'aide du kit GenElute genomic DNA kit (Sigma-Aldrich, St-Louis, MO). De jeunes feuilles de chicorées cultivées en serre ont été prélevées et immédiatement broyées à l'azote liquide. Aux 100 mg de cette poudre sont ajoutés 400 μ L d'un mélange de deux solutions de lyse avant incubation à 65°C pendant 10 minutes. Ensuite, l'ajout de 130 μ L de solution de précipitation ont permis la précipitation des débris cellulaires, des protéines et des polysaccharides après 5 minutes d'incubation dans la glace puis 5 minutes de centrifugation. Le surnageant a alors été filtré, purifié puis élué trois fois par une solution d'éluion chauffée à 65°C. La concentration de cet éluat d'ADN a été déterminée à partir des mesures prises au spectrophotomètre à 260 nm.

II.2. Les marqueurs moléculaires

II.2.A. Les marqueurs SSR

La détection du polymorphisme d'unité de répétition a été mise en œuvre à partir de trois banques enrichies (AGR, SW et EU) en motifs di-nucléotidiques de type CA et GA (travaux de Monika Mörchen et Thierry Cadalen, (Cadalen et al., 2010)). Pour la carte génétique « oxychic », 57 marqueurs microsatellites ont été répartis sur les neufs groupes de liaison.

II.2.A.a. Les amorces

Dans un premier temps, une amorce de chaque couple d'amorces (Sigma-Aldrich) a été rallongée par une queue M13F (5'-CACGACGTTGTAAAACGAC) ou M13R (5'-GGATAACAATTTTCACACAGG) (Oetting *et al.*, 1995). Ensuite, pour les couples d'amorces ayant généré des signaux trop faibles et donc des profils électrophorétiques difficiles à lire, la queue M13 a été remplacée par un fluorochrome.

II.2.A.b. Les réactions PCR

Elles sont effectuées dans un volume final de 15 μ L dont 2 μ L d'ADN à 1,5 ng/ μ L. Le tampon PCR x 1 est additionné de MgCl₂ 2mM, des quatre dNTP à 100 μ M chacun (Applied Biosystems), de BSA à 0,2 mg/mL (Sigma-Aldrich), et de Taq Polymérase 0,3 U/ μ L (Applied Biosystems).

Pour les réactions utilisant des couples d'amorces avec une queue M13, sont ajoutés l'amorce possédant la queue M13 à la concentration de 33 nM, l'amorce sans la queue M13 à 333 nM, la queue M13 couplée à un fluorochrome 6-FAMTM (bleu), VicTM (vert), PetTM (rouge) ou NedTM (jaune) à la concentration de 3 nM (Sigma-Aldrich).

Pour les réactions utilisant des amorces préalablement liées à un fluorochrome bleu 6-FAMTM ou vert HexTM (Applied Biosystems), les concentrations sont de 133 nM pour l'amorce marquée comme pour l'amorce non marquée.

Les conditions PCR établies pour les couples d'amorces munies d'une queue M13 sont les suivantes :

Dénaturation : 5 minutes à 94°C
 Dénaturation : 30 secondes à 94°C
 Hybridation : 35 secondes à 55°C
 Elongation : 45 secondes à 72°C
 Dénaturation : 30 secondes à 94°C
 Hybridation : 20 secondes à 56°C
 Elongation : 45 secondes à 72°C
 Extension terminale : 7 minutes à 72°C

} 10 cycles

} 35 cycles

Les conditions PCR établies pour les couples d'amorces préalablement marquées d'un fluorochrome :

Dénaturation : 4 minutes à 94°C
 Dénaturation : 30 secondes à 94°C
 Hybridation : 40 secondes à 60°C
 Diminution de -1°C par cycle
 Elongation : 45 secondes à 72°C
 Dénaturation : 30 secondes à 94°C
 Hybridation : 40 secondes à 55°C
 Elongation : 45 secondes à 72°C
 Extension terminale : 7 minutes à 72°C

} 5 cycles

} 35 cycles

II.2.A.c. Séparation des produits PCR par électrophorèse capillaire

Des groupes de produits PCR ont été constitués afin d'analyser plusieurs produits d'amplification en même temps. Pour ne pas être confondus, les produits PCR d'un même groupe doivent différer par leur fluorochrome et/ou leur taille en paire de bases quand celle-ci est déjà connue. Un test est d'abord effectué pour voir dans quelle proportion les produits d'amplification doivent être mélangés. Certains ayant un signal de lecture plus fort que d'autres, ils doivent être présents en plus petite concentration afin de ne pas « écraser » les signaux plus faibles.

Une fois ce paramètre optimisé, 1µL de ce mélange est associé à 10 µL de formamide et 0,1 µL de marqueur de taille interne (LYS ou ROX, Applied Biosystems). Après une dénaturation de 4 minutes à 95°C, les produits d'amplification migrent en condition

dénaturante (polymère POP4, tampon GAB x 1, Applied Biosystems) et sont séparés par le séquenceur capillaire ABI PRISM® 3100-Avant (Applied Biosystems, Foster City, CA, USA). Les résultats de ces migrations sont ensuite analysés avec le logiciel GeneMapper®v.4.0.

II.2.B. Les marqueurs SSCP (EST)

Quinze marqueurs SSCP ont été choisis pour la carte « oxychic », sept d'entre eux sont issus du travail d'Isabelle Scheer (Cadalen *et al.*, 2010), tandis que les huit autres ont été générés par Aline Clabaut (marqueurs SSCP dont le nom commence par « OC »). Comme pour les marqueurs microsatellites, notre travail a consisté à génotyper des individus utilisés pour le phénotypage et non génotypés dans les travaux de cartographie antérieurs.

II.2.B.a. Obtention des séquences EST

Au cours des travaux de doctorat de Sylvain Legrand (Etude transcriptomique au cours des phases précoces de l'embryogenèse somatique chez la chicorée, décembre 2006), une banque soustractive d'ADNc a été réalisée à partir des ARNm de K59 (embryogène) et C15 (descendant d'une autofécondation de K59, individu non embryogène), après quatre jours de culture en condition d'embryogenèse somatique (Legrand *et al.*, 2007). La technique SSH (Suppression Subtractive Hybridisation) de Diatchenko *et al.*, 1996, a permis d'obtenir 3348 EST qui sont potentiellement exprimés de façon différentielle chez ces deux génotypes.

Les EST ont été regroupés en 2302 OC (Original Cluster). A l'intérieur d'un OC, des sous-groupes ont été créés sur le critère d'au moins 95% d'homologie sur 150 pb. Ces OC ont été alignés avec des EST d'autres Astéracées et d'*Arabidopsis* (the compositae genome project, Plant Genome Database) pour connaître les fonctions éventuelles des gènes leur correspondant. Les alignements ont montré que les 3348 EST représentaient 2077 gènes. Une analyse par biopuces a révélé 105 gènes différentiellement exprimés chez K59 et C15 au quatrième jour de culture en condition d'induction de l'embryogenèse somatique (Thèse de Sylvain Legrand, université de Lille1, 2006). Ce sont les EST correspondant à ces gènes qui ont été pris comme base pour la recherche de marqueurs moléculaires polymorphes.

II.2.B.b. Recherche du polymorphisme et définition des amorces

Cette étape a été réalisée au cours de la thèse d'Aline Clabaut (2009). A l'aide du logiciel BioEdit v. 7.0.5.3 (Hall, 1999), les EST correspondants aux gènes différentiellement exprimés chez K59 et C15 ont été alignés afin de détecter du polymorphisme allélique de type SNP (Single Nucleotide Polymorphism) c'est-à-dire un changement de base. C'est de part et d'autre de ces SNP détectés *in silico* que des amorces ont été définies.

Quand aucun SNP n'a été trouvé, une recherche de polymorphisme au sein des introns, a été menée. En effet, l'expérience du laboratoire a montré que le taux de mutations introniques est généralement élevé et que l'emplacement des jonctions intron/exon entre *Cichorium intybus* et *Arabidopsis thaliana* est bien conservée. Ainsi, les EST de chicorée ont été alignés avec les gènes d'*Arabidopsis* grâce au site internet NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>). Pour déterminer les jonctions intron-exon, l'ADNc d'*Arabidopsis* a été aligné avec l'ADN correspondant via le logiciel BioEdit v.7.0.5.3 (Hall, 1999). Pareillement, les EST de chicorée ont été alignés avec l'ADNc de l'arabette et la localisation des introns répercutée sur les EST. Des amorces adjacentes aux introns ont alors été définies.

Pour les EST ne coïncidant pas avec la localisation d'introns d'*Arabidopsis*, un alignement avec des séquences de *Lactuca sp.* (the compositae genome project, université Davis, CA, <http://cgpdb.ucdavis.edu/>) a été mis en œuvre afin de construire des amorces hybrides. Les séquences de la laitue, elles-mêmes confrontées à celle d'*Arabidopsis* ont permis de définir de nouvelles amorces encadrant les introns.

Enfin, l'utilisation du logiciel (NetGene 2 v 2.4) a permis de prédire l'emplacement des jonctions intron-exon des EST de chicorées ne présentant ni SNP et ni homologie avec *Arabidopsis*.

Les amorces sens et anti-sens ont été définies, dans chaque cas, à l'aide du logiciel Primer 3 v0.3.0 (Rozen & Skaletsky, 1999) (longueur des amorces entre 19 et 23 pb, température d'hybridation (T_m) entre 55 et 60°C).

II.2.B.c. Les réactions PCR

Elles sont effectuées dans un volume final de 15 μ L dont 2 μ L d'ADN à 1,5ng/ μ L. Le tampon PCR x 1 est additionné de MgCl₂ 2mM, des quatre dNTP à 100 μ M chacun (Applied

Biosystems), de BSA à 0,2 mg/mL (Sigma-Aldrich), et de Taq Polymérase 0,3 U/μL (Applied Biosystems).

Les amorces sont concentrées à hauteur de 133 nM, que ce soit pour l'amorce marquée (fluorochrome bleu 6-FAMTM ou vert HEXTM, Applied Biosystems), ou pour l'amorce non marquée.

Les conditions PCR établies pour ces couples d'amorces sont les suivantes:

Dénaturation : 5 minutes à 94°C	} 5 cycles	
Dénaturation : 30 secondes à 94°C		
Hybridation : 30 secondes à T _m		
Diminution de -1°C par cycle		
Elongation : 30 secondes à 72°C	} 35 cycles	
Dénaturation : 30 secondes à 94°C		
Hybridation : 30 secondes à T _m minoré de 5°C		
Elongation : 30 secondes à 72°C		
Extension terminale : 7 minutes à 72°C		

II.2.B.d. Séparation des produits PCR par électrophorèse capillaire

Sur les mêmes critères que pour la technique SSR, des groupes de produits d'amplification ont été constitués afin d'analyser plusieurs produits PCR en même temps. Cependant il ne s'agit pas ici de taille en paire de base mais d'une vitesse de migration dépendante des structures secondaires formées par l'ADN simple brin.

1μL de produit d'amplification est ajouté à 10 μL de formamide et 0,5 μL de marqueur de taille interne (LYS ou ROX) et 0,5 μL de soude. Après une dénaturation de 4 minutes à 95°C, les produits PCR migrent en condition dénaturante (polymère CAP, tampon GAB x 1) et sont séparés par le séquenceur capillaire ABI PRISM® 3100-Avant. Les résultats de ces migrations sont ensuite analysés à l'aide du logiciel GeneMapper®v.4.0.

II.2.C. Les marqueurs AFLP

Pour la carte « oxychic », 56 marqueurs AFLP ont été choisis parmi les marqueurs générés au laboratoire SADV par Lucy Gonthier, Christelle Blassiau et Marie-Christine Quillet.

II.2.C.a. II.2.C.a. La restriction enzymatique

500 ng d'ADN ont été digérés pendant 3 heures, au bain-marie à 37°C, par un mélange de restriction composé de Tampon Ligase x 1 (New England Biolabs), BSA x 5, NaCl 50mM, enzymes *EcoRI* et *MseI* à la concentration de 0,333U/μL chacun (New England Biolabs), pour un volume final de 30 μL.

II.2.C.b. La ligation des adaptateurs

Le mélange de ligation, dans un volume final de 5 μL est composé comme suit : adaptateur *EcoRI* 1pmol/μL, adaptateur *MseI* 10pmol/μL, tampon ligase x 1, ATP 1mM, et T4 DNA ligase 32mU Weiss (New England Labs). Ce mélange est ajouté aux 30 μL du mélange de restriction et le tout est maintenu 16h à 37°C. Cette étape est vérifiée en déposant 5μL des produits de restriction-ligation sur un gel d'agarose 1%, TBE 5x.

II.2.C.c. La réaction PCR de pré-amplification

Elle se fait dans un volume total de 30 μL, en prenant 3 μL du produit de restriction-ligation obtenu précédemment. Le tampon PCR x 1 est additionné de MgCl₂ 2,5mM, des dNTP à 200 μM chacun, de Taq Polymérase à 0,05 U/μL (Applied Biosystems) et des amorces *EcoRI*-A (5'-GACTGCGTACCAATTCA-3') et *MseI*-C (5'-GATGAGTCCTGAGTAAC-3') à la concentration finale de 0,266 pmol/μL (Sigma-Aldrich).

Les 20 cycles de PCR se déroulent selon les trois premières étapes suivantes :

Dénaturation : 30 secondes à 94°C

Hybridation : 40 secondes à 56°C

Elongation : 1 minute à 72°C

Extension terminale : 5 minutes à 72°C

Cette étape est vérifiée en déposant 8 μL des produits de préamplification sur un gel d'agarose 1%, TBE 5x (migration 1h à 100 volts).

II.2.C.d. La réaction PCR d'amplification sélective

Elle se fait dans un volume total de 10 μL en prenant 3 μL du produit de pré-amplification dilué 50 fois. Le tampon PCR x 1 est additionné de MgCl_2 2mM, des dNTP à 200 μM chacun, et de Taq Polymérase (Applied Biosystems) à 0,025 U/ μL et des amorces *EcoRI*-ANN et *MseI*-CNN à la concentration finale de 0,24 pmol/ μL , avec N = A, C, G ou T (Sigma-Aldrich).

Les 13 premiers cycles de PCR se déroulent comme ceci :

Dénaturation : 30 secondes à 94°C

Hybridation : 30 secondes à 65°C

Elongation : 1 minute à 72°C

Diminution de -0,7°C par cycle

Pour les 23 cycles suivants :

Dénaturation : 30 secondes à 94°C

Hybridation : 30 secondes à 56°C

Elongation : 1 minute à 72°C

L'extension terminale dure 5 minutes, à une température de 72°C.

II.2.C.e. II.2.C.e. Séparation des produits PCR par électrophorèse capillaire

Un microlitre de produit d'amplification PCR est combiné à 10 μL de formamide et 0,1 μL de marqueur de taille interne (LYS ou ROX, Applied Biosystems). Après une dénaturation de 4 minutes à 95°C, les produits d'amplification migrent en condition dénaturante (polymère POP4, tampon GAB x 1) et sont séparés par le séquenceur capillaire ABI PRISM® 3100-Avant. Les résultats de ces migrations sont ensuite analysés avec le logiciel Genotyper 3.7.

II.3. Construction de la carte génétique

Les différents marqueurs, générés via les trois techniques précédemment citées, ont été liés génétiquement entre eux grâce au logiciel JoinMap v 3.0 (van Ooijen & Voorrips 2001), capable d'analyser les populations de type F1' issues de parents hétérozygotes (population appelée CP pour cross-pollinated). Les distances génétiques ont été calculées selon la méthode de Kosambi (Kosambi, 1944). La fréquence de recombinaison a été fixée à 0,45 pour tous les groupes de liaison sauf le groupe de liaison 5 (0,40). Le LOD seuil de constitution d'un groupe de liaison a été fixé à 3. Les résultats du génotypage sont codés comme l'illustre le tableau 1.

Loci parentaux	Ségrégation	Génotypes possibles	Proportion
Hétérozygote chez K59 2 allèles	<lm x ll>	ll, lm, --	1 : 1
Hétérozygote chez K28 2 allèles	<nn x np>	nn, np, --	1 : 1
Hétérozygote chez K59 et K28 2 allèles	<hk x hk>	hh, kk, hk, --	1 : 2 : 1
Hétérozygote chez K59 et K28 2 allèles	<hk x hk> (h- ; kk)	h-, kk, --	3 : 1
Hétérozygote chez K59 et K28 3 allèles	<ef x eg>	ef, eg, ee, fg, --	1 : 1 : 1 : 1
Hétérozygote chez K59 et K28 4 allèles	<ab x cd>	ac, ad, bc, bd, --	1 : 1 : 1 : 1

Tableau 1. Le type de ségrégation retrouvé dans la descendance K28xK59 est codé comme le préconise le logiciel JoinMap v 3.0 dans le cas d'une population « cross-pollinated ». Les données manquantes sont mentionnées par le symbole « -- ».

II.4. Distorsion de ségrégation : double test de χ^2

La ségrégation pour chaque marqueur a été évaluée par un test de χ^2 . Pour les marqueurs reflétant la distorsion de ségrégation (à partir de $p < 0,05$), l'origine de la distorsion (zygotique ou gamétique) a été déterminé à partir des marqueurs co-dominants en effectuant successivement deux tests de χ^2 (Lorieux *et al.*, 1995a). Le premier test χ^2 pose comme hypothèse nulle (H_0): les fréquences alléliques sont les mêmes. Le deuxième χ^2 pose H_0 = conformité de la distribution des fréquences aux classes génotypiques par rapport à une distribution théorique (par exemple, pour un marqueur à deux allèles, avec p et q étant les

fréquences alléliques, on teste la conformité à la distribution suivante : $p^2 : 2pq : q^2$). En suivant la procédure décrite par Lorieux (1995), on peut estimer qu'il s'agit d'une sélection zygotique lorsque le deuxième test de χ^2 est significatif, quelque soit la signification du premier test de χ^2 . A l'inverse, lorsque le premier test χ^2 est significatif et que le deuxième ne l'est pas, il serait question d'une sélection gamétique.

III. Matériels et méthodes, gènes candidats

III.1. Recherche des EST *Cichorium* à partir de gènes candidats identifiés chez d'autres espèces (en collaboration avec Monika Mörchen).

En synthétisant ce qui a été trouvé chez des espèces voisines ou les espèces modèles, il nous a été possible de composer une liste de gènes pouvant postuler au rôle de gènes impliqués dans la voie des composés phénoliques détectés chez la chicorée. La recherche de gènes candidats s'est appuyée sur les voies métabolites disponibles sur le site KEGG pathways (<http://www.genome.jp/kegg/pathway.html>) et les publications ayant révélé des gènes impliqués dans les voies de biosynthèse détaillées dans le paragraphe III de l'analyse bibliographique :

- la voie des phénylpropanoïdes (cf. figures 2 et 3), débutant par l'action de l'enzyme PAL sur la phénylalanine. Cette voie de métabolisme aboutissant à la synthèse des lignines et des flavonoïdes passe par la synthèse des acides hydroxycinnamiques. Parmi eux, on notera l'acide caféique et l'acide quinique qui sont les précurseurs de l'acide chlorogénique ;
- les voies métaboliques qui bordent la synthèse d'acide tartarique (cf. figures 4 et 5): la voie de dégradation de l'acide ascorbique et la voie du glyoxylate. L'acide tartarique associé à une molécule d'acide caféique (voie des phénylpropanoïdes) forme l'acide caftarique alors qu'il faut deux molécules d'acide caféique pour former l'acide chicorique ;
- les facteurs de transcription régulant l'expression des gènes codant pour les enzymes impliquées dans les voies métaboliques précédemment citées.

Dans les figures 2 à 5, les cadres en gras indiquent les enzymes dont les séquences étaient disponibles et qui ont ainsi pu servir à la recherche de gènes candidats.

La recherche des EST de *Cichorium intybus* s'est effectuée à partir de la banque EST de l'université de Davis (Californie, Etats-Unis, the compositae genome project,

CHAPITRE II

<http://cgpdb.ucdavis.edu/>). Si aucun EST n'a été trouvé pour *C. intybus*, la recherche a été dirigée vers les EST de *Cichorium endivia*. Parmi les EST de *Cichorium intybus*, sont également présents ceux issus des travaux de Sylvain Legrand (thèse université de Lille1, 2006 ; (Legrand *et al.*, 2007)).

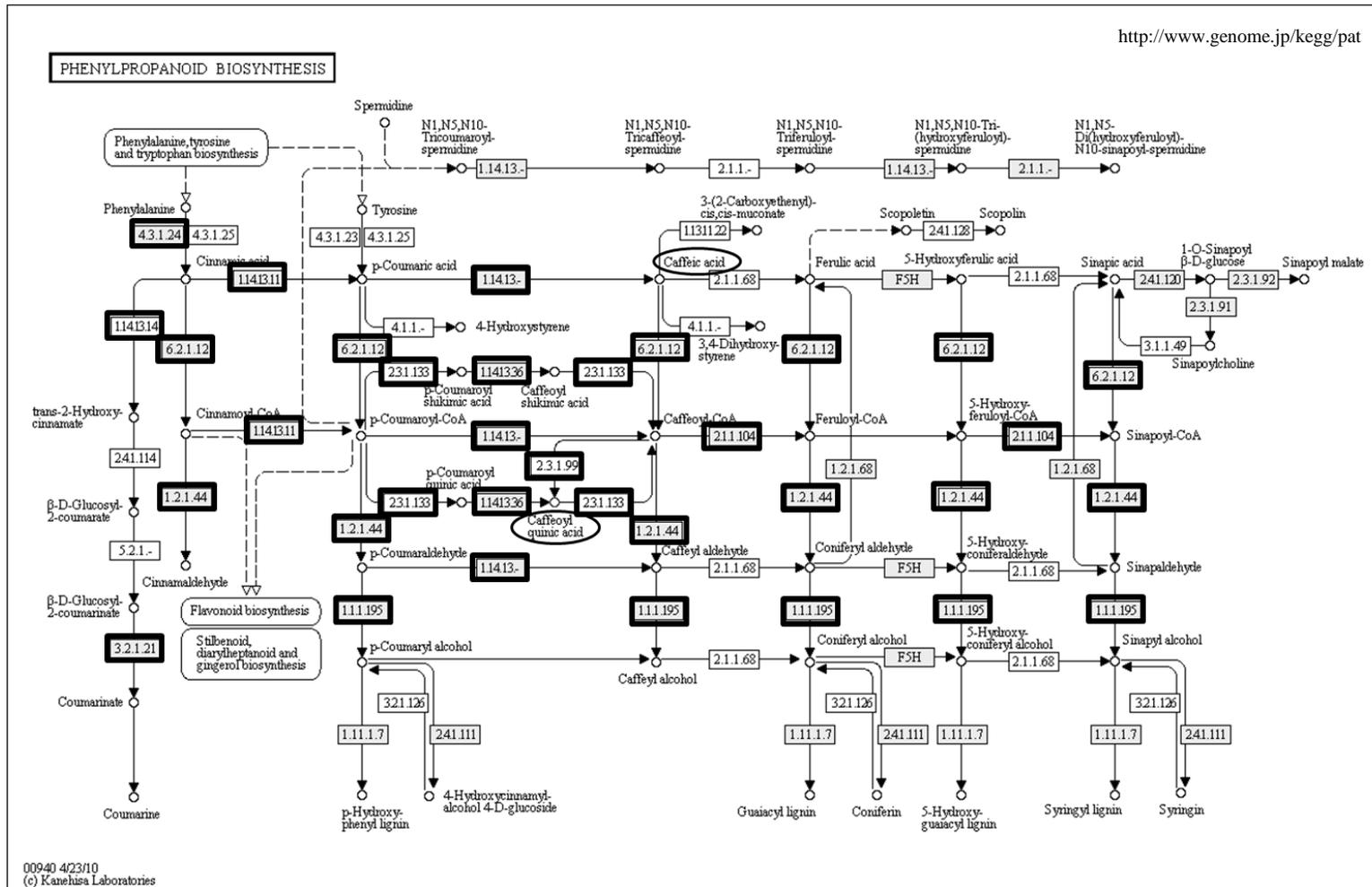


Figure 2. Voie de biosynthèse des phénylpropanoïdes (KEGG).

- Code Enzyme encadrée:
- 4.3.1.24 : phenylalanine ammonia-lyase (PAL)
 - 1.14.13.11 : cinnamate-4-hydroxylase (C4H)
 - 1.14.13.- : p-coumarate 3-hydroxylase (C3H)
 - 6.2.1.12 : 4-coumarate-CoA-ligase (4CL)
 - 1.14.13.14 : cinnamate-4-hydroxylase (C4H)
 - 3.2.1.21 : beta-xylosidase, acyl activating enzyme (AAE)
 - 1.2.1.44 : cinnamoyl-CoA reductase (CCR)
 - 1.1.1.195 : cinnamyl-alcohol-dehydrogenase (CAD)
 - 1.14.13.36 : p-coumarate 3-hydroxylase (C3H)
 - 2.3.1.99 : hydroxycinnamoyl-CoA-quininate-transferase (HQT)
 - 2.3.1.133 : hydroxycinnamoyl-CoA-shikimate/quininate transferase (HCT)
 - 2.1.1.104 : caffeoyl-CoA 3-o-methyltransferase (CCoAOMT)

Les cadres en gras indiquent les enzymes dont les séquences étaient disponibles et qui ont ainsi pu servir à la recherche de gènes candidats.

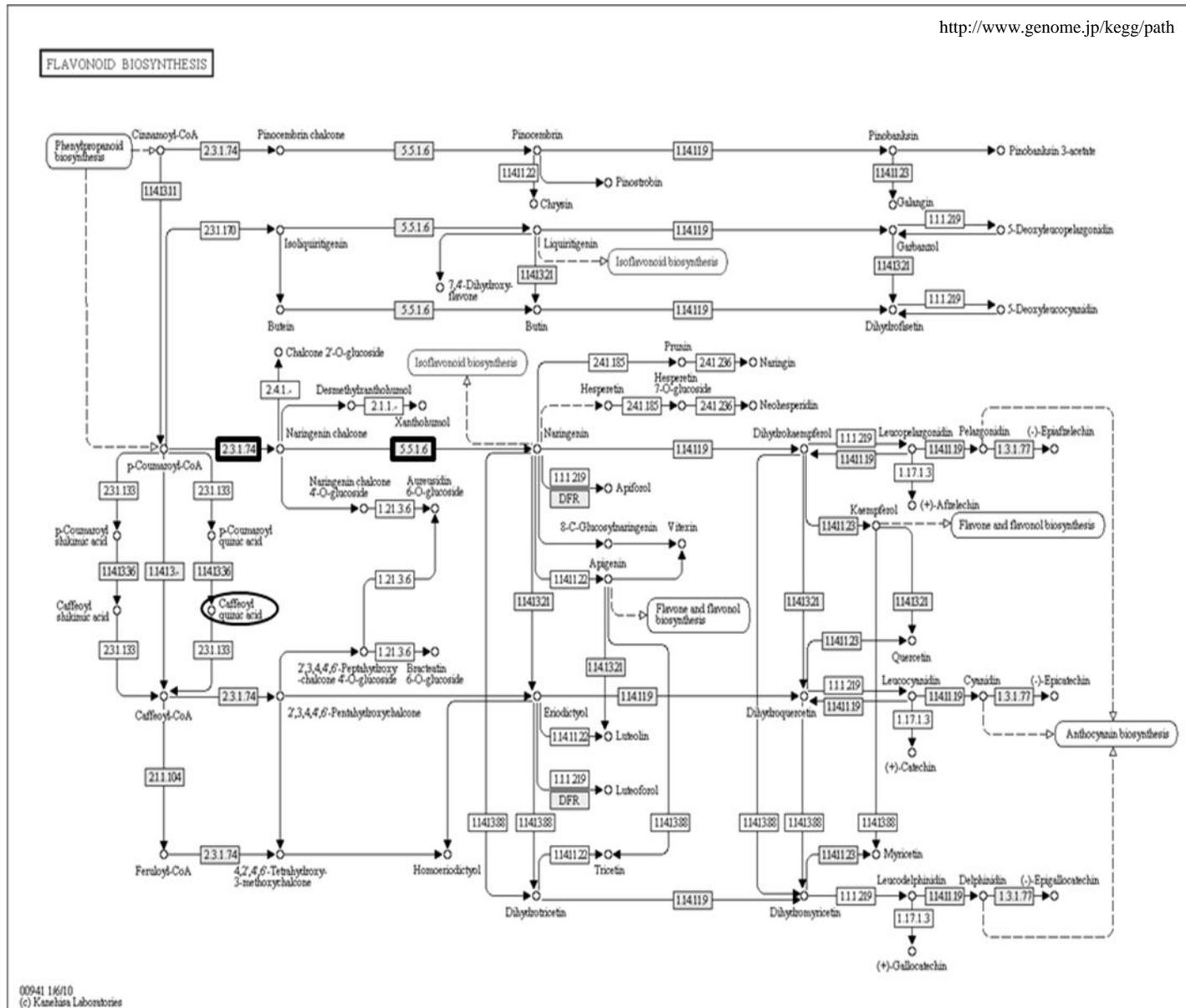


Figure 3. Voie de biosynthèse des flavonoïdes (KEGG).

Code Enzyme encadrée:

2.3.1.74 : chalcone synthase (CHS)

5.5.1.6 : chalcone isomerase (CHI)

Les cadres en gras indiquent les enzymes dont les séquences étaient disponibles et qui ont ainsi pu servir à la recherche de gènes candidats.

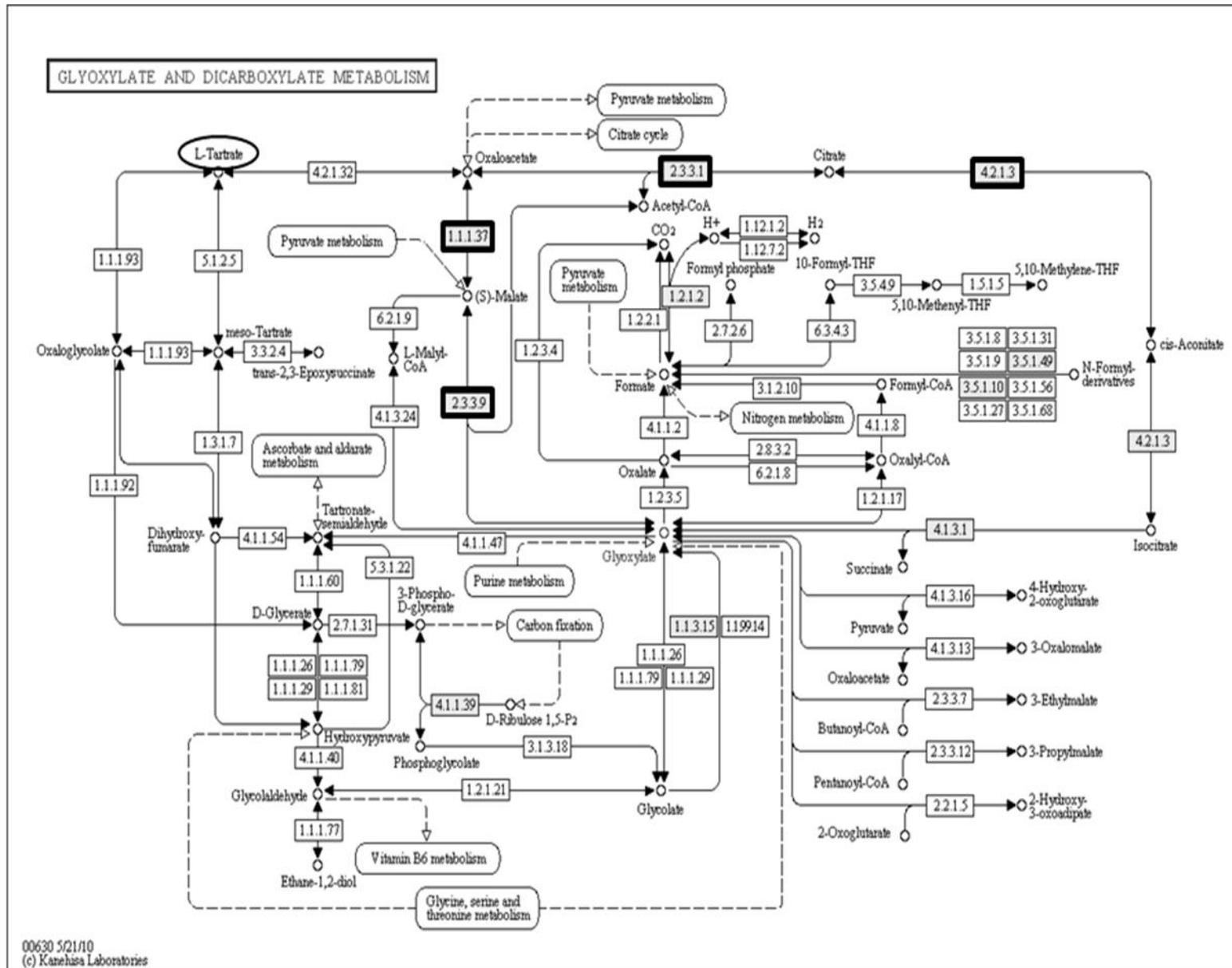


Figure 5. Voie de biosynthèse du glyoxylate et du dicarboxylate (KEGG).

Code Enzyme encadrée:
 2.3.3.1 : citrate synthase (CSY)
 1.1.1.37 : malate dehydrogenase (MDH)
 2.3.3.9 : malate synthase (MLS)
 4.2.1.3 : aconite hydratase/citrate-hydrolyase (ACO)

Les cadres en gras indiquent les enzymes dont les séquences étaient disponibles et qui ont ainsi pu servir à la recherche de gènes candidats.

III.2. Définition des amorces

La base de données de l'université de Davis inclut une comparaison des EST de chicorée avec le génome d'*Arabidopsis thaliana* et donne les références des gènes homologues trouvés. Les séquences EST de chicorée ont été alignées aux séquences ADN et ARNm d'*Arabidopsis* (<http://www.ncbi.nlm.nih.gov/BLAST/>) grâce au logiciel BioEdit v.7.0.5.3 (Hall, 1999), afin d'identifier les positions de régions introniques chez la chicorée. En effet, pour des gènes conservés, l'emplacement des introns est moins soumis à la sélection naturelle (Rogozin *et al.*, 2003). Ceci a été conforté par le fait de trouver des séquences spécifiques aux bordures exon/intron dans les EST de chicorée là où les séquences d'*Arabidopsis* suggéraient la présence d'un intron.

Des amorces flanquant ces introns suggérés ont été définis à partir des séquences EST de chicorée. Si l'EST de chicorée ne couvrait pas l'intron, des amorces ont été définis dans l'exon (logiciel Primer 3 v0.3.0 (Rozen & Skaletsky, 1999)).

III.3. Recherche de polymorphisme

III.3.A. Polymorphisme SSCP

Les réactions PCR sont effectuées dans un volume final de 15 μL en prenant 2 μL d'ADN à 2 $\text{ng}/\mu\text{L}$. Les concentrations finales correspondent à : tampon PCR x 1, 2mM MgCl_2 , 0,3 U/ μL de Taq Polymérase (Applied Biosystems), et 100 μM de chaque dNTP (Promega). Au final, les amorces sont concentrées à hauteur de 166 nM, que ce soit pour l'amorce marquée (fluorochrome bleu 6-fam ou vert Hex), ou pour l'amorce non marquée (Sigma-Aldrich).

Les conditions PCR établies pour ces couples d'amorces se distinguent ainsi:

Dénaturation : 4 minutes à 94°C	} 5 cycles
Dénaturation : 30 secondes à 94°C	
Hybridation : 30 secondes à T_m	
Diminution de -1°C par cycle	
Elongation : 30 secondes à 72°C	

Dénaturation : 30 secondes à 94°C	}	35 cycles
Hybridation : 30 secondes à 58°C		
Elongation : 30 secondes à 72°C		
Extension terminale : 7 minutes à 72°C		

1 µL du produit d'amplification est ajouté à 10 µL de formamide, 0,5 µL de NaOH 0,3N et entre 0,1 et 0,4 µL de marqueur de taille interne. Après une dénaturation de 4 minutes à 95°C, les produits PCR sont séparés sur le séquenceur capillaire (ABI PRISM® 3100-Avant) à 18 ou 25°C (polymère POPTM Conformational Analysis Polymer). Les résultats de ces migrations sont ensuite analysés avec le logiciel GeneMapper®v.4.0.

III.3.B. Polymorphisme HRM

Le principe de la technique HRM (High Resolution Melting) est de visualiser des différences de stabilité d'allèles causées par des changements nucléotidiques. Lors de la PCR, l'intercalation d'un fluorochrome spécifique de l'ADN double brin permet de suivre l'amplification et ensuite la dénaturation. Après la PCR, lors d'une phase d'augmentation de la température, les deux brins d'ADN se séparent et l'intensité de la fluorescence décroît. On peut ainsi suivre les courbes de dénaturation. Les individus homozygotes montrent des courbes d'allure identique alors que les individus hétérozygotes se distinguent par des courbes à l'allure différente. Plus il y a de combinaison d'allèles différents, plus il y a de courbes de fusion différentes. La technique HRM est réalisée sur la machine HRM Rotor-Gene Q (Qiagen) avec le kit LightCycler®480 High Resolution Melting Master (Roche). Les mix PCR sont effectués comme le montre le tableau 2.

solution kit	volume pour 15µl
2 x master mix	7,5 µl
25 mM MgCl ₂	1,2 µl
amorce F 5 µM	0,6 µl
amorce R 5 µM	0,6 µl
H ₂ O	qsp 15 µl

Tableau 2. Mix PCR utilisé pour la technique HRM.

Le master mix contient une polymérase « hotstart », du tampon, des nucléotides et un fluorochrome. La réaction PCR se déroule selon les étapes suivantes :

Dénaturation : 5 minutes à 95°C
 Dénaturation : 30 secondes à 94°C
 Hybridation : 30 secondes à 63°C
 Diminution de -1°C par cycle
 Elongation : 5 minutes à 72°C

} 5 cycles

Dénaturation : 30 secondes à 94°C
 Hybridation : 30 secondes à 58°C
 Elongation : 30 secondes à 72°C

} 35 cycles

Extension terminale : 7 minutes à 72°C
 Elongation finale : 5 minutes à 72°C
 Dénaturation : 1 minute à 95°C
 Hybridation : 1 minute à 40°C
 Lente dénaturation de 65 à 95°C (observation des différentes courbes).

IV. Résultats

IV.1. Recherche de gènes candidats

Le tableau 3 liste les gènes candidats étudiés, selon leur implication dans les différentes voies de métabolisme d'intérêt (phénylpropanoïdes, flavonoïdes, acide ascorbique, glyoxylate). Les différentes voies de biosynthèse (selon <http://www.genome.jp/kegg/pathway.html>) utilisées pour la recherche de gènes candidats sont illustrées dans figures 2, 3, 4 et 5.

Parmi ces 30 gènes candidats, quatre codent pour des facteurs de transcription. Pour un dixième des gènes candidats, aucune séquence n'a été trouvée chez *Cichorium* dans la banque EST de l'Université de Davis. Parmi les 27 gènes candidats ayant une séquence chicorée correspondante et pour lesquels des amorces ont pu être définies, huit n'ont pas présenté de polymorphisme, ni en SSCP ni en HRM. Pour les 19 gènes candidats restants (polymorphes), 15 ont été cartographiés sur la carte « oxychic ». Leur localisation est détaillée dans les paragraphes suivants. Les amorces utilisées pour les 19 gènes candidats présentant du polymorphisme sont disponible dans l'annexe 3.

Parmi les marqueurs disponibles sur la carte de référence, OC0603 correspond à un gène codant l'enzyme cobalamine-indépendante méthionine synthase aussi appelée

homocystéine S-méthyltransferase (HMT, EC 2.1.1.14,). L'enzyme HMT catalyse la réaction entre la L-homocystéine et la L-méthionine. La L-méthionine est ensuite catalysée en S-Adenosyl-L-méthionine par la S-adenosyl-L-méthionine synthétase. Le gène correspondant à cette enzyme n'a pas pu être cartographié car aucun polymorphisme n'a été détecté (cf. tableau 3, gène candidat *SAM*). Ces réactions font partie de la voie de biosynthèse de la méthionine, impliquée dans la formation de substrat pour la voie des phénylpropanoïdes, plus exactement au niveau de deux *O*-méthyltransférases : CCoAOMT et OMT qui catalysent la formation de féruloyl CoA à partir du caféoyl CoA et la formation de sinapaldéhyde à partir de la 5-hydroxyconiféraldéhyde, respectivement (Kirst *et al.*, 2004).

CHAPITRE II

nom	gène codant l'enzyme	Code enzyme	voie métabolique	polymorphisme	séquence <i>Arabidopsis</i> correspondant à celles de <i>C. intybus</i>
ADT	arogenate hydratase/prephenate	4.2.1.91	Amont voie phénylpropanoïdes	non	AT3G07630
4CL	4-coumarate-CoA-ligase	6.2.1.12	phénylpropanoïdes	SSCP	AT4G05160, AT1G51680, AT5G63380, AT3G21240
AAE	beta-xylosidase, acyl activating enzyme	3.2.1.21	phénylpropanoïdes	SSCP	AT5G49360, AT1G20560, AT4G34230
C3H	<i>p</i> -coumarate 3-hydroxylase (CYP98A3)	1.14.13.-/36	phénylpropanoïdes	SSCP	AT2G40890
C4H	cinnamate-4-hydroxylase	1.14.13.11/14	phénylpropanoïdes	SSCP	AT2G30490
CAD	cinnamyl-alcohol-déhydrogénase	1.1.1.195	phénylpropanoïdes	SSCP	AT3G19450, AT1G68540
CCoAOMT	cafféoyl-CoA 3- <i>o</i> -méthyltransférase	2.1.1.104	phénylpropanoïdes	non	AT5G48930
CCR	cinnamoyl-CoA réductase	1.2.1.44	phénylpropanoïdes	SSCP	AT1G15950, AT1G80820
ExpProt	Protéine impliquée dans la voie des phénylpropanoïdes	-	phénylpropanoïdes	SSCP	AT2G48110
HCT	hydroxycinnamoyl-CoA-Shikimate/quinate transférase	2.3.1.133	phénylpropanoïdes	HRM	AT5G48930
HQT	hydroxycinnamoyl-CoA-quinate-transférase	2.3.1.99	phénylpropanoïdes	pas de séquence correspondante dans la banque chicorée	
MYB4	facteur de transcription type MYB (répression C4H)	-	phénylpropanoïdes	non	AT5G26660, AT4G38620
NAD	NADPH- P450 cytochrome réductase	1.6.2.4	phénylpropanoïdes	SSCP	AT4G30210
PAL	phénylalanine ammonia-lyase	4.3.1.24	phénylpropanoïdes	non	AT2G37040, AT3G53260
SAM	S-adenosyl-L-méthionine synthétase	2.5.1.6	phénylpropanoïdes	non	AT3G17390, AT2G36880
Anthoc	anthocyanin transférase family protein	2.3.1.-	flavonoïdes	HRM	AT1G03940
CHI	chalcone isomérase	5.5.1.6	flavonoïdes	SSCP	AT3G55120, AT3G6620, AT1G53520, AT5G05270
CHS	chalcone synthase	2.3.1.74	flavonoïdes	SSCP	AT5G13930
MYB12	facteur de transcription de la CHS	-	flavonoïdes	SSCP	AT2G47460
PAP1	facteur de transcription MYB, production of anthocyanin pigment	-	flavonoïdes	non	AT2G16430, AT2G27190
PAP2	facteur de transcription MYB, production of anthocyanin pigment	-	flavonoïdes	non	AT3G50920, AT3G54020, AT1G51110
AOX	ascorbate oxydase	1.10.3.3	dégradation ascorbate	SSCP	AT1G55570, AT5G21105, AT5G21100
APX	ascorbate peroxydase	1.11.1.11	dégradation ascorbate	SSCP	AT1G07890, AT1G77490, AT4G09010, AT4G35000
GLDH	L-gulono-1,4-lactone déhydrogénase/galactonolactone déhydrogénase	1.3.2.3	dégradation ascorbate	pas de séquence correspondante dans la banque chicorée	
ID	idonate déhydrogénase	1.1.1.-	dégradation ascorbate	pas de séquence correspondante dans la banque chicorée	
MDA	monodéhydroascorbate réductase	1.6.5.4	dégradation ascorbate	SSCP	AT3G27820, AT3G52880
MLS	Malate synthase	2.3.3.9	glyoxylate	non	AT5G03860
ACO	aconite hydratase/citrate-hydrolyase	4.2.1.3	glyoxylate	SSCP	AT2G05710, AT4G26970, AT4G35830
CSY	citrate synthase	2.3.3.1	glyoxylate	SSCP	AT3G58740, AT3G58750, AT2G42790, AT3G60100, AT2G44350
MDH	malate déhydrogénase	1.1.1.37	glyoxylate	SSCP	AT1G04410, AT1G53240, AT2G22780, AT5G09660, AT3G15020, AT3G47520, AT5G56720, AT5G43330, AT5G58330

Tableau 3. Résultat de la recherche de gènes candidats pour cartographie sur la carte « oxychic ». Trois gènes candidats sur 30 n'ont pas de séquence correspondante dans la banque EST chicorée (<http://cgpdb.ucdavis.edu/>). Huit gènes candidats sur 27 n'ont pas présenté de polymorphisme SSCP ou HRM avec les amorces définies comme expliqué dans le paragraphe III.2.

IV.2. Construction de la carte « oxychic »

Les marqueurs sélectionnés pour construire la carte du projet « oxychic » ont été choisis en fonction du type de ségrégation et de leur emplacement sur les groupes de liaison (LG), afin de recouvrir de la façon la plus régulière possible les neuf LG de *Cichorium intybus* L., (Cadalen *et al.*, 2010). Comme 111 individus étaient préalablement génotypés dans le cadre de la thèse d'Aline Clabaut, nous avons repris, pour les 90 individus restants, essentiellement les mêmes marqueurs et nous en avons rajouté/supprimé lorsque cela était nécessaire pour la construction de la carte « oxychic ». Quinze marqueurs supplémentaires ont pu être placés suite à la recherche de gènes candidats. Quatre gènes candidats n'ont pas pu être placés correctement. *Anthoc* ferait partie du LG9 mais sa localisation reste difficile à déterminer. Les gènes candidats *CHI*, *CHS* et *MDA* font partie des LG1, LG7 et LG8, respectivement. Ces marqueurs ont présenté une lecture de génotypage difficile, générant ensuite une ambiguïté quant à l'exactitude de leur placement dans le groupe de liaison. En regardant à quels autres marqueurs de la carte ils sont le plus liés génétiquement, il a été possible de connaître leur localisation la plus probable. Ils ont donc été rajoutés en retrait sur la carte « oxychic », présentée dans la figure 6.

Ainsi, cette nouvelle carte génétique (figure 6) est construite sur la base de 142 marqueurs dont 58 sont dominants. On compte 44 de type <lm x ll>, 44 de type <nn x np>, 5 de type <hk x hk>, 24 de type <ef x eg>, 16 de type <hk x hk> (h- ; kk) et 9 de type <ab x cd>, dont le polymorphisme a été détecté par SSR (51 marqueurs), SSCP (32), AFLP (55) et HRM (4). On retrouve une moyenne de 15,77 marqueurs par groupe de liaison (minimum 11, maximum 21), définissant un intervalle moyen de 6,08 cM (distance maximale entre les marqueurs FDCB98b/EU07B09 et OC2060 : 21 cM, LG8). On remarque des zones particulièrement concentrées en marqueurs, aussi appelées clusters (LG3, LG6, LG8, LG9).

CHAPITRE II

	LG1	LG2	LG3	LG4	LG5	LG6	LG7	LG8	LG9	Total
SSR	6	7	6	4	10	5	6	5	7	56
SSCP	4	0	3	1	5	2	2	7	3	27
AFLP	5	5	5	5	5	9	6	7	8	55
HRM	1	0	1	1	1	0	0	0	0	4
Total marqueurs	15	12	15	11	21	16	14	19	18	142
longueur (cM)	101	74	88	72	77	93	55	92	80	81,3
intervalle moyen (cM)	7,77	6,73	6,28	8	4,05	6,2	4,23	6,13	5,33	6,08

Tableau 4. Nombre, type et intervalle moyen entre deux marqueurs, pour chaque groupe de liaison de la carte « oxychic ».

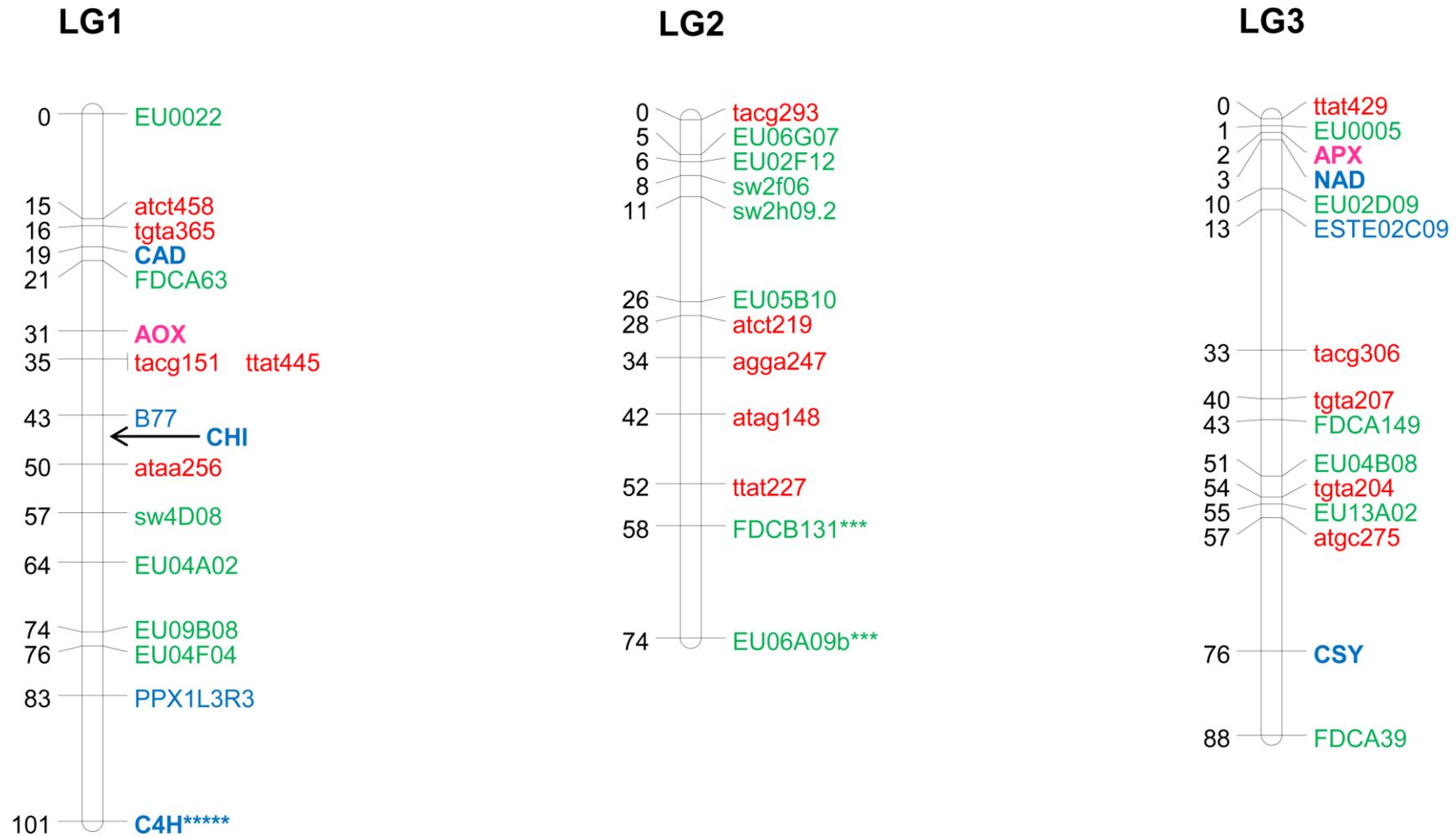


Figure 6. Carte « oxychic » K59xK28, 9 groupes de liaisons, 142 marqueurs moléculaires (les gènes candidats sont en gras). En rouge, les marqueurs AFLP ; en vert, les marqueurs SSR ; en rose, les marqueurs HRM ; en bleu, les marqueurs SSCP. Les marqueurs présentant des distorsions sont accompagnés d'astérisques. Le nombre d'astérisques indique le niveau de probabilité associé au test χ^2 : * = $p < 0,05$; ** = $p < 0,01$; *** = $p < 0,005$; **** = $p < 0,001$; ***** = $p < 0,0005$.

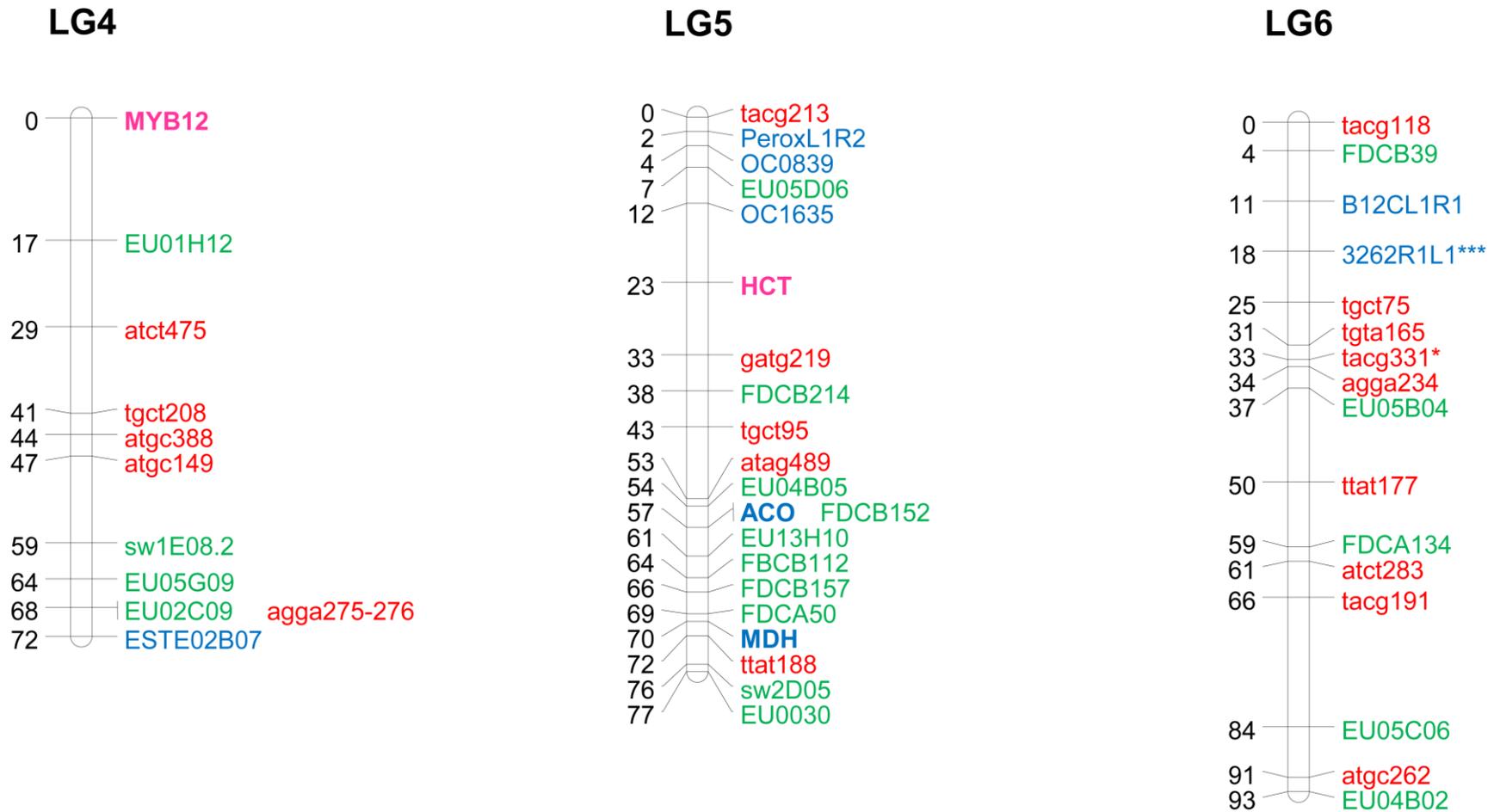


Figure 6 (suite). Carte « oxychic » K59xK28, 9 groupes de liaisons, 142 marqueurs moléculaires (les gènes candidats sont en gras). En rouge, les marqueurs AFLP ; en vert, les marqueurs SSR ; en rose, les marqueurs HRM ; en bleu, les marqueurs SSCP. Les marqueurs présentant des distortions sont accompagnés d'astérisques. Le nombre d'astérisques indique le niveau de probabilité associé au test χ^2 : * = $p < 0,05$; ** = $p < 0,01$; *** = $p < 0,005$; **** = $p < 0,001$; ***** = $p < 0,0005$.

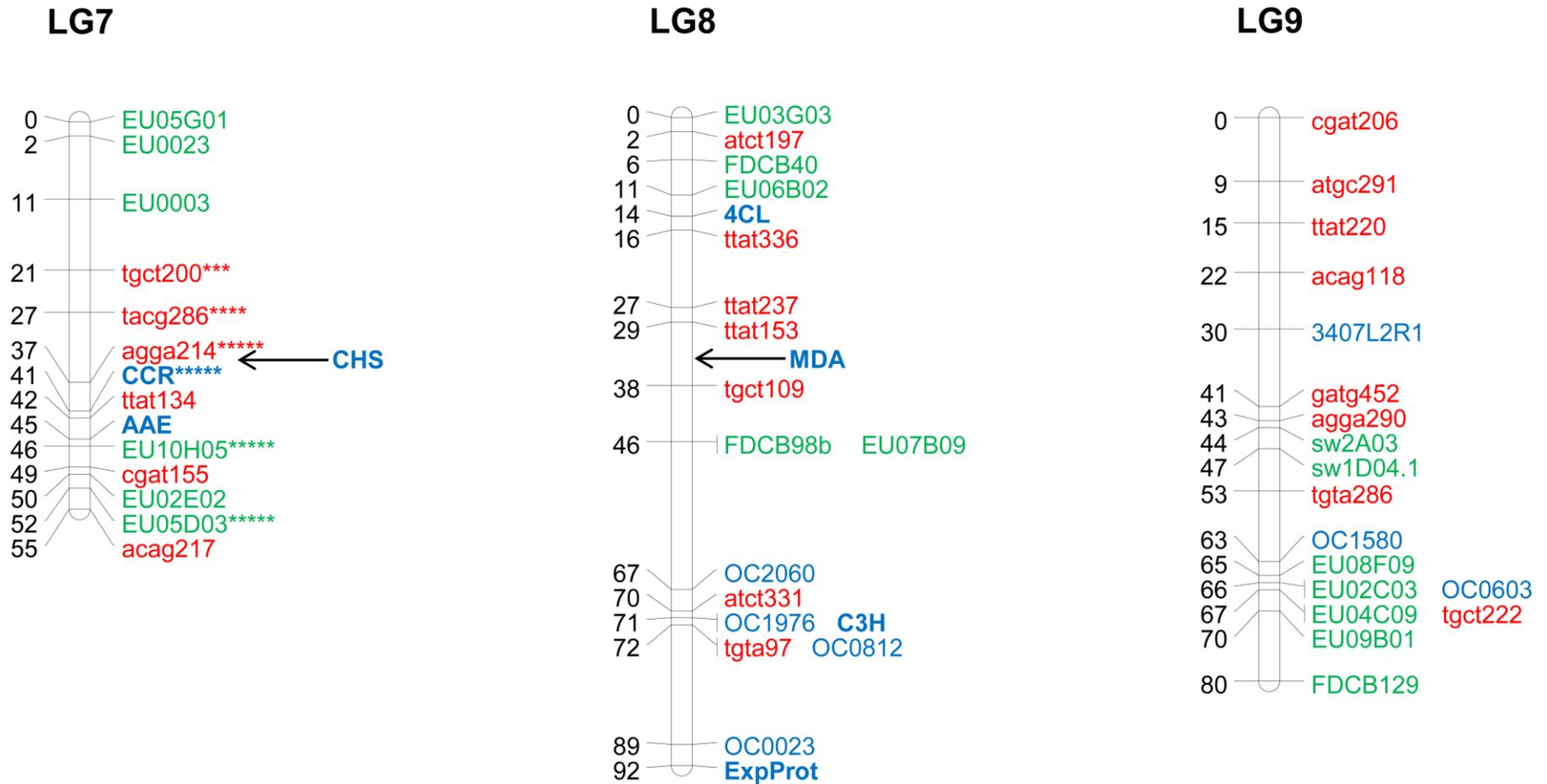


Figure 6 (suite). Carte « oxychic » K59xK28, 9 groupes de liaisons, 142 marqueurs moléculaires (les gènes candidats sont en gras). En rouge, les marqueurs AFLP ; en vert, les marqueurs SSR ; en rose, les marqueurs HRM ; en bleu, les marqueurs SSCP. Les marqueurs présentant des distorsions sont accompagnés d'astérisques. Le nombre d'astérisques indique le niveau de probabilité associé au test χ^2 : * = $p < 0,05$; ** = $p < 0,01$; *** = $p < 0,005$; **** = $p < 0,001$; ***** = $p < 0,0005$.

IV.3. Ordre des marqueurs

Par rapport à la carte de référence K28xK59 (Cadalen *et al.*, 2010), l'ordre des marqueurs de la carte « oxychic » est sensiblement le même.

On remarque quelques inversions de marqueurs très proches : entre atct458 et tgta365 (LG1), entre EU09B08 et EU04F04 (LG1), entre EU06GO7 et EU02F12 (LG2), entre EU05B10 et atct219 (LG2), entre ttat227 et FDCB131 (LG2), entre tacg118, FDCB39 et B12CL1R1 (LG6), entre tacg331, agga234 et EU05B04 (LG6), entre atct283 et tacg191 (LG6), entre atgc262 et EU04B02 (LG6), entre EU05G01, EU0023 et EU0003 (LG7), et entre atct197 et FDCB40 (LG8).

Certains marqueurs co-localisent sur la carte de référence mais pas sur la carte « oxychic »: tgta204 et FDCA149 (LG3), EU13A02 et atgc 275 (LG3), FDCB112 et FDCB157 (LG5), FDCA50 et ttat188 (LG5), 3262R1L1 et tgct75 (LG6), cgat155 et EU02E02 (LG7), OC1976, OC2060 et tgta97 (LG8). Ces marqueurs restent néanmoins très accolés, entre 1 et 3 cM sauf pour 3262R1L1 et tgct75 (LG6) qui sont éloignés de 7 cM sur la carte « oxychic ».

A l'inverse, des marqueurs co-localisent sur la carte « oxychic » et non sur la carte de référence bien qu'ils soient proches: EU02C09 et agga275-276 (LG4), FDCB98b et EU07B09 (LG8), à 1 et 7 cM près, respectivement.

Finalement, le groupe de liaison présentant le plus de différences avec la carte de référence est le LG6. Cela est sans doute dû au fait qu'il s'agisse du LG qui présente le plus de clusters : 3 selon la carte de référence K28 x K59 (Cadalen *et al.*, 2010).

Le LG9 n'a pas pu être comparé à celui présenté dans la publication de Cadalen *et al.* (2010) car l'ordre des marqueurs est complètement différent. En effet, ce LG présentait beaucoup de marqueurs à seulement deux allèles et informatifs pour un seul des parents (de type <nn x np> et <lm x ll>). Les autres types de marqueurs (<hk x hk>, <ef x eg>, <ab x cd>) étant réduits, deux solutions étaient proposées sur la base de l'algorithme de cartographie utilisé par logiciel JoinMap v 3.0. La première option consiste à regrouper les marqueurs type <hk x hk>, <ef x eg> et <ab x cd> au centre du LG et répartir les marqueurs type <nn x np> d'un côté et les marqueurs type <lm x ll> de l'autre (comme dans la publication). La seconde option est aussi celle qui a été trouvée lors de ces travaux de thèse : les marqueurs de type <hk

x hk>, <ef x eg> et <ab x cd> sont regroupés dans la même région et les marqueurs <nn x np> et <lm x ll> sont mélangés. La figure 7 illustre ces propos.

La première option avait été choisie par l'équipe de Cadalen *et al.* (2010) car un locus semblait être dupliqué (AF497999L2/R2 et AF497999L3/R3). Il s'agissait de deux marqueurs produits à partir de la séquence EST de *Cichorium* correspondant au gène *germacrene A synthase* (forme longue). De plus, le LG9 ayant très peu de marqueurs « ancre », reliant les intervalles communs entre K28xK59 et Rubis 118 ou K28xK59 et BR, les approches comparatives ne permettaient pas de faire un choix entre les solutions 1 et 2 présentées dans la figure 7. Après la publication de l'article, le génotypage a été recommencé et les résultats ont contredit la duplication du locus AF497999 et conforté l'hypothèse d'un gène unique et donc de la solution 2 (figure 7).

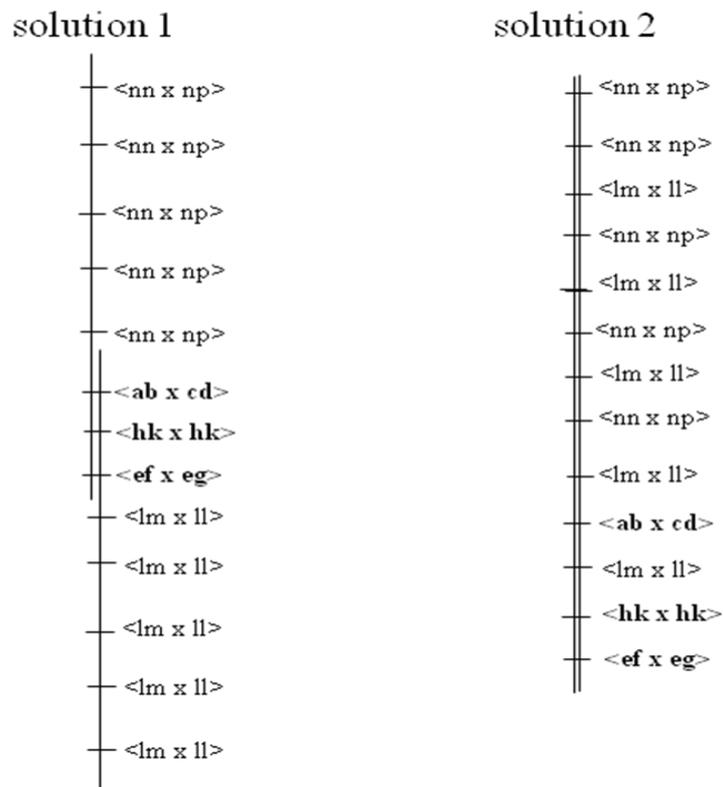


Figure 7. Illustration des deux types de solutions proposé par JoinMap v 3.0 pour le placement des marqueurs moléculaires du LG9 sur la carte de référence. Les marqueurs informatifs pour les deux parents (en gras) permettent de faire le lien entre les marqueurs informatifs pour un seul des deux parents (<lm x ll> et <nn x np>).

IV.4. Distorsion

Onze marqueurs liés à des loci en distorsion de ségrégation (LDS) présentent des écarts plus ou moins importants par rapport aux effectifs attendus. Ces informations sont

visibles sur la figure 6 : le nombre d'étoiles adjacentes au nom du marqueur indique le niveau de la distorsion (test χ^2).

Parmi les 56 marqueurs SSR utilisés pour la carte « oxychic », quatre montrent une distorsion de ségrégation (7,1%), quatre parmi les 55 marqueurs AFLP (7,3%) et trois sur les 27 marqueurs SSCP (11,1%).

Ces distorsions, particulièrement fortes sur le LG7, ont également été remarquées lors des travaux de thèse sur l'embryogenèse somatique (Aline Clabaut, thèse 2009) dont la carte génétique a été construite à partir d'individus de la population F1'a (54%) et de la population F1'b (46%), comme expliqué dans la figure 1. Ainsi, dans le cas de cette carte comme dans le cas de la notre (47% de F1'a et 53% de F1'b), c'est la sous-descendance F1'b qui semble apporter ces loci en distorsion de ségrégation pour le LG7. Le tableau 5 liste les marqueurs biaisés de notre carte et compare les résultats des tests χ^2 significatifs avec les tests obtenus pour ces marqueurs sur la carte de référence K28xK59. Les marqueurs distordus de la carte « oxychic » ne le sont pas ou très peu sur la carte de référence. A contrario, notons que la carte de référence K28xK59 possède des marqueurs distordus (maximum *** c'est-à-dire $p < 0,005$) alors que ces mêmes marqueurs ne présentent pas de distorsion pour la carte « oxychic ». C'est en effet le cas pour EU02F12*, sw2f06*, sw2h09.2* sur le LG2, ttat134** sur le LG7, EU03G03* et ttat336* sur le LG8, EU02C03**, sw2A03**, cgat206*** sur le LG9.

LG	Marqueur	Type	Position (cM)	Ségrégation	Probabilité associée au test χ^2 pour les cartes :	
					« oxychic »	de référence K28xK59
LG1	C4H	SSCP	100,81	<ef x eg>	*****	non génotypé
LG2	FDCB131	SSR	57,723	<nn x np>	***	-
LG2	EU06A09b	SSR	73,587	<nn x np>	***	-
LG6	3262R1L1	SSCP	18,033	<ab x cd>	***	*
LG6	tacg331	AFLP	33,176	<nn x np>	*	-
LG7	tgct200	AFLP	21,37	<lm x ll>	***	-
LG7	tacg286	AFLP	26,67	<lm x ll>	*****	-
LG7	agga214	AFLP	36,98	<lm x ll>	*****	-
LG7	CCR	SSCP	41,02	<lm x ll>	*****	non génotypé
LG7	EU10H05	SSR	45,84	<ab x cd>	*****	-
LG7	EU05D03	SSR	51,60	<ef x eg>	*****	-

Tableau 5. Marqueurs présentant des distorsions sur l'ensemble des neuf groupes de liaison (7,7%). Les astérisques indiquent le niveau de probabilité associé au test χ^2 : * = $p < 0,05$; ** = $p < 0,01$; *** = $p < 0,005$; **** = $p < 0,001$; ***** = $p < 0,0005$.

Pour le LG2, ce sont des marqueurs de type <nn x np> (informatif pour le parent K28) qui présentent des effectifs distordus ($p < 0,005$) alors que pour le LG1 et LG6 (sauf tacg331),

se sont des marqueurs hétérozygotes chez les deux parents (<ef x eg> et <abxcd>) qui reflètent la distorsion. Dans le cas du LG7, quatre marqueurs informatifs pour le parent K59 <lm x ll> et deux types de marqueurs informatifs chez les deux parents (<ef x eg> et <ab x cd>) montrent une distorsion des effectifs trouvés par rapport aux effectifs attendus très importante (de $p < 0,005$ à $p < 0,0005$).

Il est possible de savoir quel allèle et donc quel parent possède le locus responsable de la distorsion de ségrégation en comparant les effectifs attendus aux effectifs obtenus (tableau 6). Dans le cas des marqueurs de type <nn x np> EU06A09b et tacg331, l'allèle associé au génotype en sous-effectif est l'allèle « p », porté par le parent K28. A contrario, pour FDCB131, l'allèle « p » de K28 serait impliqué dans la hausse des effectifs obtenus. Pour le marqueur EU05D03 de type <ef x eg>, on remarque que les classes « ef » et « fg » ont des effectifs faibles comparés aux classes « ee » et « eg ». L'allèle « f » apporté par le parent K59 semble donc impliqué dans cette baisse d'effectif. Sur le même groupe de liaison, le marqueur CCR <lm x ll> montre aussi un effectif faible rattaché à l'allèle « m » de K59. Pour le marqueur C4H, c'est l'allèle « g » porté par K28 qui serait impliqué dans la baisse des effectifs. Un raisonnement analogue appliqué aux marqueurs tacg200 (<lm x ll>), tacg286 (<lm x ll>), agga214 et EU10H05 (<ab x cd>) permet d'identifier les allèles « m », « m », « m » et « b » respectivement, comme impliqués dans l'augmentation des effectifs attendus. Ces allèles sont portés par le parent K59, sur le même chromosome (tableau 7). Pour les marqueurs 3262R1L1, les informations concernant les effectifs faibles ou forts ne permettent pas non plus de conclure.

CHAPITRE II

LG	Marqueur	Allèles parentaux	Ségrégation	ac	ad	bc	bd	ee	ef	eg	fg	hh	hk	kk	h- k-	ll	lm	nn	np	--	<i>p</i>	effet des allèles parentaux
1	C4H	ef x eg	1:1:1:1					43	76	45	29									8	*****	g K28 ↘
								<i>48</i>	<i>48</i>	<i>48</i>	<i>48</i>											
2	FDCB131	nn x np	1:1															77	119	5	***	p K28 ↗
																		<i>98</i>	<i>98</i>			
2	EU06A09b	nn x np	1:1															113	73	15	***	p K28 ↘
																		<i>93</i>	<i>93</i>			
6	3262R1L1	ab x cd	1:1:1:1	22	44	54	33													48	***	non déterminé
				<i>38</i>	<i>38</i>	<i>38</i>	<i>38</i>															
6	tacg331	nn x np	1:1															112	79	10	*	p K28 ↘
																		<i>95</i>	<i>95</i>			
7	tgct200	lm x ll	1:1															73	119	9	***	m K59 ↗
																		<i>96</i>	<i>96</i>			
7	tacg286	lm x ll	1:1															69	122	10	****	m K59 ↗
																		<i>95</i>	<i>95</i>			
7	agga214	lm x ll	1:1															53	133	15	*****	m K59 ↗
																		<i>93</i>	<i>93</i>			
7	CCR	lm x ll	1:1															134	58	9	*****	m K59 ↘
																		<i>96</i>	<i>96</i>			
7	EU10H05	ab x cd	1:1:1:1	24	30	62	77													8	*****	b K59 ↗
				<i>48</i>	<i>48</i>	<i>48</i>	<i>48</i>															
7	EU05D03	ef x eg	1:1:1:1					54	31	75	31									10	*****	f K59 ↘
								<i>48</i>	<i>48</i>	<i>48</i>	<i>48</i>											

Tableau 6. Fréquence de distribution des génotypes par classe pour les marqueurs présentant des effectifs distordus par rapport aux effectifs attendus. Les astérisques indiquent le niveau de probabilité associé au test χ^2 : * = $p < 0,05$; ** = $p < 0,01$; *** = $p < 0,005$; **** = $p < 0,001$; ***** = $p < 0,0005$. Les effectifs attendus sont indiqués en gris et en italique. La colonne « -- » indique le nombre de données manquantes sur 201 individus.

Marqueurs LG7	K59		K28	
EU05G01	k	h	h	k
EU0023	b	a	d	c
EU0003	n	n	n	p
tgct200***	l	m	l	l
tacg286****	l	m	l	l
agga214*****	l	m	l	l
CCR*****	<u>m</u>	l	l	l
ttat134	n	n	n	p
AAE	n	n	p	n
EU10H05*****	a	b	c	d
cgat155	n	n	n	p
EU02E02	n	n	n	p
EU05D03*****	<u>f</u>	e	e	g
acag217	n	n	n	p

Tableau 7. Exemple d'un fichier excel « génotype graphique ». Les deux blocs représentent les deux chromosomes des deux parents : K59 et K28. Les marqueurs sont classés selon l'ordre de la carte « oxychic ». En gras, les allèles associés à une distribution des effectifs génotypiques en hausse par rapport aux distributions attendues. Souligné, les allèles associés à une distribution des effectifs génotypiques en baisse par rapport aux distributions attendues. Les astérisques indiquent le niveau de probabilité associé au test χ^2 : * = $p < 0,05$; ** = $p < 0,01$; *** = $p < 0,005$; **** = $p < 0,001$; ***** = $p < 0,0005$.

L'intensité des distorsions observées n'est pas proportionnelle au nombre de données manquantes (tableau 6). Ces distorsions de ségrégation ne semblent pas être responsables des petites modifications dans l'ordre des marqueurs évoqués dans le paragraphe IV.2.

Pour les marqueurs hétérozygotes chez les deux parents (<hk x hk>, <ef x eg>, <ab x cd>) un double test χ^2 permet de déterminer l'origine de la distorsion de ségrégation (Lorieux *et al.*, 1995a). Il révèle que les marqueurs C4H (LG1) et 3262R1L1 (LG6) sont liés à des LDS agissant comme des facteurs de sélection zygotiques. A contrario, les marqueurs du LG7 (EU05D03 et EU10H05), seraient associés à des LDS intervenant comme des facteurs de sélection gamétiques.

V. Discussion

Nous avons construit une carte génétique spécialement dédiée au projet « oxychic », incluant 142 marqueurs dont 15 gènes candidats à la biosynthèse des hydroxycinnamates détectés dans les feuilles de chicorée. Parmi les marqueurs, 36% présentent un polymorphisme d'unité de répétition, 22 % un polymorphisme de conformation simple brin

d'ADN, 3% un polymorphisme de courbe de fusion et 39% un polymorphisme de longueur de fragment d'amplification. Ce dernier type de marqueurs (dominants), bien que moins informatifs que les autres, permet d'augmenter rapidement la densité de la carte.

V.1. Région riche/pauvre en marqueurs moléculaires

Les neuf groupes de liaison ont en moyenne 15,8 marqueurs et un intervalle moyen entre deux marqueurs de 6,08 cM. Malgré nos efforts de couvrir de façon homogène les différents groupes de liaison, certaines zones restent peu denses, comme les zones entre les marqueurs PPX1L3R3 et C4H (LG1), entre FDCB131 et EU06A09 (LG2), entre ESTE02C09 et tacg306, entre atgc275 et CSY (LG3), entre MYB12 et EU01H12 (LG4), entre tacg191 et EU05C06 (LG6), ou encore FDCB98b/EU07B09 et OC2060 du LG8. Ces régions peu denses peuvent correspondre soit à des zones de fortes recombinaisons, soit à un manque de marqueurs polymorphes trouvés dans ces zones. Ces marqueurs qui semblent éloignés sur la carte génétique peuvent, sur une carte physique, être en fait relativement proches et vice versa (Sourdille *et al.*, 2004). Actuellement, le manque d'information concernant la carte physique de *Cichorium intybus* ne nous permet pas d'aller plus loin dans ces hypothèses.

A l'inverse, nous observons des zones riches en marqueurs, évoquées dans le paragraphe IV.I. Ce phénomène est également visible sur les cartes génétiques de la laitue (Jeuken *et al.*, 2001), de la tomate (Areshchenkova & Ganal, 1999) et chez la vigne, par exemple (Costantini *et al.*, 2008). Ces zones denses en marqueurs, donc peu sujet à la recombinaison, pourraient correspondre aux zones centromériques comme indiqué chez l'orge (Ramsay *et al.*, 1999) et chez la tomate (Areshchenkova & Ganal, 1999). L'équipe de Chen (Chen *et al.*, 2002), a également appuyé cette hypothèse en comparant une carte génétique et une carte physique du riz.

V.2. Cas particulier du groupe de liaison 9

Sur la carte « oxychic », le LG9 ne comporte que trois marqueurs hétérozygotes chez les deux parents : EU02CO3 et EU04C09 de type <hk x hk> et EU08F09 de type <ef x eg>. Sur la carte de référence, le LG9, contient également peu de marqueurs informatifs chez les deux parents par rapport aux autres LG, expliquant les deux possibilités suggérées par le logiciel JoinMap v 3.0 (figure 7) quant à l'ordre des marqueurs. Pour vérifier l'ordre

finalement choisi, il faudrait essayer d'augmenter le nombre de marqueurs de type <hk x hk>, <ef x eg> et <ab x cd>.

Pour expliquer le manque de marqueurs hétérozygotes chez les deux parents, la première hypothèse impliquerait le hasard : lors des recherches de polymorphisme, peu de marqueurs hétérozygotes pour K28 et K59 ont été trouvés. La seconde hypothèse suggère que ce groupe de liaison comporte, chez les parents, des zones importantes d'homozygotie, donnant alors peu de loci à plus de deux allèles différents. Bien que cette dernière explication paraît peu probable sachant que la chicorée est allogame stricte, le LG9 pourrait comporter des loci de domestication. Ce qui est appelé syndrome de domestication correspondrait au fait qu'il existe des régions chromosomiques impliquées dans la variation de plusieurs caractères d'intérêt agronomique (à cause du phénomène de pléiotropie ou d'un lien physique entre plusieurs gènes) et qu'ainsi, les sélections successives de ce caractère ont influencé un ou plusieurs autres caractères apparemment sans relation avec le premier (Burke *et al.*, 2002). La chicorée, sélectionnée depuis relativement longtemps pour des caractères agronomiques comme la taille de la racine par exemple, pourrait peut-être comporter des zones particulièrement homozygotes ou avec peu de diversité allélique.

Si l'on considère la première hypothèse (« hasard ») comme étant la plus probable et pour vérifier le bon placement des marqueurs du LG9, il faudrait augmenter le nombre de marqueurs informatifs pour les deux parents. Pour cela, les techniques rapides visant à accroître la part de marqueurs co-dominants, comme la SSCP ou la HRM peuvent être utilisées. Pour chercher du polymorphisme uniquement sur le LG9, nous pouvons utiliser la banque BAC de *Cichorium intybus* L. disponible au sein de notre laboratoire (Gonthier *et al.*, 2010). Les marqueurs déjà attribués au LG9 serviraient de sondes pour détecter les clones positifs de la banque BAC, c'est-à-dire les clones possédant un insert correspondant à une région du LG9. Les inserts faisant environ 100 000 pb, nous envisagerions de séquencer les extrémités sur 1000 pb par exemple. A partir de ces extrémités séquencées, il sera possible de définir des amorces et de chercher ensuite du polymorphisme SSCP ou HRM au sein des amplicons.

V.3. Distorsion de ségrégation

Le phénomène de distorsion de ségrégation, mentionné dans le paragraphe III.3, est défini comme étant une déviation des fréquences génotypiques observées par rapport aux

fréquences attendues, transgressant les lois de ségrégation et rendant les théories et analyses de la génétique conventionnelles non applicables (Lu *et al.*, 2002). Les distorsions de ségrégation observées sur la carte génétique « oxychic » sont pourtant un phénomène commun chez les plantes (Song *et al.*, 2006). Mis à part la chicorée, il a déjà été relaté chez des astéracées comme le tournesol (rétrocroisement interspécifique, (Quillet *et al.*, 1995)) et la laitue (F2, croisement intraspécifique, Kesseli *et al.*, 1994). Dans une synthèse bibliographique, Song *et al.* (2006) explique que la proportion des loci montrant une ségrégation significative varie grandement en fonction de l'espèce, le type de population et le croisement. Les descendances d'haploïdes doublés et les lignées recombinantes semblent être plus touchées par ce phénomène que les autres types de population. De même, les croisements interspécifiques semblent être plus atteints que les croisements intraspécifiques selon l'expérience de Kianian & Quiros 1992 : 59% des marqueurs présentant des effectifs distordus pour une F2 interspécifique contre 7% pour une F2 intraspécifique. Notre cas ne correspond pas au croisement ou au type de population les plus sensibles à ce phénomène : seulement 7,7% des marqueurs de la carte « oxychic » sont le reflet de LDS. D'autre part, les marqueurs co-dominants seraient moins touchés par la distorsion que les marqueurs dominants (Lorieux *et al.*, 1995a ; Lorieux *et al.*, 1995b). Cette observation ne peut pas totalement s'accorder à notre situation puisque seulement 36,4% de marqueurs aux effectifs distordus sont dominants (AFLP).

Les distorsions de ségrégation regroupées sur une portion d'un groupe de liaison, comme c'est ici le cas pour le LG7 (figure 6), sont souvent reliées à l'existence de loci de distorsion de ségrégation (LDS) associés à des facteurs de sélection gamétique ou zygotique. Ces facteurs seraient responsables de la majorité des distorsions de ségrégation. Le double test de χ^2 proposé par Lorieux *et al.* (1995) nous a permis de constater que les distorsions observées sur le LG7 seraient dues à la sélection gamétique, pouvant aboutir à des phénomènes de compétition pollinique, de compétition au moment de la fécondation ou du pollen non viable (Song *et al.*, 2006). Ces informations sont en adéquation avec les résultats présentés dans la thèse d'Aline Clabaut (2009). On remarque aussi une distorsion de même type sur le LG7 de Rubis118 (Cadalen *et al.*, 2010). Pour le marqueur 3262R1L1 (LG6), les résultats du double test de χ^2 indiquent que des facteurs de sélection post-fécondation ou zygotiques sont à l'origine des distorsions observées. Cela concorde avec les résultats trouvés pour le LG6 de la carte de référence K28xK59. Pour le marqueur C4H (LG1), ce sont également les facteurs de sélection zygotiques qui semblent impliqués. A l'extrémité du LG1

de la carte de Rubis 118, on constate également un marqueur en distorsion associé à des facteurs de sélection zygotique (donnée non publiée, Thierry Cadalen).

Les mécanismes ou raisons expliquant la distorsion de ségrégation restent néanmoins encore nombreux et difficilement identifiables. Il existe des facteurs environnementaux ou des facteurs biologiques déclenchant ce phénomène comme la complémentarité de gène, l'existence de gènes dupliqués, d'anomalies chromosomiques (Song *et al.*, 2006) ou encore le phénomène d'hétérosis (Brummer *et al.*, 1993). Cependant, une des premières causes serait l'erreur humaine, lors de la lecture du génotypage (Sibov *et al.*, 2003). C'est en effet peut-être le cas pour certains de nos marqueurs, sachant que la population exploitée a été génotypée par plusieurs personnes différentes, pouvant engendrer des différences d'interprétation concernant la lecture du génotypage, plus particulièrement pour des profils ambigus. Toutefois, cela ne devrait pas concerner C4H (LG1, génotypé uniquement pour la carte « oxychic ») et les marqueurs déjà distordus sur la carte de référence K28xK59 et la carte « embryogénèse somatique ». Les marqueurs éventuellement concernés par l'erreur de lecture lors du génotypage seraient donc FDCB131 et EU06A09b du LG2.

La production de plusieurs sets de descendants pour la carte « oxychic » génère des phénomènes localisés de distorsions dans la ségrégation des marqueurs dus à des facteurs de sélection gamétique ou zygotique. Ces sélections pourraient être dues à des conditions environnementales différentes entre les deux croisements (faits en 2001 pour F1'a et en 2004 pour F1'b). Pourtant, ces deux croisements ont été faits en serre, c'est à dire dans un environnement contrôlé et à la même période (juillet/août). On remarque néanmoins une différence d'ensoleillement selon les relevés météorologique : 236h en 2001 et 163h en 2004 (thèse Aline Clabaut, université de Lille1, 2009).

V.4. Gènes candidats

Pour 30 gènes candidats référencés comme étant impliqués dans la biosynthèse des hydroxycinnamates d'intérêt, seulement trois n'ont pas trouvé de séquence correspondante dans la banque EST de *Cichorium*. Parmi les 27 gènes candidats restant, huit n'ont pas présenté de polymorphisme, ni en SSCP ni en HRM (tableau 3). Notons qu'avant d'arriver à cette conclusion, différents couples d'amorces ont été testés quand cela était envisageable (plusieurs introns dans la séquence). Ainsi, pour les gènes candidats monomorphes pour le croisement K28xK59 (ADT, CCoAOMT, MYB4, PAL, SAM, PAP1, PAP2, MLS), il serait

intéressant de les tester sur les autres croisements utilisés pour la carte consensus (Rubis 118 et BR) (Cadalen *et al.*, 2010). La détection de polymorphisme sur une des deux autres descendance indiquerait leur localisation. En effet, en fonction des marqueurs communs, aussi appelés marqueurs « ancrés », qui ont permis de relier les trois cartes génétiques, il serait alors possible de connaître la région où se situe le gène candidat chez K28xK59.

Parmi les 19 gènes candidats polymorphes, 15 ont été cartographiés sur la carte « oxychic ». Les quatre autres, bien qu'ils montrent du polymorphisme, n'ont pas pu être localisés avec assez de précision pour être inclus dans la carte. Cette situation est due à l'ambiguïté des profils de génotypage, générant des erreurs de lecture. Tester d'autres couples d'amorces pour ces gènes candidats afin de rechercher à nouveau de polymorphisme pourrait résoudre ce problème. Aux 15 gènes candidats cartographiés, nous pouvons ajouter le marqueur OC0603, déjà présent sur la carte « embryogenèse somatique » (thèse d'Aline Clabaut, 2009) et qui correspond à une enzyme impliquée dans la voie de la méthionine (HMT), impliquée dans la biosynthèse de substrats pour la voie des phénylpropanoïdes.

Pour compléter nos travaux, la cartographie de gènes candidats impliqués dans la dégradation des molécules polyphénoliques devra être envisagée. Par exemple, notre recherche pourra s'orienter vers les polyphénols oxydases (Nüsslein *et al.*, 2000; Martinez & Duvnjak, 2006). Les gènes candidats ont été cartographiés grâce au polymorphisme détecté à partir d'amorces définies sur des EST de chicorée. Comme il a souvent été observé chez d'autres espèces, il n'est pas exclu que ces gènes fassent partie d'une famille multigénique comme cela a déjà été observé par exemple chez *Arabidopsis* (cf. tableau 3), la tomate (Guo & Wang, 2009), le tabac (Lee & Douglas, 1996) ou le maïs (Szalma *et al.*, 2002).

VI. Conclusion

Nous avons complété le génotypage pour les individus dont les données phénotypiques ont été obtenues dans le chapitre I. La carte K28xK59 « oxychic » est ainsi composée de 142 marqueurs moléculaires, établis grâce aux techniques d'AFLP, SSR et SSCP. Les marqueurs dominants (41%) et co-dominants (59%) se répartissent sur les neuf LG à un intervalle moyen entre deux marqueurs de 6,08 cM ; 7,7% des marqueurs présentent des effectifs en distorsion. Parmi les 142 marqueurs, 16 sont des gènes candidats impliqués

CHAPITRE II

dans les voies de biosynthèse formant les molécules identifiées lors du phénotypage. Ces informations génotypiques, à relier aux caractères quantitatifs, vont permettre de détecter les loci responsables de la variation de l'activité antiradicalaire, de la biosynthèse de l'acide caftarique, chlorogénique et chicorique détectés dans les feuilles de chicorée.

CHAPITRE III

Détection de QTL impliqués dans
l'activité antiradicalaire et le
métabolisme des hydroxycinnamates

Dans le chapitre I, nous avons rendu compte de la mise au point d'un protocole d'extraction de molécules polyphénoliques présentes dans la feuille de *Cichorium intybus* L. Ces extractions ont été suivies d'une analyse CLHP et d'un test d'activité antiradicalaire sur une descendance de type F1' de 201 individus (x 5 blocs). Ces analyses nous ont permis de mettre en évidence de la variabilité phénotypique pour plusieurs caractères : l'activité antiradicalaire (AAR) mesurée par le test DPPH, la quantité d'acide caftarique (ACAFT), chlorogénique (ACHLO), chicorique (ACHIC), la somme des ces trois quantités (ATOT), ainsi que la proportion de molécules d'acide caftarique (PCAFT) et chlorogénique (PCHLO) par rapport à l'ensemble des molécules détectées à 320 nm par CLHP.

Dans le cadre d'une analyse QTL, la variabilité phénotypique observée doit être mise en relation avec la variabilité génotypique afin d'identifier des loci responsables du déterminisme génétique des caractères quantitatifs. Les individus non génotypés préalablement (44%) ont été génotypés à leur tour en utilisant des marqueurs présentant du polymorphisme AFLP, SSCP, SSR et HRM. Ces marqueurs, comprenant 16 gènes candidats, ont pu alors être placés sur la carte « oxychic », présentée dans le chapitre II.

Ce nouveau chapitre associe les données phénotypiques et génotypiques dans le but d'identifier des QTL pouvant expliquer en partie les phénotypes ACAFT, ACHLO, ACHIC, ATOT, PCAFT, PCHLO et AAR retrouvés dans la feuille de chicorée.

I. Matériels et méthodes

I.1. Données utilisées pour la détection de QTL

Pour la détection de QTL, le logiciel MapQTL@5.0 (Van Ooijen *et al.*, 2004) a été utilisé. Les informations concernant la détermination des phases de liaison génétique et la position des marqueurs sur chaque groupe de liaison proviennent du logiciel JoinMap@3.0. (Van Ooijen & Voorrips, 2001).

Les valeurs phénotypiques, détaillées dans le chapitre I (paragraphe II.5), correspondent à l'activité antiradicalaire (AAR) mesurée par le test DPPH, la quantité d'acide caftarique (ACAFT), chlorogénique (ACHLO), chicorique (ACHIC), la somme des ces trois quantités (ATOT), ainsi que la proportion de molécules d'acide caftarique (PCAFT) et chlorogénique (PCHLO) par rapport à l'ensemble des molécules détectées à 320 nm par CLHP. Toutes ces données ont préalablement été ajustées en fonction du modèle statistique

prenant en compte l'effet « bloc » (LSmeans du logiciel, Systat 12 version 12.00.08, SPSS Inc., Chicago, USA).

I.2. Tests statistiques pour la détection de QTL

Le premier outil statistique employé pour la recherche de QTL correspond à une recherche marqueur par marqueur : le test de Kruskal-Wallis (Lehmann, 1975). Il s'agit de l'équivalent non paramétrique d'une analyse de variance : une relation statistique entre le génotype de chaque marqueur et la valeur du caractère quantitatif étudié est recherchée pour l'ensemble des individus d'une descendance donnée.

La seconde analyse statistique utilisée est celle de la cartographie d'intervalle (IM, Interval Mapping, Lander & Botstein, 1989). Cette analyse teste la présence d'un QTL entre deux marqueurs, reliés entre eux par un taux de recombinaison r , en calculant un LOD score. En tout point du génome, on calcule ainsi le logarithme décimal du rapport de vraisemblances, c'est-à-dire tester l'hypothèse nulle « absence d'un QTL » contre l'hypothèse « présence d'un QTL ». Ce test permet de sélectionner les marqueurs les plus proches des QTL pour le test suivant utilisant des cofacteurs. L'intervalle d'analyse choisi pour ce test est de 1 cM.

Le troisième test statistique est un test de type Composite Interval Mapping (CIM) (Jansen & Stam, 1994; Zeng, 1994): le Multiple QTL Model (MQM, Van Ooijen *et al.*, 2004). L'analyse MQM s'appuie sur les marqueurs liés à un QTL selon l'approche IM. La sélection de ces marqueurs en tant que co-facteurs permet de rechercher des QTL mineurs non détectés par l'IM. Un jeu initial de co-facteurs est choisi par l'utilisateur en fonction des résultats IM, correspondants aux marqueurs moléculaires situés « sous » le QTL (à partir du seuil_{LG}). La fonction « Automatic Cofactor Selection » du logiciel élimine un par un les cofacteurs et un rapport de vraisemblance (tester l'hypothèse nulle « absence d'un QTL » contre l'hypothèse « présence d'un QTL ») est calculé pour chaque sous-groupe ainsi créé. Il y a donc autant de sous-groupes qu'il y a de co-facteurs choisis. Le rapport de vraisemblance de chaque sous-groupe est comparé au rapport de vraisemblance du modèle avec tous les co-facteurs sélectionnés par l'utilisateur. Le sous-groupe dont le rapport de vraisemblance est le plus proche du modèle avec tous les co-facteurs est gardé et pris comme point de départ pour la création de nouveaux sous-groupes et ainsi de suite. Le processus s'arrête quand le rapport de vraisemblance du dernier sous-groupe incluant un nouveau co-facteur devient

significativement différent (valeur de $p < 0,02$) du sous-groupe de référence ou quand il n'y a plus de co-facteurs à tester. Un test MQM est alors réalisé en fonction du ou des co-facteur(s) suggéré(s) par le logiciel. Si un nouveau QTL apparaît, un nouveau co-facteur correspondant à ce nouveau QTL est testé. Pour cette analyse, nous avons opté pour un intervalle de 1 cM.

Pour fixer le seuil de détection des QTL, mille tests de permutation ont été effectués (Churchill & Doerge, 1994). Les valeurs phénotypiques sont permutées mille fois au hasard. La distribution des statistiques calculée sur les données réellement observées est alors comparée à la distribution des statistiques calculées à partir des données redistribuées au hasard, permettant de définir un seuil de signification pour rejeter les faux QTL. Ces tests sont effectués pour chaque groupe liaison et pour le génome entier, un seuil de détection leur est donc spécifiquement attribué : seuil_{LGn} et seuil_{GW}, respectivement (n = numéro du groupe liaison et GW = genome wide). Ces deux seuils ont été établis en prenant en compte le seuil de signification de 5% : la probabilité de trouver, par hasard, un LOD score au dessus des seuil_{LGn} ou seuil_{GW} est seulement de 5%.

I.3. Identification de QTL

Les valeurs de LOD score, calculées tous les cM dans un intervalle entre deux marqueurs, sont reliées entre elles pour former une courbe de LOD. Le point le plus haut dépassant le seuil de détection a été considéré comme la meilleure position pour les QTL identifiés. Un intervalle de confiance a été défini en retranchant deux unités de LOD à la valeur maximale.

Pour connaître la contribution de chaque QTL dans la variation phénotypique totale observée, le coefficient de détermination R^2 est calculé comme suit:

$R^2 = 100 \times (H0var - \text{variance totale}) / \text{variance totale}$, avec H0var la variance résiduelle sous l'hypothèse nulle (Van Ooijen *et al.*, 2004).

Les QTLs identifiés ont été nommés par l'abréviation donnée au caractère phénotypique et en indice est indiqué le groupe de liaison (LG) sur lequel il a été détecté. Exemple pour un QTL correspondant à l'aire de l'acide chicorique, retrouvé sur le groupe de liaison 8 : ACHIC_{LG8}.

I.4. Calcul des interactions entre marqueurs

Les tests d'interaction ont été effectués avec le programme R (R Development Core Team, 2009). Les 142 marqueurs moléculaires de la carte « oxychic » sont testés deux à deux, pour chaque caractère. Les 10011 interactions ont été testées suivant le modèle linéaire suivant : $P_{ijk} = \mu + M_j + M_k + M_j M_k + \sum_{ijk}$, avec P_{ijk} = valeur phénotypique du descendant i en considérant dans le modèle les marqueurs j et k ; μ = moyenne phénotypique ; M_j = valeur génotypique au marqueur j ; M_k = valeur génotypique au marqueur k ; \sum_{ijk} = résiduelle du modèle. Les interactions entre marqueurs ont été considérées comme significatives à partir de $p < 0,0001$.

Pour chaque interaction significative, un test de Tukey a été effectué (procédure GLMSAS 9.1 2002-2003) pour comparer les moyennes phénotypiques obtenues pour chaque combinaison allélique des marqueurs présentant des effets épistatiques selon le test d'interaction : on teste les comparaisons entre paires en choisissant la valeur critique en fonction de l'étendue de la comparaison (hypothèse nulle « pas de différence significative entre les moyennes »). Les combinaisons alléliques ont été considérées comme significativement différentes à partir de $p < 0,05$.

II. Résultats

II.1. Détections de QTL

II.1.A. Test de Kruskal-Wallis (K-W)

Le test de K-W permet de connaître quel marqueur est associé de manière significative à la variation du phénotype étudié. Le tableau 1 montre les associations phénotype-marqueur significatives. En effet, comme ce test est effectué sur tous les marqueurs de façon indépendante, le manuel de MapQTL suggère de ne prendre en compte que les associations présentant un niveau de signification élevé, c'est-à-dire ceux dont la valeur de p est inférieure à 0,005 (à partir de trois astérisques).

CHAPITRE III

Caractère: ACAFT				Caractère: ATOT				Caractère: PCHLO			
LG	marqueurs	position	<i>p</i>	LG	marqueurs	position	<i>p</i>	LG	marqueurs	position	<i>p</i>
1	EU0022	0.000	*	3	NAD	2.698	**	1	AOX	30.920	*
6	ttat177	50.474	*	6	ttat177	50.474	*	2	tacg293	0.000	**
9	tgta286	52.604	*	8	atct331	70.294	**	2	EU06G07	5.447	**
9	<u>OC1580</u>	62.885	***	8	OC1976	71.237	*	2	EU05B10	26.492	**
9	EU02C03	66.244	*	9	gatg452	41.206	*	2	<u>atct219</u>	28.031	***
9	<u>OC0603</u>	66.439	***					2	agga247	34.177	**
9	EU04C09	66.707	*	Caractère: PCAFT				2	ttat227	51.674	**
9	tgct222	67.116	**	LG	marqueurs	position	<i>p</i>	2	<u>EU06A09b</u>	73.587	***
9	<u>EU09B01</u>	69.798	***	2	EU05B10	26.492	*	3	FDCA149	43.128	*
				3	NAD	2.698	*	3	<u>FDCA39</u>	88.416	**
Caractère: ACHLO				5	gatg219	33.001	*	6	B12CL1R1	10.913	*
LG	marqueurs	position	<i>p</i>	5	MDH	70.496	*	6	<u>3262R1L1</u>	18.033	*****
1	FDCA63	21.473	*	6	tacg331	33.176	**	6	<u>tgct75</u>	24.661	****
1	AOX	30.920	*	6	agga234	34.001	*	6	<u>tgta165</u>	30.538	*****
1	tacg151	34.780	*	7	CCR	41.018	*	6	<u>ttat177</u>	50.474	*****
2	tacg293	0.000	**	7	EU10H05	45.837	*	6	<u>FDCA134</u>	59.300	***
2	EU06G07	5.447	*	8	ttat237	26.662	*	6	atct283	61.428	*
2	EU05B10	26.492	**	8	FDCB98b	45.697	*	6	tacg191	65.883	**
2	<u>atct219</u>	28.031	*****	9	acag118	21.804	*	8	<u>OC0812</u>	72.393	***
2	agga247	34.177	*	9	<u>sw2A03</u>	43.777	****	8	ExpProt	91.632	*
6	B12CL1R1	10.913	*	9	tgta286	52.604	*	9	cgat206	0.000	*
6	<u>tgct75</u>	24.661	*	9	OC1580	62.885	*	9	sw1D04.1	46.729	*
9	gatg452	41.206	*	9	<u>EU02C03</u>	66.244	***	9	<u>OC1580</u>	62.885	*****
9	<u>OC1580</u>	62.885	****	9	<u>OC0603</u>	66.439	***	9	EU02C03	66.244	*
9	<u>OC0603</u>	66.439	***	9	<u>EU04C09</u>	66.707	****	9	<u>OC0603</u>	66.439	*****
9	<u>tgct222</u>	67.116	***	9	tgct222	67.116	*	9	EU04C09	66.707	*
9	<u>EU09B01</u>	69.798	****	9	EU09B01	69.798	*	9	<u>tgct222</u>	67.116	*****
9	FDCB129	80.049	*	9	<u>FDCB129</u>	80.049	***	9	<u>EU09B01</u>	69.798	*****
								9	<u>FDCB129</u>	80.049	***
Caractère: ACHIC				Caractère: AAR							
LG	marqueurs	position	<i>p</i>	LG	marqueurs	position	<i>p</i>				
3	<u>NAD</u>	2.698	***	1	CAD	19.157	*				
6	ttat177	50.474	*	2	EU05B10	26.492	*				
8	atct197	2.234	*	2	<u>atct219</u>	28.031	***				
8	<u>atct331</u>	70.294	***	2	<u>agga247</u>	34.177	****				
8	OC1976	71.237	**	2	ttat227	51.674	*				
9	gatg452	41.206	*	4	atct475	29.353	*				
				6	<u>ttat177</u>	50.474	*****				
				6	<u>FDCA134</u>	59.300	*****				
				6	<u>tacg191</u>	65.883	***				
				8	EU07B09	46.083	*				

Tableau 1. Résultat de test de K-W sur chaque caractère ayant montré de la variabilité (Chapitre I). Les astérisques indiquent le niveau de probabilité significative :

* = $p < 0,05$; ** = $p < 0,01$; *** = $p < 0,005$; **** = $p < 0,001$; ***** = $p < 0,0005$; ***** = $p < 0,0001$.

Si l'on considère uniquement les associations avec $p < 0,005$, les tests de K-W suggèrent l'existence de QTL liés à la quantité d'acide caftarique (ACAFT) sur le LG9 (OC1580, OC0603, EU09B01). Le marqueur atct219 du LG2 et les marqueurs OC1580, OC0603, tgct222, et EU09B01 du LG9 seraient liés de façon très significative au caractère ACHLO. De manière analogue, le marqueur NAD du LG3 et le marqueur atct331 du LG8 sont significativement liés au contenu d'acide chicorique (ACHIC). Par contre, aucun marqueur n'est statistiquement lié à la concentration des trois molécules étudiées (ATOT) pour une valeur de p inférieure à 0,005. La variation de l'activité antiradicalaire quant à elle, semble affiliée aux marqueurs atct219 et agga247 du LG2 et aux marqueurs ttat177, FDCA134 et tacg191 du LG6. Le caractère PCAFT serait significativement relié au LG9 par les marqueurs sw2A03, EU02C03, OC0603, EU04C09, FDCB129. La variation de la proportion en acide chlorogénique semble associée à plusieurs groupes de liaison : le LG2 (EU06A09b), le LG6 (3262R1L1, tgct75, tgta165, ttat177, FDCA134), le LG8 (OC0812) et le LG9 (OC1580, OC0603, tgct222, EU09B01, FDCB129).

II.1.B. Test de permutation

Mille tests de permutations ($\alpha = 0,05$) ont permis de définir un LOD seuil au-delà duquel les QTL sont significatifs. Deux seuils ont été pris en compte : le seuil calculé sur le génome entier (Seuil_{GW}), et le seuil calculé pour chaque groupe de liaison (Seuil_{LGn}). Leurs valeurs respectives sont mentionnées dans le tableau 2.

Seuil _{GW}	Seuil _{LG1}	Seuil _{LG2}	Seuil _{LG3}	Seuil _{LG4}	Seuil _{LG5}	Seuil _{LG6}	Seuil _{LG7}	Seuil _{LG8}	Seuil _{LG9}
3,9	2,8	2,6	2,7	2,7	2,95	2,8	2,6	3	2,8

Tableau 2. Résultat des tests de permutation (1000 permutations, $\alpha = 0,05$). La valeur de chaque seuil est donnée en unité de LODscore pour le génome entier (GW) et pour chaque groupe de liaison (LGn).

Plus il y a de marqueurs impliqués par groupe de liaison, plus la chance de trouver un QTL par hasard est élevée (erreur de type I), ce qui explique pourquoi le seuil de signification augmente avec le nombre de marqueurs. Les courbes de LOD obtenues avec les analyses IM et MQM sont présentées avec les deux seuils.

II.1.C. Interval-Mapping (IM)

La figure 1 permet de visualiser les résultats de l'analyse d'IM sur chaque groupe de liaison, pour tous les caractères quantitatifs à variabilité significative (ACAFT, ACHLO, ACHIC, ATOT, PCAFT, PCHLO et AAR). Cette analyse teste la présence d'un QTL entre deux marqueurs (intervalle d'analyse = 1 cM). Les deux seuils de signification sont représentés (Seuil_{GW} et $\text{Seuil}_{\text{LGN}}$).

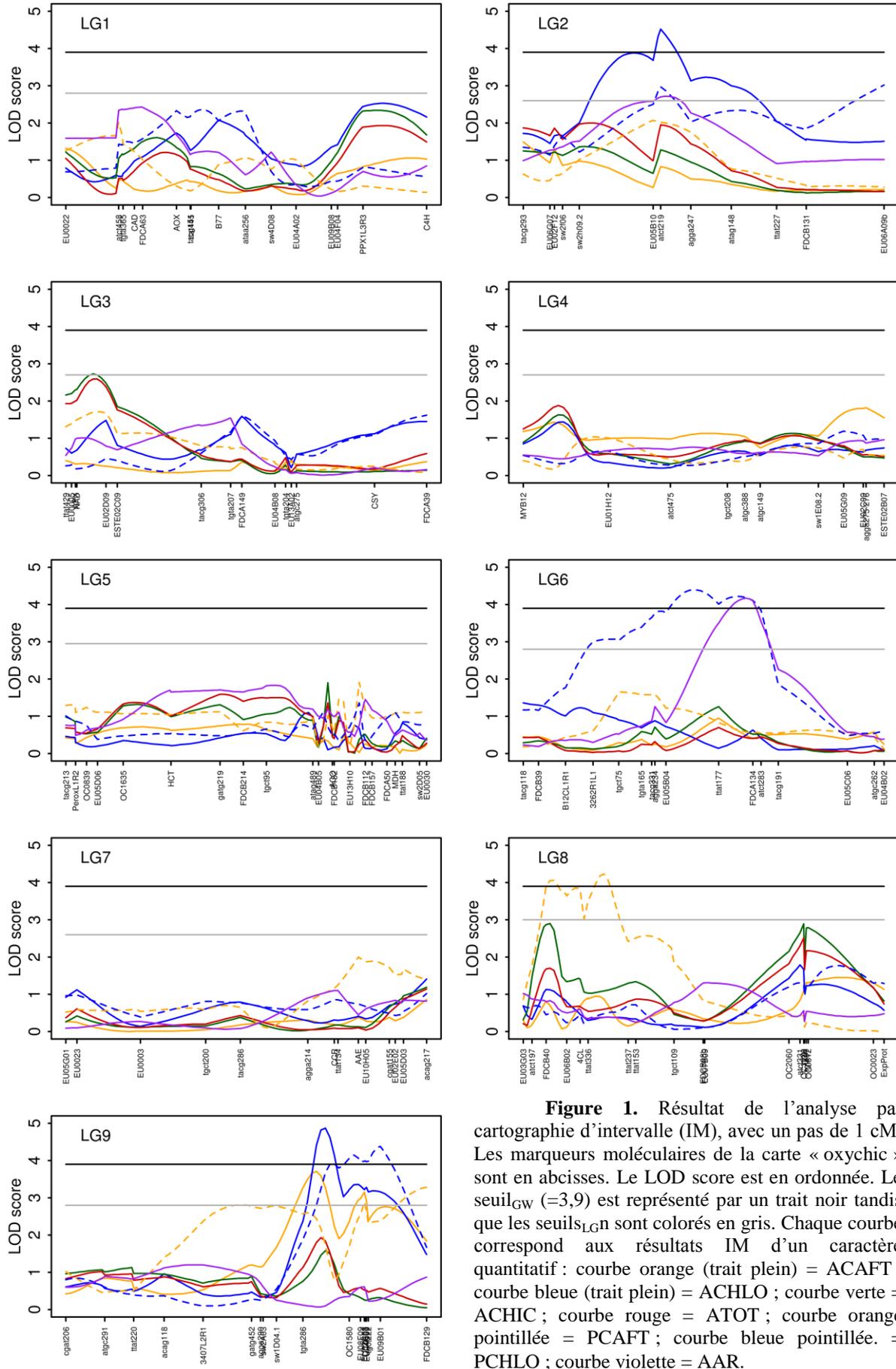


Figure 1. Résultat de l'analyse par cartographie d'intervalle (IM), avec un pas de 1 cM. Les marqueurs moléculaires de la carte « oxychic » sont en abscisses. Le LOD score est en ordonnée. Le seuil_{GW} (=3,9) est représenté par un trait noir tandis que les seuils_{LGn} sont colorés en gris. Chaque courbe correspond aux résultats IM d'un caractère quantitatif : courbe orange (trait plein) = ACAFT ; courbe bleue (trait plein) = ACHLO ; courbe verte = ACHIC ; courbe rouge = ATOT ; courbe orange pointillée = PCAFT ; courbe bleue pointillée. = PCHLO ; courbe violette = AAR.

Certains groupes de liaison ne semblent pas, dans un premier temps, associés aux caractères quantitatifs d'intérêt selon le seuil de signification associé à chaque groupe de liaison : LG1, LG4 et LG5 et LG7. Ces premières informations sont en adéquation avec les résultats donnés par le test de K-W.

Dans le tableau 3 ci-dessous la position et le pourcentage de variation phénotypique expliquée sont détaillés pour les 16 QTL associés aux différents caractères mesurés. Parmi ces 16 QTL, 7 ont un LOD score supérieur au seuil_{GW} et 9 ont un LOD score compris entre le seuil_{GW} et le seuil_{LGn}.

Caractères	LG	Position (cM)	LOD score	R ²
ACAFT	LG9	55,6	3,71	16,4
		66,2	3,15	7
ACHLO	LG2	28	4,52	10,4
	LG9	57,6	4,87	27,9
ACHIC	LG3	6,7	2,73	6,9
ATOT	x	x	x	x
PCAFT	LG8	7,8	4,06	12,7
		20,5	4,23	19,4
	LG9	38,4	2,81	8
		66,2	2,89	6,5
PCHLO	LG2	80	3,28	9,3
		28	2,97	6,7
	73,6	3,02	9,7	
	LG6	43,9	4,39	11,3
	LG9	69,8	4,38	9,7
AAR	LG2	29	2,72	6,5
	LG6	57,5	4,17	9,9

Tableau 3. Résultats de l'analyse IM. Les QTL identifiés pour chaque caractère et indiqué en noir ont un LOD score supérieur au seuil_{GW}. Les QTL dont le LOD score est inférieur au seuil_{GW} mais supérieur au seuil_{LGn} sont grisés. R² : pourcentage de la variation phénotypique expliquée. x = absence de détection d'un QTL.

Pour les QTL détectés, ceux qui partagent les mêmes régions chromosomiques correspondent aux caractères : (1) ACHLO, PCHLO et AAR sur le LG2 aux alentours de 28-29 cM ; (2) PCHLO et AAR sur le LG6, entre 44,9 et 57,5 cM; (3) ACAFT, ACHLO, PCAFT et PCHLO sur le LG9 entre 55,6 et 69,8 cM. Les pourcentages de variation phénotypique expliquée (R²) restent assez faibles avec une moyenne de 11,14%. Le minimum correspond à 6,5 % pour les QTL associés à PCAFT (LG9) et à AAR (LG2) et le maximum correspond à 27,9% pour le QTL associé au phénotype ACHLO sur le LG9, ce qui pourrait correspondre à l'expression d'un gène majeur pour ce caractère.

L'analyse IM n'a révélé aucun QTL pour le caractère ATOT. On peut constater aussi que les courbes de LOD correspondant à ce caractère ont la même allure que celles associées au caractère ACHIC. Cela n'est pas surprenant dans la mesure où l'acide chicorique est la molécule détectée la plus abondante parmi les trois composés additionnés pour constituer le caractère ATOT. Comme vu dans le chapitre I, la variation d'ACHIC explique 90% de la variation d'ATOT (coefficient de corrélation $r = 0,95$, $p < 0,001$). La variabilité trouvée pour le caractère ATOT dépend donc quasi-totalement d'ACHIC et ne reflète pas une variation propre à l'addition des trois molécules d'intérêt. Pour ces deux raisons (absence de QTL et dépendance vis-à-vis du caractère ACHIC), l'analyse QTL n'a pas pu être approfondie pour ce caractère.

Ce premier jeu de QTL et les marqueurs voisins correspondants vont nous permettre de choisir les co-facteurs pour l'analyse MQM.

II.1.D. Multiple QTL Model (MQM)

La stratégie employée est la suivante: un jeu initial de co-facteurs est choisi, en fonction des résultats IM, correspondants aux marqueurs moléculaires situés « sous » le QTL. La fonction « Automatic Cofactor Selection » du logiciel élimine un par un les cofacteurs comme expliqué dans le paragraphe I. Matériels et Méthodes, et un ou deux co-facteurs sont proposés à l'utilisateur.

La courbe de LOD est alors à nouveau recalculée en fonction du ou des co-facteur(s) proposé(s), grâce à la fonction « MQM mapping ». En fonction des courbes obtenues, nous essayons d'autres combinaisons de co-facteurs et effectuons d'autres analyses MQM jusqu'à obtenir les courbes de LOD les plus précises au niveau de l'emplacement des QTL, nous permettant de rétrécir l'intervalle de confiance autour du QTL.

Les graphes présentés dans les paragraphes suivant montrent :

- la courbe IM, en pointillés : courbe de LOD donnée par l'IM (en pointillés) ;
- la courbe MQM_{autom} , avec des tirets : la courbe de LOD résultante de l'analyse MQM avec le co-facteur suggéré par l'« Automatic Cofactor Selection »;
- la courbe MQM_{user} : la courbe de LOD rapportée par l'analyse MQM et une combinaison de co-facteurs la plus adéquate quant à la précision de l'emplacement du QTL (trait plein) ;
- les deux seuils de signification des QTL sont indiqués (seuil_{GW} en noir et seuil_{LGn} en gris) ;

- en abscisses, les marqueurs moléculaires mis en évidence par le test de K-W sont précédés d'un astérisque.

II.1.D.a. QTL liés au caractère ACAFT

Pour le caractère ACAFT et les deux QTL détectés par l'analyse IM sur le LG9, le jeu de co-facteurs initial est composé des marqueurs moléculaires sw1D04.1, tgta286, OC1580, EU08F09, EU02C03, OC0603, EU04C09, tgct222 et EU09B01. Le co-facteur suggéré par le logiciel - tgta286 - nous a permis de tracer une courbe de LOD (courbe MQM_{autom}) affinée autour des marqueurs tgta286 et sw1D04.1 mais le QTL détecté par l'IM à la position 66,2 cM a disparu. La combinaison des co-facteurs tgta286 et OC1580 a précisé d'avantage la détection du QTL à la position 55,6 cM (courbe MQM_{user}), comme l'illustre la figure 2. Ce QTL est supérieur au seuil_{LG9} mais inférieur au seuil_{GW}. La position, le LOD score et le R² du QTL est identique à ce qui avait été trouvé lors de l'analyse IM (tableau 4). Le gène candidat *HMT* (= OC0603) est localisé sur le LG9. Le test de K-W avait révélé qu'il était significativement lié au caractère ACAFT. Cependant, *HMT* ne se situe pas dans l'intervalle de confiance pour le QTL ACAFT_{LG9} selon l'analyse MQM.

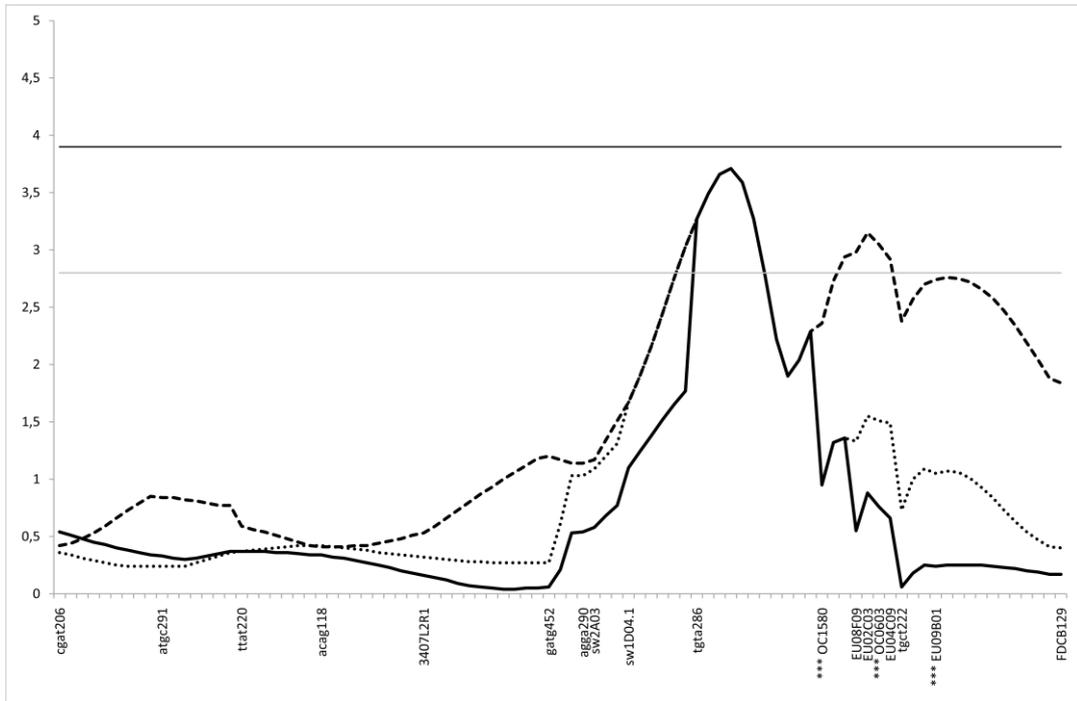


Figure 2. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère ACAFT, sur le LG9. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir (Seuil_{GW} = 3,9), le seuil calculé pour le LG9 est coloré en gris (2,8).

	ACAFT	IM	MQM_{autom}	MQM_{user}	Position (cM)
$ACAFT_{\text{LG9}}$	LOD score maximal	3,71	3,71	3,71	55,6
	R ² (%)	16,4	16,4	16,4	

Tableau 4. LOD score maximal et R² correspondant obtenus suite aux analyses IM, MQM_{autom} et MQM_{user} concernant le caractère ACAFT, pour le LG9. Les LOD scores inférieurs au seuil_{GW} (3,9) sont grisés.

II.1.D.b. QTL liés au caractère ACHLO

L'analyse IM a permis la détection de deux QTL pour le caractère ACHLO sur les LG2 et LG9. Le jeu de co-facteurs initial choisi est donc le suivant : EU05B10, atct219, agga247, atag 148 (LG2), tgta286, OC1580, EU08F09, EU02C03. Le logiciel nous a proposé d'utiliser atct219 et EU08F09 comme co-facteurs. Ceci a permis de tracer une courbe de LOD (courbe MQM_{autom}) affinée autour du marqueur atct219 pour le LG2. Le LOD score du QTL placé sur le LG9 est descendu en dessous du seuil_{GW}. Sur le LG1, un nouveau QTL au LOD score supérieur au seuil_{LG1} a été détecté au niveau du marqueur PPX1L3R3. C'est finalement la combinaison des co-facteurs C4H (LG1), atct219 (LG2), OC1580 (LG9) qui a permis

d'obtenir les courbes (courbe MQM_{user}) illustrées dans les figures 3 à 5. Ces trois QTL ont tous un LOD score supérieur ou égal au seuil_{GW}. Le QTL $ACHLO_{LG1}$ co-localise avec le gène candidat *C4H* pour un intervalle de confiance égale au LOD score maximal retranché d'une unité de LOD score. Pour le LG9, comme pour *ACAFT*, le gène candidat *HMT* (= OC0603) avait été révélé significativement lié au caractère *ACHLO* par le test de K-W. Cependant, une fois l'analyse MQM effectuée, *HMT* ne se situe pas dans l'intervalle de confiance du QTL $ACHLO_{LG9}$.

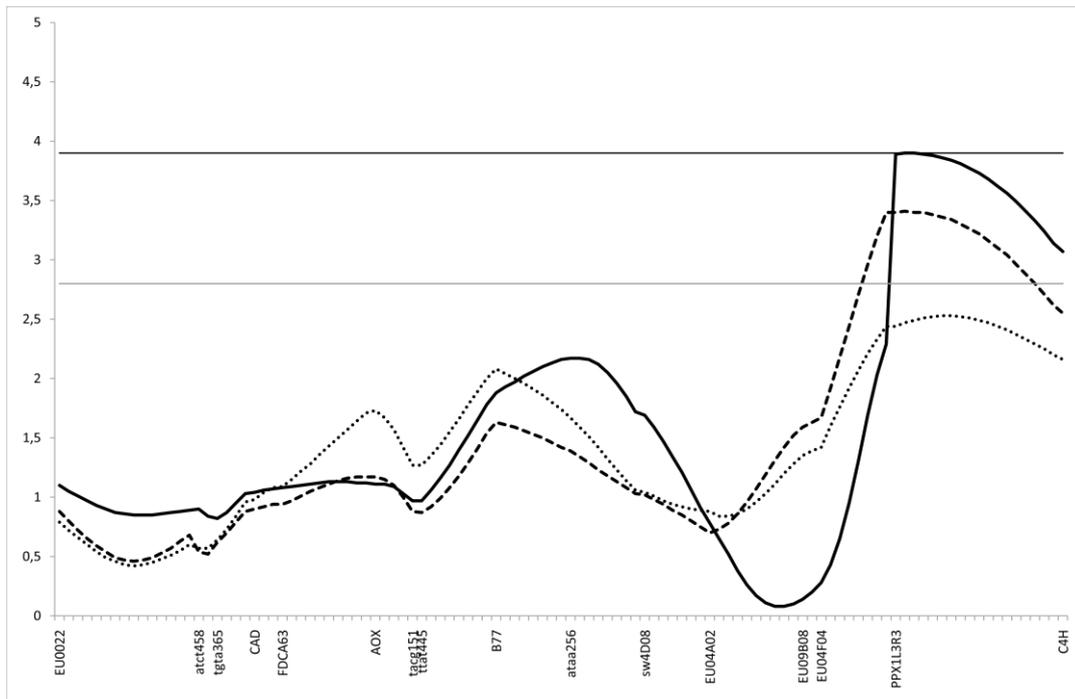


Figure 3. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère *ACHLO*, sur le LG1. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir (Seuil_{GW} = 3,9), le seuil calculé pour le LG1 est coloré en gris (2,8).

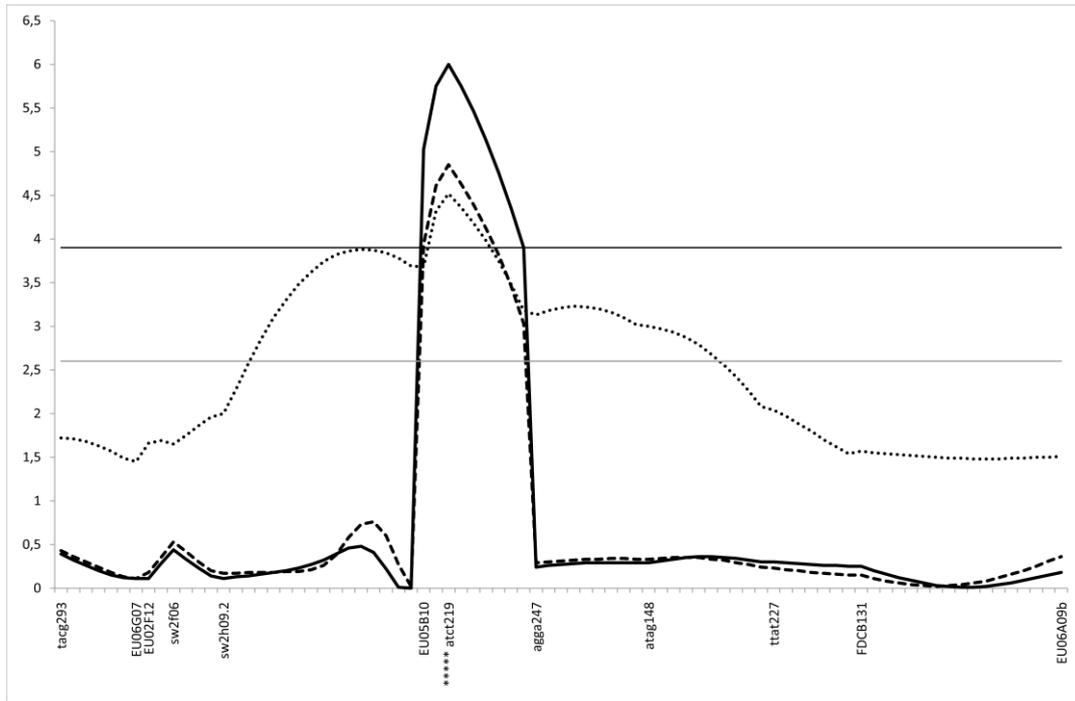


Figure 4. Résultat des analyses IM (pointillés) et MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère ACHLO, sur le LG2. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG2 est coloré en gris (2,6).

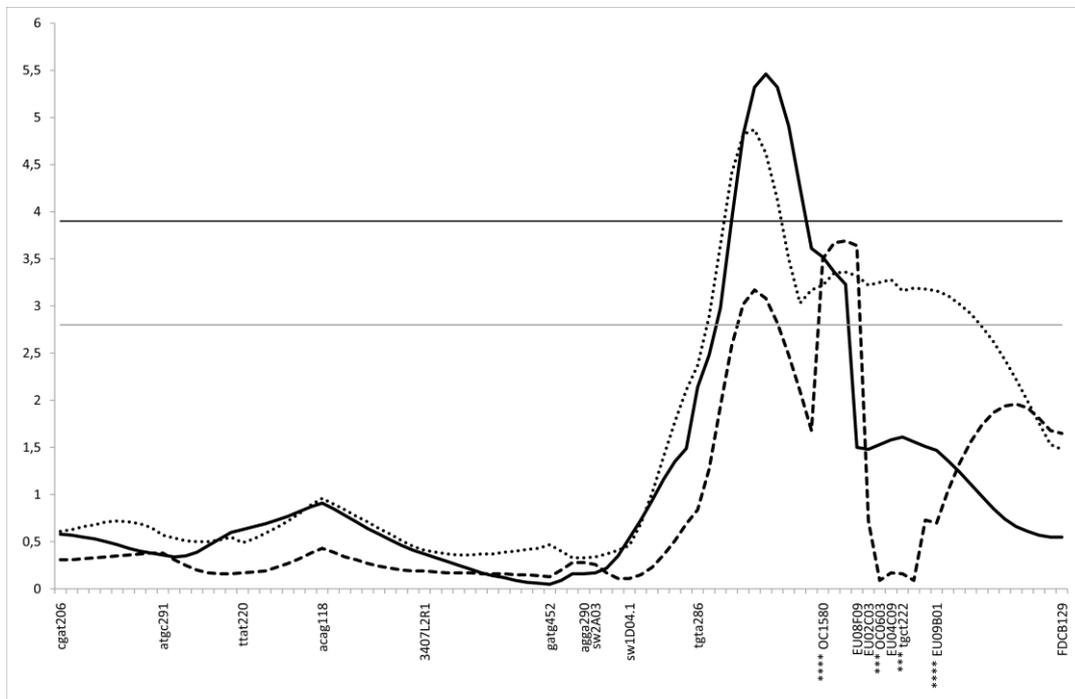


Figure 5. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère ACHLO, sur le LG9. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG9 est coloré en gris (2,8).

Le tableau 5 récapitule les LOD score maximaux obtenus pour les différentes analyses et les pourcentages de variation phénotypique expliquée correspondants (R^2), pour les trois QTL détectés pour le caractère ACHLO selon l'approche MQM_{user}. La somme des R^2 pour ce caractère est égale à 50 %.

	ACHLO	IM	MQM _{autom}	MQM _{user}	Position (cM)
ACHLO _{LG1}	LOD score maximal	-	3,41	3,9	84 (<i>C4H</i>)
	R^2 (%)	-	6,5	8,8	
ACHLO _{LG2}	LOD score maximal	4,52	4,85	6	28
	R^2 (%)	10,4	10,2	14	
ACHLO _{LG9}	LOD score maximal	4,87	3,17	5,46	58,6
	R^2 (%)	27,9	26,1	27,2	

Tableau 5. LOD score maximal et R^2 correspondant obtenus suite aux analyses IM, MQM_{autom} et MQM_{user} concernant le caractère ACHLO, pour les LG1, LG2 et LG9. Les LOD scores inférieurs au seuil_{GW} (3,9) sont grisés. Le symbole « - » correspond aux LOD scores inférieurs au seuil_{GW} et au seuil_{LG} correspondant. Dans la colonne position, figure en italique le nom des gènes candidats qui se situent dans l'intervalle de confiance (LOD score maximal - 1 unité de LOD score).

II.1.D.c. QTL liés au caractère ACHIC

Suite aux résultats des analyses IM, le jeu de co-facteurs initial est constitué des marqueurs moléculaires APX, NAD, EU02D09 sur le LG3. L'utilisation d'EU02D09 en tant que cofacteur, suggéré par le logiciel (MQM_{autom}) a permis de réduire l'intervalle de confiance autour de ce QTL déjà détecté par l'analyse IM. Cela a également permis de révéler un autre QTL, situé sur le LG8, à 71,23 cM. En rajoutant à EU02D09 le marqueur C3H (LG8) comme co-facteur, la courbe de LOD score a également été affinée sur le LG8. Aucun des deux QTL détectés n'atteint le seuil_{GW} (figures 6 et 7). Les QTL ACHIC_{LG3} et ACHIC_{LG8} co-localisent avec le gène candidat *NAD* et *C3H*, respectivement, pour un intervalle de confiance égale au LOD score maximal retranché d'une unité de LOD score.

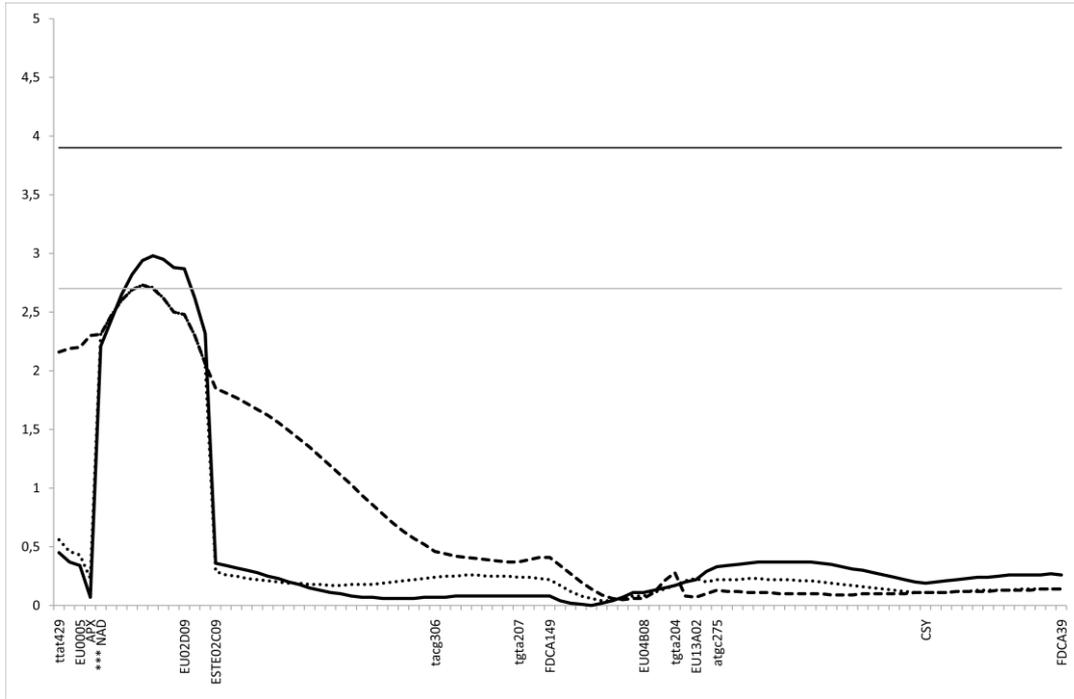


Figure 6. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère ACHIC, sur le LG3. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnées, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG3 est coloré en gris (2,7).

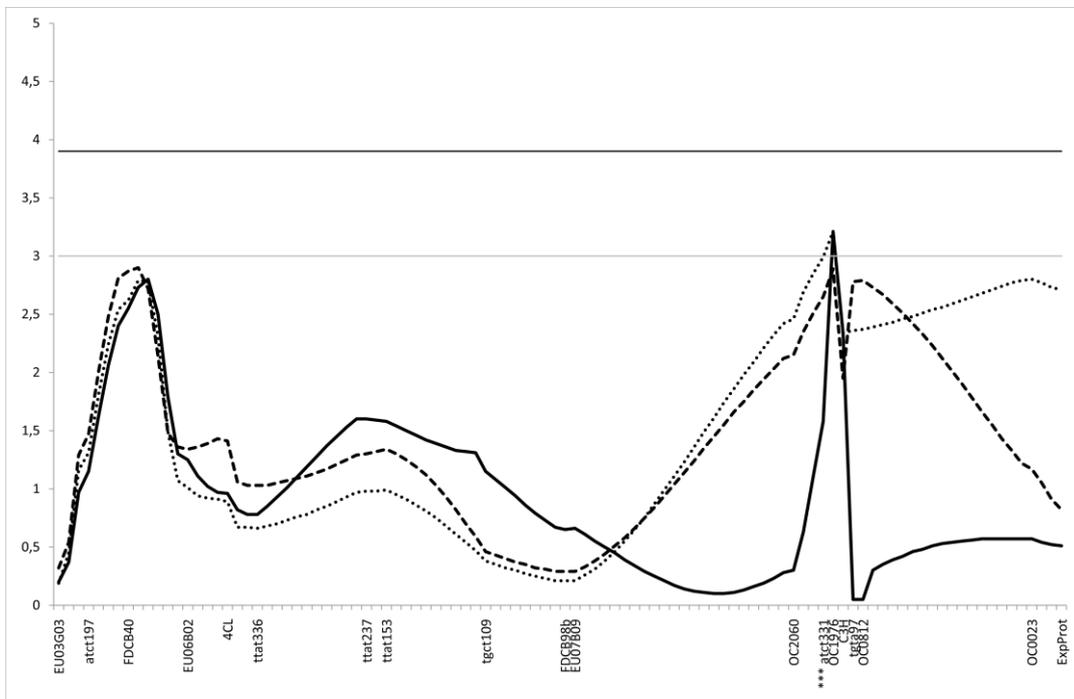


Figure 7. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère ACHIC, sur le LG8. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnées, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG8 est coloré en gris (3).

Le tableau 6 récapitule les LOD maximaux obtenus pour les différentes analyses et les pourcentages de variation phénotypique expliquée correspondants (R^2) pour les deux QTL trouvés selon l'approche MQM_{user} . La somme des R^2 pour le caractère ACHIC est égale à 13,9%.

	ACHIC	IM	MQM_{autom}	MQM_{user}	Position (cM)
ACHIC_{LG3}	LOD score maximal	2,73	2,73	2,98	7,7 (<i>NAD</i>)
	R² (%)	6,9	6,9	7,1	
ACHIC_{LG8}	LOD score maximal	-	3,21	3,21	71,2 (<i>C3H</i>)
	R² (%)	-	6,8	6,8	

Tableau 6. LOD score maximal et R^2 correspondant obtenus suite aux analyses IM, MQM_{autom} et MQM_{user} concernant le caractère ACHIC, pour les quatre QTL détectés. Les LOD scores inférieurs au seuil_{GW} (3,9) sont grisés. Le symbole « - » correspond aux LOD scores inférieurs au seuil_{GW} et au seuil_{LG} correspondant. Dans la colonne position, figure en italique le nom des gènes candidats qui se situent dans l'intervalle de confiance (LOD score maximal - 1 unité de LOD score).

II.1.D.d. QTL liés au caractère PCAFT

Pour le caractère PCAFT, cinq QTL ont été détectés par l'analyse IM : les deux QTL du LG8 sont très proches l'un de l'autre et peuvent laisser supposer qu'il ne s'agit en réalité que d'un QTL ; pareillement, les trois QTL du LG9 pourraient correspondre en fait à deux QTL et non trois). Le jeu de co-facteurs initial est composé des marqueurs moléculaires FDCB40, EU06B02, 4CL, ttat336 (LG8) et gatg452, sw1D04.1, EU02C03, FDCB129 (LG9). Les marqueurs ttat336 et OC0603 correspondent aux co-facteurs suggérés par le logiciel. Leur sélection a permis d'affiner l'intervalle de confiance du QTL autour de la position 20,5 cM du LG8. Il reste néanmoins deux pics très proches. Pour les LG9, la courbe MQM_{autom} s'est précisé au niveau du QTL placé à 66,2 cM, à la défaveur de deux autres QTL détectés aux positions 38,4 cM et 80 cM. Ce choix de co-facteurs a également permis de détecter un nouveau QTL sur le LG2 à la position de 26,5 cM. Ensuite, la combinaison des marqueurs EU05B10 (LG2), 4CL (LG8) et OC0603 (LG9) comme co-facteurs a mis en valeur un nouveau QTL sur le LG1 et un autre sur le LG7. A partir de ces nouvelles informations, nous avons opté pour la combinaison de co-facteurs suivante : atct458 (LG1), EU05B10 (LG2), AAE (LG7), 4CL (LG8) et OC0603 (LG9). Cela a permis de réduire les intervalles de confiance de chaque QTL détecté. Le QTL du LG8 reste encore double et le LOD score maximal correspond à la position 14,05 (au lieu de 20,5 cM suite aux analyses IM) ; bien que

le LOD score maximal soit augmenté, le R^2 baisse à 15% au lieu des 19,4% trouvés en IM. Les résultats des cinq QTL détectés pour le caractère PCAFT sont exposés dans le tableau 7 et dans les figures 8 à 12. Les QTL situés sur les LG2, LG8 et LG9 dépassent le seuil_{GW} alors que les QTL sur les LG1 et LG7 sont uniquement supérieurs au seuil_{LG} qui leur est associé. Les QTL PCAFT_{LG1}, PCAFT_{LG7}, PCAFT_{LG8} et PCAFT_{LG9} co-localisent respectivement avec les gènes candidats *CAD*, *AAE*, *4CL* et *HMT* pour un intervalle de confiance égale au LOD score maximal retranché d'une unité de LOD score.

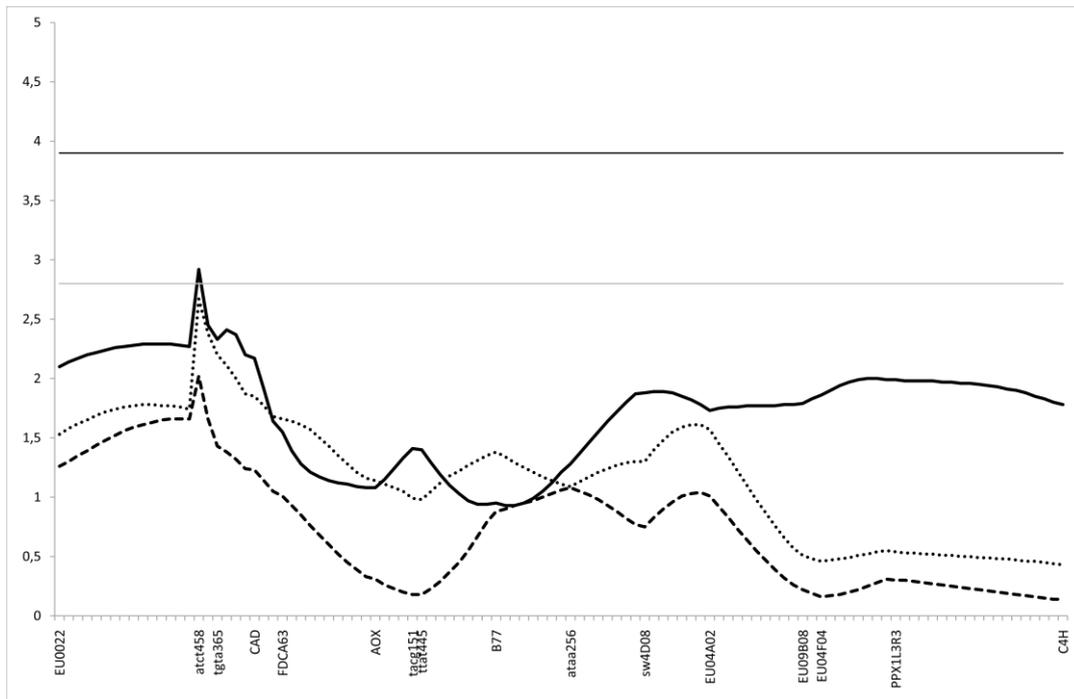


Figure 8. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère PCAFT, sur le LG1. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir (Seuil_{GW} = 3,9), le seuil calculé pour le LG1 est coloré en gris (2,8).

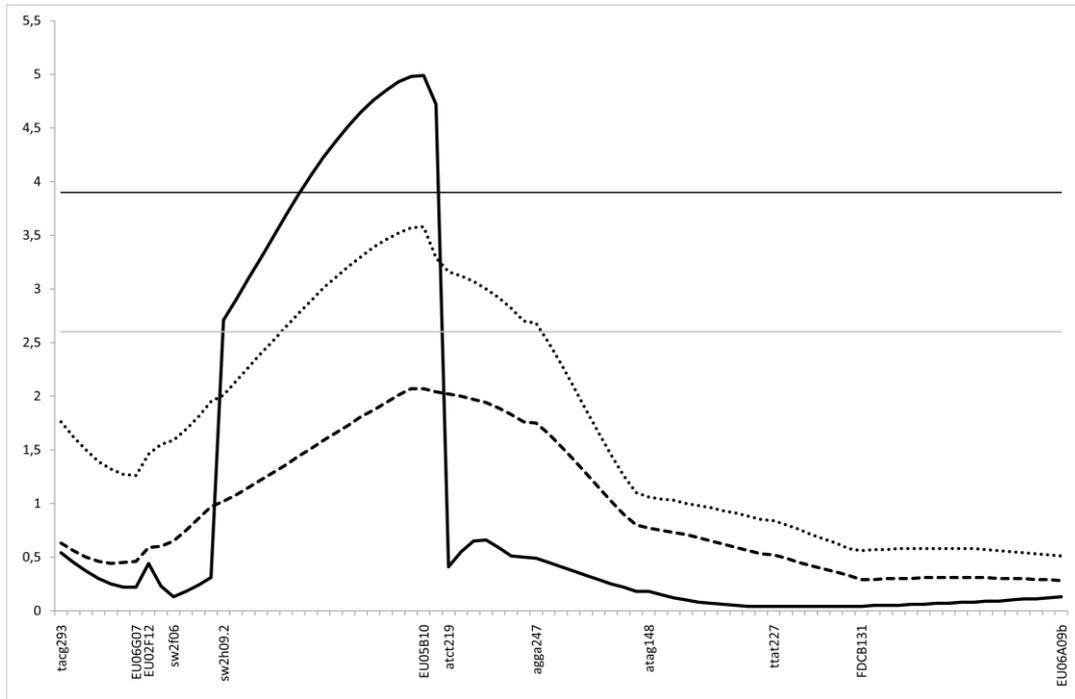


Figure 9. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère PCAFT, sur le LG2. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG2 est coloré en gris (2,6).

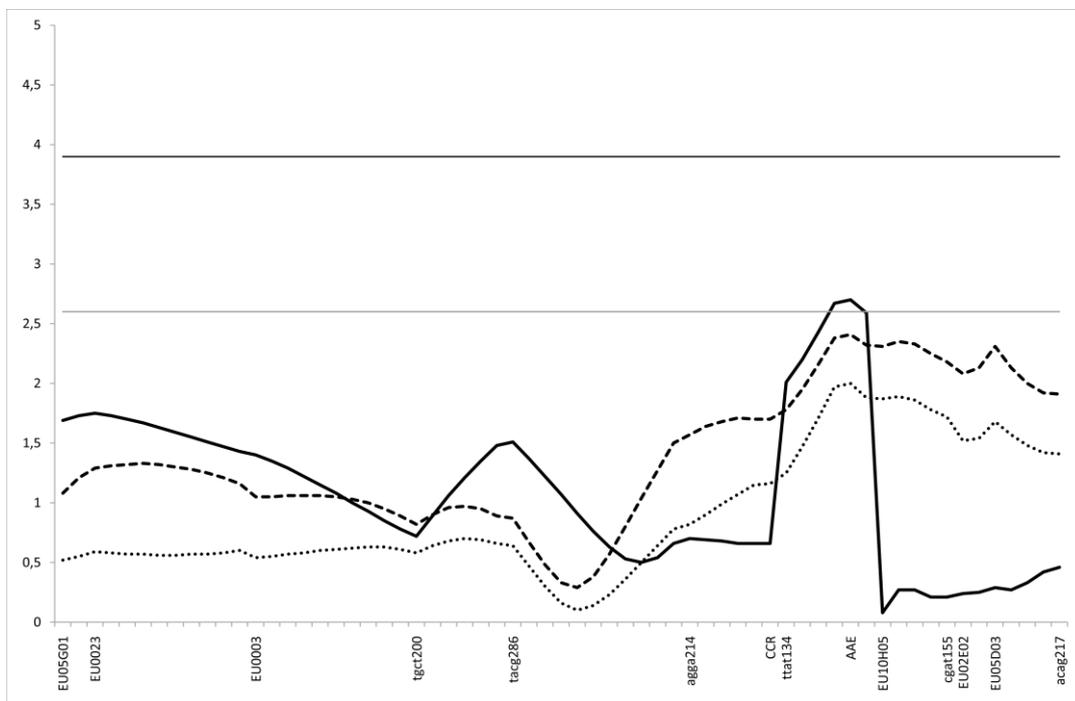


Figure 10. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère PCAFT, sur le LG7. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG7 est coloré en gris (2,7).

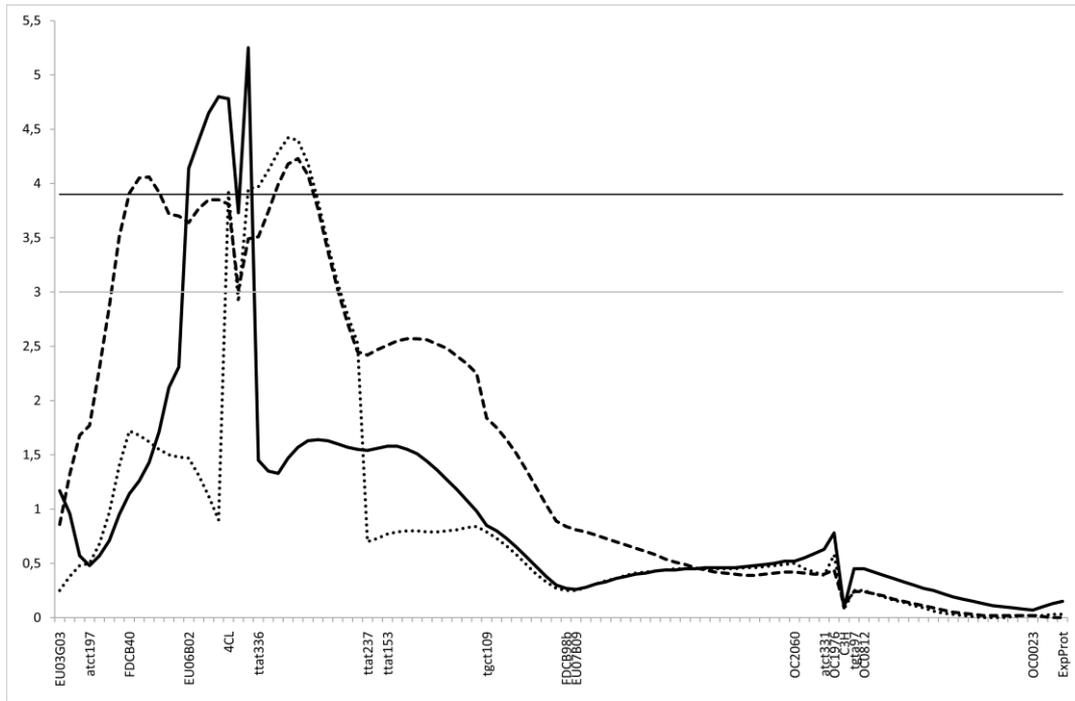


Figure 11. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère PCAFT, sur le LG8. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG8 est coloré en gris (3).

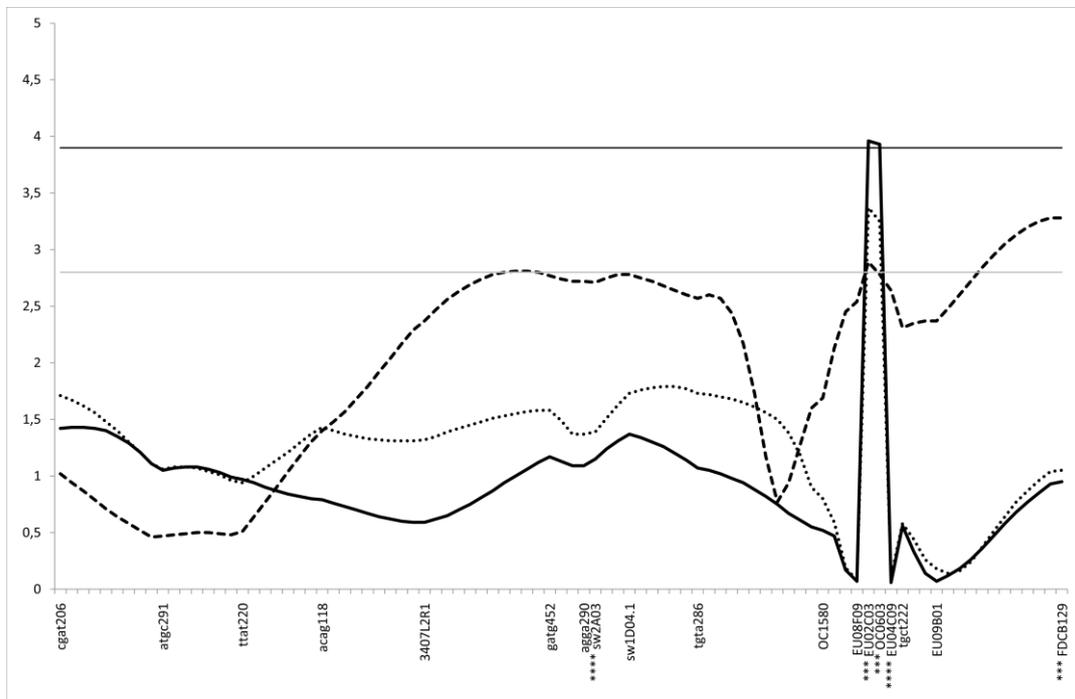


Figure 12. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère PCAFT, sur le LG9. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG9 est coloré en gris (2,8).

Le tableau 7 récapitule les LOD maximum obtenus pour les différentes courbes et les pourcentages de variation phénotypique expliquée correspondants (R^2), pour les cinq QTL associés au caractère PCAFT selon l'approche MQM_{user} . La somme des R^2 pour ce caractère est égale à 44,4 %.

	PCAFT	IM	MQM_{autom}	MQM_{user}	Position (cM)
PCAFT _{LG1}	LOD score maximal	-	-	2,92	14,8 (<i>CAD</i>)
	R^2 (%)	-	-	8,2	
PCAFT _{LG2}	LOD score maximal	-	3,58	4,99	26,5
	R^2 (%)	-	7,8	10,4	
PCAFT _{LG7}	LOD score maximal	-	-	2,7	44,7 (<i>AAE</i>)
	R^2 (%)	-	-	3,8	
PCAFT _{LG8}	LOD score maximal	4,23	4,4	5,25	14 (<i>4CL</i>)
	R^2 (%)	19,4	18,9	15	
PCAFT _{LG9}	LOD score maximal	2,89	3,37	3,96	66,2 (<i>HMT</i>)
	R^2 (%)	6,5	7,9	7	

Tableau 7. LOD score maximal et R^2 correspondant obtenus suite aux analyses IM, MQM_{autom} et MQM_{user} concernant le caractère PCAFT, pour les LG1, LG2, LG7, LG8 et LG9. Les LOD scores inférieurs au seuil_{GW} (3,9) sont grisés. Le symbole « - » correspond aux LOD scores inférieurs au seuil_{GW} et au seuil_{LG} correspondant. Dans la colonne position, figure en italique le nom des gènes candidats qui se situent dans l'intervalle de confiance (LOD score maximal - 1 unité de LOD score).

II.1.D.e. QTL liés au caractère PCHLO

Suite aux analyses IM, le jeu de co-facteurs initial a été constitué comme suit : atct219, agga247 (LG2), EU05B04, ttat177, FDCA134 (LG6), OC1580, OC0603, tgct222, EU09B01 (LG9). Parmi ces marqueurs moléculaires, le logiciel a suggéré d'utiliser atct219, EU05B04, tgct222, EU09B01. Ce set de co-facteurs a permis de réduire l'intervalle de confiance des QTL des LG2, LG6 (mais avec un LOD score descendu en dessous du seuil_{GW}) et LG9 (pic double). En plus, deux nouveaux QTL sur le LG4 ont été révélés: le premier à la position de 11 cM et le second à 67,8 cM. Ensuite, la sélection de deux co-facteurs supplémentaires situés au niveau des QTL du LG4 précédemment détectés (EU01H12 et EU05G09), un QTL a été détecté sur le LG8, à 83,4 cM. C'est finalement le choix des marqueurs en tant que co-facteurs EU05B10 (LG2), EU05G09 (LG4), EU05B04 (LG6), OC0812 (LG8) et EU04C09 (LG9) qui nous a permis d'obtenir les résultats illustrés par la

courbe MQM_{user} (figures 13 à 17) et détaillés dans le tableau 8. Sur le LG9, comme pour ACAFT et ACHLO, le gène candidat *HMT* (= OC0603) avait été révélé très significativement lié au caractère PCHLO par le test de K-W. Pourtant, une fois l'analyse MQM effectuée, *HMT* ne se situe pas dans l'intervalle de confiance du QTL $ACHLO_{LG9}$.

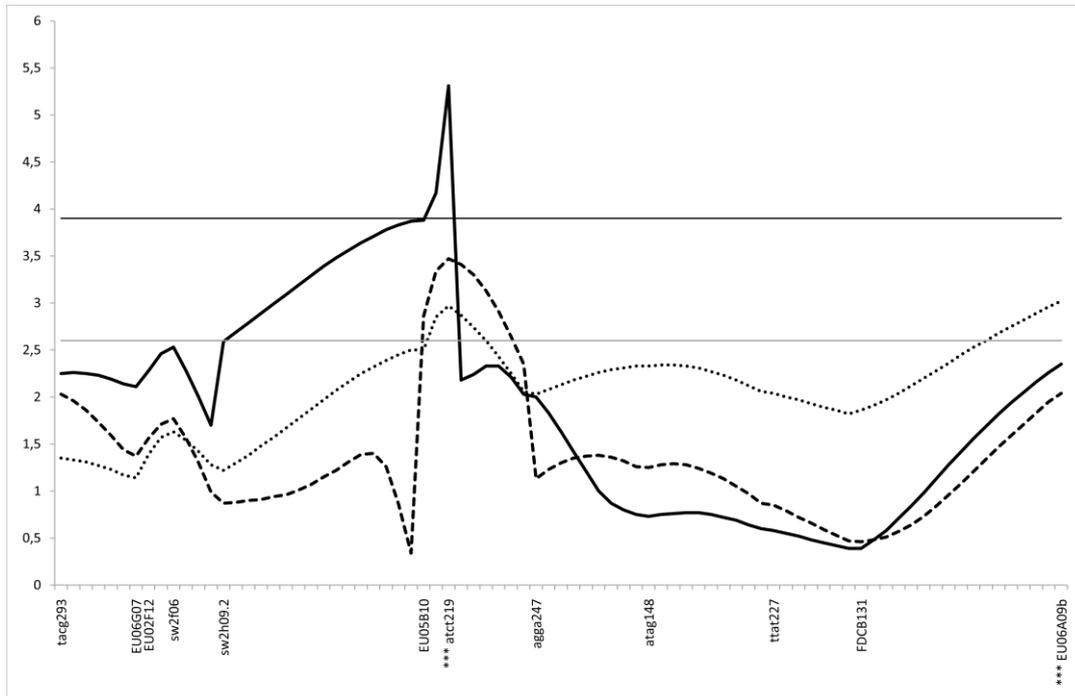


Figure 13. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère PCHLO, sur le LG2. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($Seuil_{GW} = 3,9$), le seuil calculé pour le LG2 est coloré en gris (2,6).

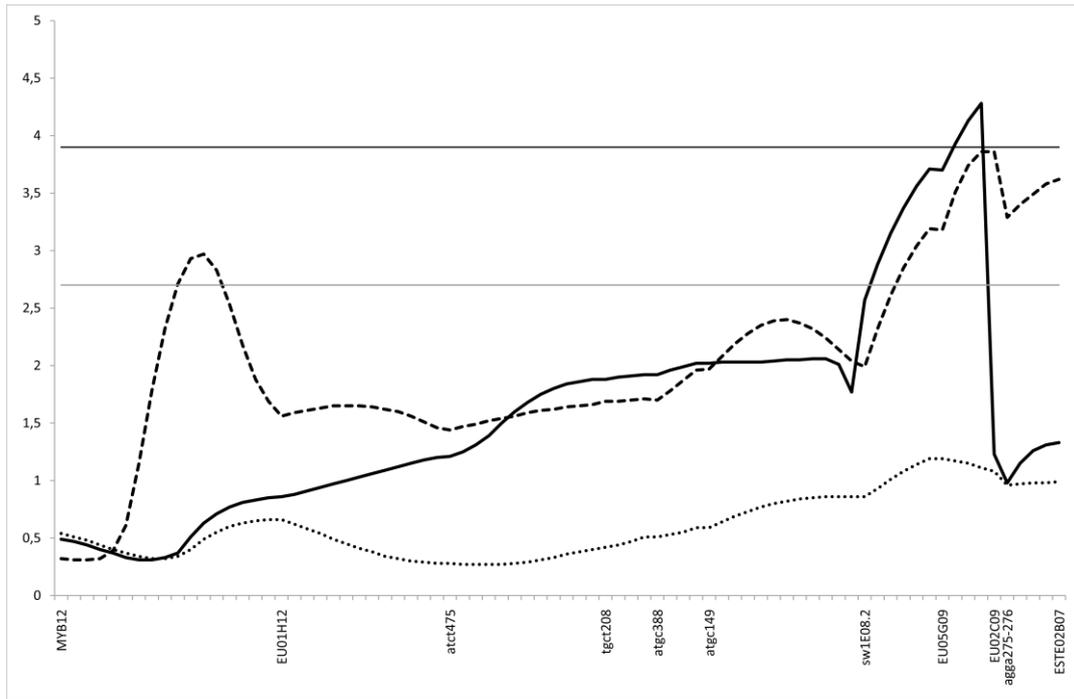


Figure 14. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère PCHLO, sur le LG4. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG4 est coloré en gris (2,7).

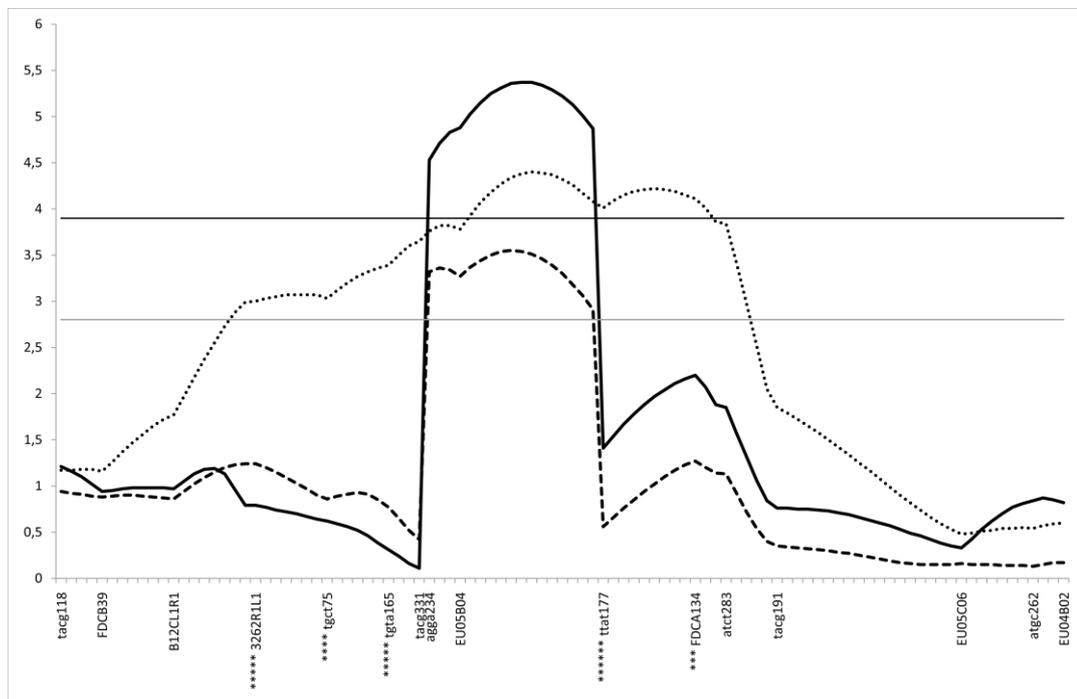


Figure 15. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère PCHLO, sur le LG6. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG6 est coloré en gris (2,8).

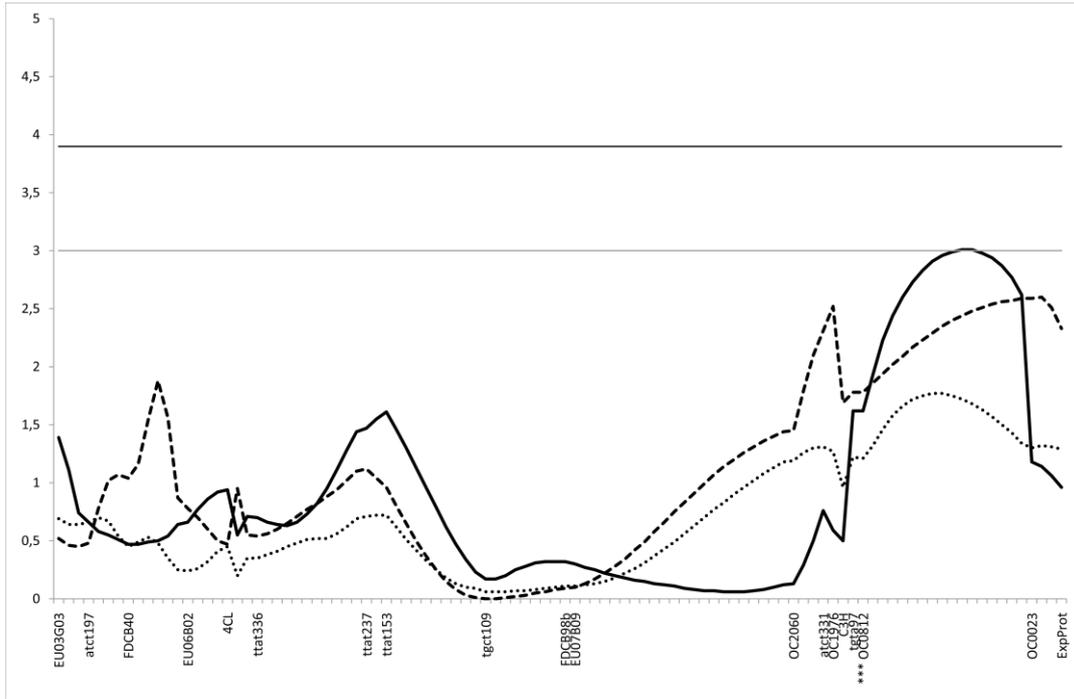


Figure 16. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère PCHLO, sur le LG8. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir (Seuil_{GW} = 3,9), le seuil calculé pour le LG8 est coloré en gris (3).

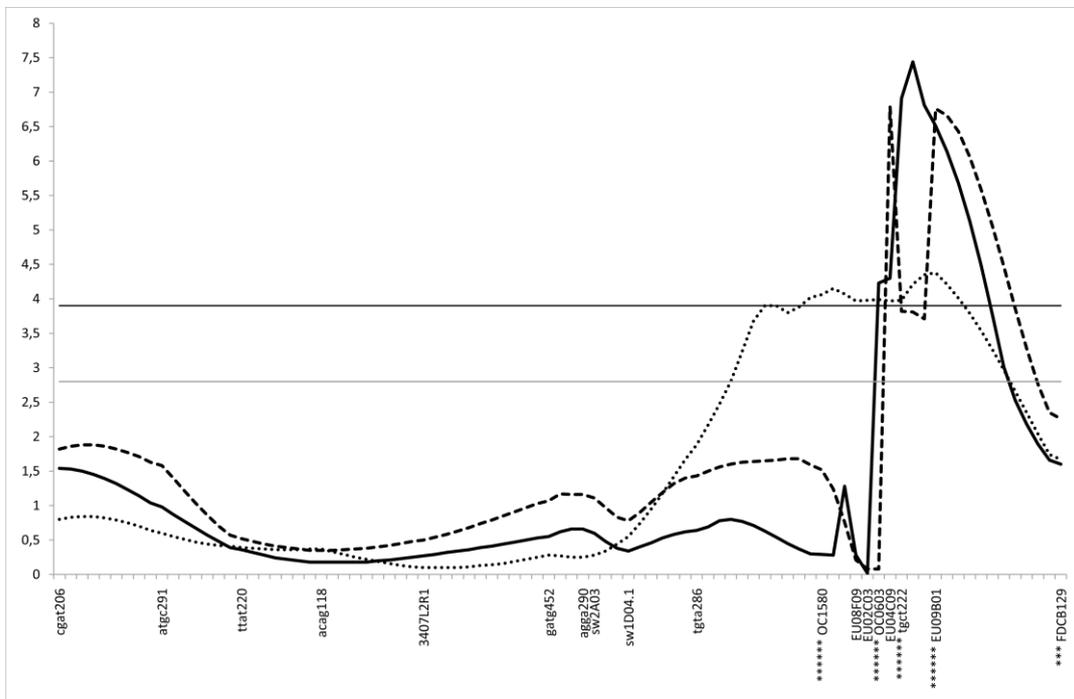


Figure 17. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère PCHLO, sur le LG9. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir (Seuil_{GW} = 3,9), le seuil calculé pour le LG9 est coloré en gris (2,8).

Concernant les cinq QTL détectés pour le caractère PCHLO selon l'approche MQM_{user}, le tableau 8 récapitule les LOD scores maximaux obtenus pour les différentes courbes et les pourcentages de variation phénotypique expliquée correspondants (R²). La somme des R² pour ce caractère est égale à 60,7 %.

	PCHLO	IM	MQM _{autom}	MQM _{user}	Position (cM)
PCHLO _{LG2}	LOD score maximal	2,97	3,47	5,3	28
	R ² (%)	6,7	3,7	13,9	
PCHLO _{LG4}	LOD score maximal	-	3,86	4,28	66,9
	R ² (%)	-	7,2	8,3	
PCHLO _{LG6}	LOD score maximal	4,39	3,55	5,37	43,9
	R ² (%)	43,9	5,6	11,1	
PCHLO _{LG8}	LOD score maximal	-	-	3,01	82,4
	R ² (%)	-	-	7	
PCHLO _{LG9}	LOD score maximal	4,38	6,76	7,44	68,1
	R ² (%)	69,8	20,6	20,4	

Tableau 8. LOD score maximal et R² correspondant obtenus suite aux analyses IM, MQM_{autom} et MQM_{user} concernant le caractère PCHLO, pour les LG2, LG4, LG6 et LG9. Les LOD scores inférieurs au seuil_{GW} (3,9) sont grisés. Le symbole « - » correspond aux LOD scores inférieurs au seuil_{GW} et au seuil_{LG} correspondant.

II.1.D.f. QTL liés au caractère AAR

Les marqueurs moléculaires EU05B10, atct219, agga247 (LG2), ttat177, FDCA134 et atct283 et tacg191 (LG6) composent le jeu de co-facteurs initial pour les analyses MQM. L'emploi des co-facteurs proposés par le logiciel - atct219 et FDCA134 - nous a permis de préciser la position des deux QTL révélés après les analyses IM. L'addition du co-facteur EU05B10 pour le LG2 et de ttat177 pour le LG6 ont encore réduit l'intervalle des QTL détectés. L'ajout du co-facteur CAD pour le LG1 a permis de la révélation d'un autre QTL sur le LG1, à la position de 57,4 cM. L'intervalle de confiance autour de ce marqueur a été rétréci lorsque le marqueur sw4D08 a été ajouté en tant que co-facteur pour l'analyse MQM. Les courbes MQM_{user} exposées dans les figures 18, 19 et 20 ont été enregistrées à partir des données obtenues avec les co-facteurs suivants : CAD, sw4D08 (LG1), EU05B10, atct219 (LG2), FDCA134 et ttat177 (LG6). Le QTL AAR_{LG1} co-localise avec le gène candidat CAD

pour un intervalle de confiance égale au LOD score maximal retransché d'une unité de LOD score.

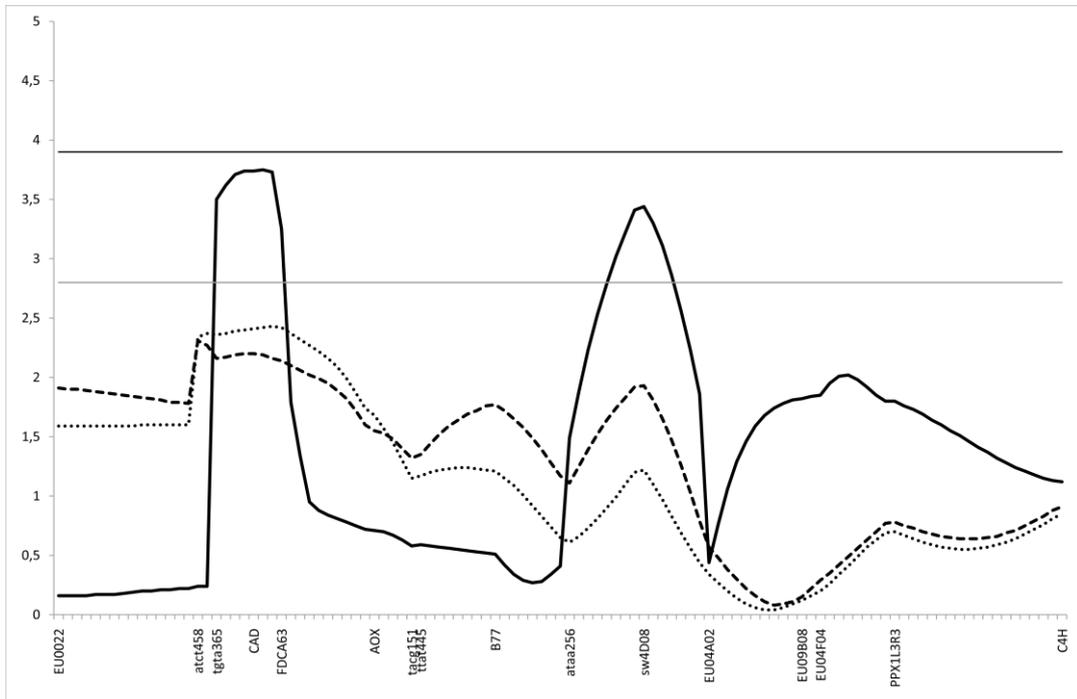


Figure 18. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère AAR, sur le LG1. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG1 est coloré en gris (2,8).

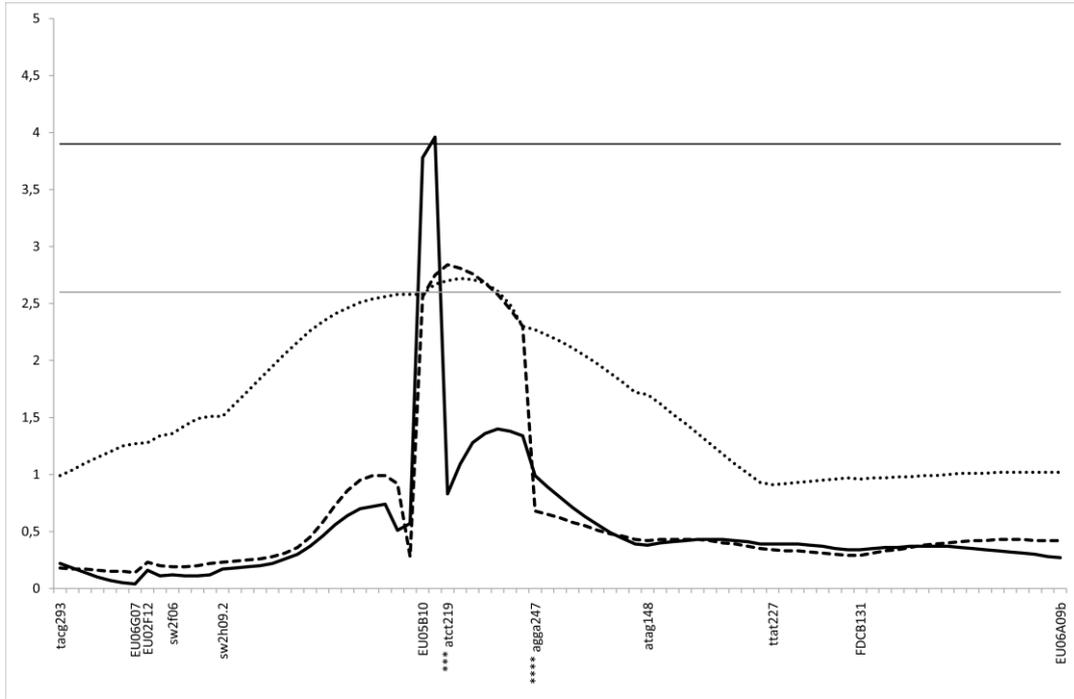


Figure 19. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère AAR, sur le LG2. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG2 est coloré en gris (2,6).

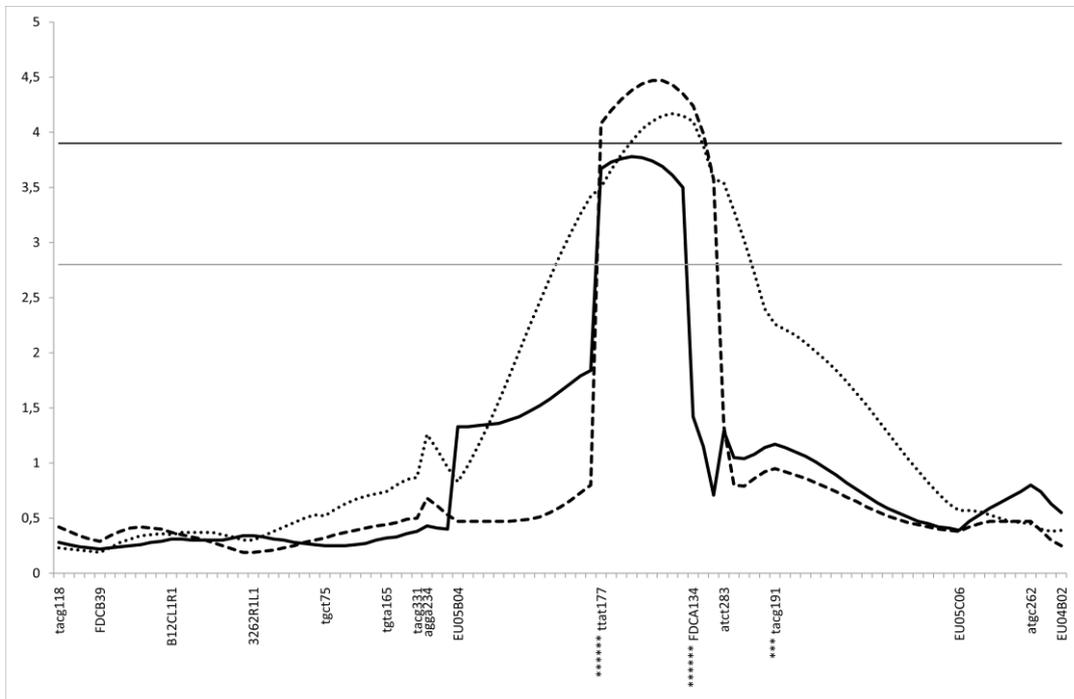


Figure 20. Résultat des analyses IM (pointillés), MQM_{autom} (tirets) et MQM_{user} (trait plein) pour le caractère AAR, sur le LG6. En abscisse, les marqueurs moléculaires de la carte « oxychic » ; les marqueurs précédés d'un astérisque sont significativement liés au phénotype selon le test de K-W. En ordonnée, le LOD score. Le seuil de signification des QTL pour le génome entier est représenté par un trait noir ($\text{Seuil}_{\text{GW}} = 3,9$), le seuil calculé pour le LG2 est coloré en gris (2,8).

Pour les quatre QTL détectés pour le caractère AAR (approche MQM_{user}), le tableau 9 récapitule les LOD scores maximaux obtenus pour les différentes courbes et les pourcentages de variation phénotypique expliquée correspondants (R^2). La somme des R^2 pour ce caractère est égale à 31,1 %.

	AAR	IM	MQM_{autom}	MQM_{user}	Position (cM)
AAR_{LG1-I}	LOD score maximal	-	-	3,75	20,1 (<i>CAD</i>)
	R^2 (%)	-	-	7,6	
AAR_{LG1-II}	LOD score maximal	-	-	3,44	57,4
	R^2 (%)	-	-	9,4	
AAR_{LG2}	LOD score maximal	2,72	2,84	3,96	27,5
	R^2 (%)	6,5	5,9	8,2	
AAR_{LG6}	LOD score maximal	4,17	4,47	3,78	53,5
	R^2 (%)	9,9	10,2	5,9	

Tableau 9. LOD score maximal et R^2 correspondant obtenus suite aux analyses IM, MQM_{autom} et MQM_{user} concernant le caractère AAR, pour les LG1, LG2 et LG6. Les LOD scores inférieurs au seuil_{GW} (3,9) sont grisés. Le symbole « - » correspond aux LOD scores inférieurs au seuil_{GW} et au seuil_{LG} correspondant. Dans la colonne position, figure en italique le nom des gènes candidats qui se situent dans l'intervalle de confiance (LOD score maximal - 1 unité de LOD score).

II.1.E. Effets alléliques

Les valeurs phénotypiques moyennes associées aux classes génotypiques des marqueurs liés aux différents QTLs sont exposés dans le tableau 10. Par ce biais, il a été possible de savoir quel parent porte l'allèle favorable (amenant une augmentation de la valeur phénotypique) ou défavorable (baisse de la valeur phénotypique) aux différents caractères observés.

CHAPITRE III

	nom des QTL					parent responsable de l'effet
LGI		tgta365 16 cM <nn x np>				
		nn	np			
	PCAFT _{LG1}	0,403443	0,375488			K28↘
		CAD 19,2 cM <nn x np>				
		nn	np			
	AAR _{LG1-I}	0,765192	0,745826			K28↘
		sw4D08 57,4 cM <efx eg>				
		ee	eg	ef	fg	
AAR _{LG1-II}	0,382865	0,377987	0,365846	0,373875	K59↘	
	PPX1L3R3 83 cM <nn x np>					
	nn	np				
ACHLO _{LG1}	48707,4	43306,5			K28↘	

	nom des QTL					parent responsable de l'effet
LG2		EU05B10 26,5 cM <ab x cd>				
		ac	ad	bc	bd	
	PCAFT _{LG2}	0,211267	0,218411	0,202926	0,220193	K28↗
		atct219 28 cM <lm x ll>				
		ll	lm			
	ACHLO _{LG2}	53818,9	69981,5			K59↗
	PCHLO _{LG2}	0,1076932	0,13239891			K59↗
AAR _{LG2}	0,782373	0,798835			K59↗	

	nom des QTL					parent responsable de l'effet
LG3		EU02D09 9,8 cM <efx eg>				
		ee	eg	ef	fg	
ACHIC _{LG3}	232311	264223	250600	280648	K28↗	

	nom des QTL					parent responsable de l'effet
LG4		EU02C09 67,8 cM <nn x np>				
		nn	np			
PCHLO _{LG4}	0,247596	0,245444			K28↘	

CHAPITRE III

	nom des QTL					parent responsable de l'effet
LG6		EU05B04 36,9 cM <efx eg>				
		ee	eg	ef	fg	
	PCHLO _{LG6}	0,116072	0,125766	0,131856	0,133672	K59↗
		ttat177 50,5 cM <lm x ll>				
		ll	lm			
	AAR _{LG6}	0,768725	0,745302			K59↘

	nom des QTL					parent responsable de l'effet
LG7		AAE 44,7 cM <nn x np>				
		nn	np			
	PCAFT _{LG7}	0,412302	0,423382			K28↗

	nom des QTL					parent responsable de l'effet	
LG8		4CL 14,4 cM <nn x np>					
		nn	np				
		PCAFT _{LG8}	0,418352	0,418533			K28↗
			C3H 71,4 cM <efx eg>				
			ee	eg	ef	fg	
		ACHIC _{LG8}	232311	263762	223939	250965	K28↗
			OC0023 88,9 cM <lm x ll>				
			ll	lm			
	PCHLO _{LG8}	0,225574	0,242345			K59↗	

	nom des QTL			parent responsable de l'effet
LG9		tgta286 52,6 cM <lm x ll>		
		ll	lm	
	ACAFT _{LG9}	117976,1	129986,8	K59↗
	ACHLO _{LG9}	56057,4	67608,1	K59↗
		OC0603 = HMT 66,4 <nn x np>		
		nn	np	
	PCAFT _{LG9}	0,415451	0,429615	K28↗
		tgct222 67,1 cM <nn x np>		
		nn	np	
	PCHLO _{LG9}	0,11625732	0,05966903	K28↘

Tableau 10. Effet alléliques observés pour les marqueurs liés significativement aux QTL détectés. Pour chaque QTL, les valeurs moyennes phénotypiques ont été rapportées au génotype du marqueur le plus proche pouvant indiquer quel allèle intervient dans la baisse ou la hausse du phénotype. Les données phénotypiques les plus élevées sont en gras. La colonne de droite mentionne quel parent est favorable (↗) ou défavorable (↘) au caractère étudié.

Pour chaque marqueur lié aux différents loci détectés, les allèles des parents sont associés aussi bien à des effets positifs que négatifs pour chaque caractère, expliquant peut-être l'absence de variabilité phénotypique observée entre les parents K59 et K28 pour les caractères ACAFT, ACHLO et AAR (cf. Chapitre I). Les nouvelles combinaisons obtenues en croisant K28 et K59 ont ainsi permis d'observer des phénomènes de transgression dans la descendance (valeurs phénotypiques supérieures et inférieures aux valeurs parentales).

II.1.F. Interactions entre marqueurs

La recherche d'interaction entre les 142 marqueurs a révélé des phénomènes d'épistasie pour chaque caractère sauf ACHLO, comme le présente le tableau 11. Les interactions significatives ($p < 0,0001$) ont été détectées entre des marqueurs autres que ceux significativement liés aux QTL correspondants à chaque caractère (cf. tableau 10) sauf pour le marqueur sw1E08.2 du LG4 associé au QTL PCHLO_{LG4}. Ces marqueurs, aux effets épistatiques, sont eux aussi liés génétiquement aux loci associés aux caractères d'intérêt et peuvent expliquer la part de variation phénotypique non attribués aux QTL détectés précédemment. Aucun gène candidat n'a été recensé dans cette étude d'interaction.

ACAFT	OC0812 ↔ 3407L2R1 LG8 LG9	$F = 20,51$
ACHLO	aucune interaction détectée	x
ACHIC	sw2h09.2 ↔ EU05B10 LG2 LG2	$F = 5,22$
PCAFT	tgta204 ↔ OC0812 LG3 LG8	$F = 19,69$
	PPX1L3R3 ↔ EU03G03 LG1 LG8	$F = 4,77$
PCHLO	FDCB40 ↔ ttat336 LG8 LG8	$F = 14,10$
	ttat429 ↔ EU03G03 LG3 LG8	$F = 10,17$
	sw1E08.2 ↔ EU0003 LG4 LG7	$F = 8,23$
AAR	EU04A02 ↔ agga234 LG1 LG6	$F = 16,38$

Tableau 11. Résultats des tests d'interaction entre les 142 marqueurs de la carte « oxychic » ($p < 0,0001$). $F = F$ de Snedecor.

Les valeurs de F permettent de juger de l'effet de ces interactions sur le phénotype. On retrouve des effets assez importants pour l'interaction entre OC0812 et 3407L2R1 (ACAFT) et tgta204 et OC0812 (PCAFT). Pour ces marqueurs révélateurs d'effets épistatiques, les différentes combinaisons alléliques sont représentées dans la figure 21. Ensuite, le test de Tukey a permis de comparer les moyennes obtenues pour chaque combinaison et ainsi de connaître quelles sont les combinaisons alléliques significativement différentes des autres. Pour illustrer cela, nous prendrons l'exemple simple de la première interaction de la figure 21 (caractère ACAFT, OC0812 - 3407L2R1) qui ne comporte que quatre combinaisons alléliques possibles : ll/nn, lm/nn, ll/np ou lm/np (tableau 12). On constate que l'interaction ll/np est significativement différente de ll/nn et de lm/np ($p < 0,05$) mais qu'elle n'est pas significativement différente de lm/nn. Les résultats des tests de Tukey pour les autres interactions sont disponibles dans l'annexe 4.

ACAFT	ll / nn	ll / np	lm / nn	lm / np
OC0812 / 3407L2R1	52474,0460 ^a	70039,8359 ^b	67090,6347 ^{b,c}	57270,0396 ^{a,c}

Tableau 12. Résultats du test de Tukey pour l'interaction entre le marqueur OC0812 - 3407L2R1 pour le caractère ACAFT. Chaque colonne indique la valeur phénotypique associée à la combinaison entre les allèles du marqueur OC0812 (<lm x ll>) et 3407L2R1 (<nn x np>). Les lettres en exposant indiquent les catégories qui sont significativement différentes des autres ($p < 0,05$).

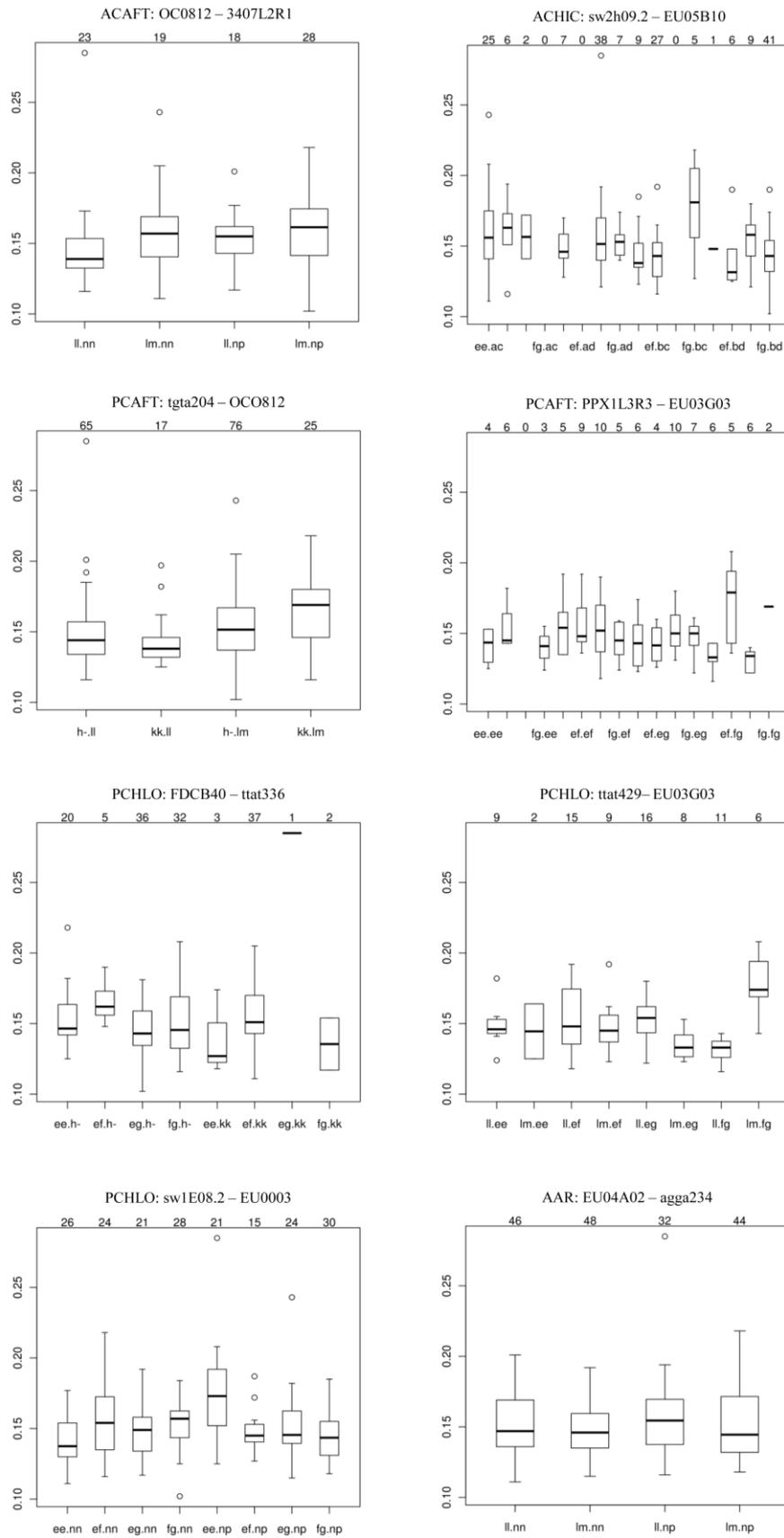


Figure 21. Résultat des interactions. En abscisses, les différentes combinaisons alléliques et les effectifs correspondant. En ordonnées, les valeurs phénotypiques. Les petits cercles représentent les valeurs extrêmes (« outliers »).

III. Discussion

III.1. Méthode de détection et robustesse des QTL

Vingt QTL ont été détectés pour les six caractères étudiés : un pour ACAFT ($R^2 = 16,4\%$), trois pour ACHLO ($R^2 = 50\%$), deux pour ACHIC ($R^2 = 13,9\%$), cinq pour PCAFT ($R^2 = 44,4\%$) et PCHLO ($R^2 = 60,7\%$) et quatre pour AAR ($R^2 = 31,1\%$). Le tableau 13 reprend les informations obtenues grâce aux analyses MQM_{user} (LOD score maximal, R^2 , position sur le LG). Parmi les QTL identifiés, ACHLO_{LG9} et PCHLO_{LG9} peuvent être considérés comme des QTL majeurs (R^2 supérieur à 20%) alors que AAR_{LG6}, PCAFT_{LG7}, ACHIC_{LG8} semblent contrôler une part mineure du caractère qui leur est associé.

QTL	LOD	R ² (%)	Position (cM)	K-W	seuil _{GW}	g. c.	parent responsable de l'effet
PCAFT _{LG1}	2,92	8,2	14,8			CAD	K28↘
AAR _{LG1-I}	3,75	7,6	20,2			CAD	K28↘
AAR _{LG1-II}	3,44	9,7	57,4				K59↘
ACHLO _{LG1}	3,9	8,4	84		=	C4H	K28↘
PCAFT _{LG2}	4,99	10,4	26,5		>		K28↗
AAR _{LG2}	3,96	8,2	27,5		>		K59↗
ACHLO_{LG2}	6	14	28	*	>		K59↗
PCHLO_{LG2}	5,31	13,9	28	*	>		K59↗
ACHIC _{LG3}	2,98	7,1	7,7	*		NAD	K28↗
PCHLO _{LG4}	4,28	8,7	66,9		>		K28↘
PCHLO _{LG6}	5,37	11,1	43,9		>		K59↗
AAR _{LG6}	3,78	5,9	53,5	*			K59↘
PCAFT _{LG7}	2,7	3,8	44,7			AAE	K28↗
PCAFT _{LG8}	5,25	15	16,4		>	4CL	K28↗
ACHIC _{LG8}	3,21	6,8	71,2	*		C3H	K28↗
PCHLO _{LG8}	3,01	7	83,4	*			K59↗
ACAFT _{LG9}	3,71	16,4	55,6				K59↗
ACHLO_{LG9}	5,46	27,2	58,6	*	>		K59↗
PCAFT_{LG9}	3,96	7	66,2	*	>	HMT	K28↗
PCHLO_{LG9}	7,44	20,4	68,1	*	>		K28↘

Tableau 13. Tableau récapitulatif des données obtenues suite aux analyses MQM_{user}, pour les groupes de liaison où des QTL ont été identifiés. R^2 (%) = pourcentage de variabilité phénotypique expliqué. Le symbole « * » est indiqué pour les QTL détecté au niveau d'un marqueur moléculaire significativement liés aux valeurs phénotypiques ($p < 0,005$, test K-W). Le symbole « > » est précisé lorsque le QTL détecté dépasse le seuil_{GW} (3,9). La colonne « g.c » précise si un QTL co-localise avec un gène candidat. Les QTL dont les LOD score sont supérieurs au seuil_{GW} et significativement liés à un marqueur révélé par le test K-W sont en gras. La colonne de droite mentionne quel parent est favorable (↗) ou défavorable (↘) au caractère étudié.

L'analyse MQM permet de détecter des QTL à effet faible pouvant être masqués par des QTL expliquant une large portion de la variance. Ces QTL à effet mineur se retrouvent en effet dans la résiduelle des modèles IM, d'où l'intérêt des modèles incluant des co-facteurs représentant l'effet des QTL majeurs. La façon de procéder pour cette analyse, en choisissant un jeu de co-facteurs initial correspondant aux QTL détectés en IM, puis en rajoutant d'autres co-facteurs si de nouveaux QTL sont détectés, reste néanmoins relativement subjective : avec le même jeu de données, un autre utilisateur pourrait présenter des résultats quelque peu différents.

Les analyses de type CIM étant dépendantes des marqueurs choisis en co-facteur par l'utilisateur et pouvant donner des résultats sujet à controverse, l'équipe de Li (Li *et al.*, 2007) a proposé une alternative baptisée ICIM pour Inclusive Composite Interval Mapping. Ils ont développé un algorithme qui, à chaque pas, procède à une régression linéaire incluant tous les marqueurs simultanément, sauf les deux marqueurs flquant l'intervalle testé. Les valeurs phénotypiques sont ajustées en fonction de cette régression linéaire puis testées par IM. Cette méthode non arbitraire quant au choix des co-facteurs a été développée à partir de QTL Cartographer (Wang *et al.*, 2005b), logiciel gratuit et donc très couramment utilisé mais ne prenant pas en compte les descendance de type F1'.

En ce qui nous concerne, nous avons choisi de présenter les QTL détectés accompagnés des résultats du test K-W et de deux seuils de signification. Ces deux critères permettent de classer les QTL trouvés par l'analyse MQM selon leur robustesse.

Premièrement, opter pour un LOD score correspondant à celui trouvé pour le génome entier (1000 tests de permutation) nous permet d'éviter les erreurs de type I, c'est-à-dire le rejet d'une hypothèse H_0 qui en fait, est vraie (absence d'un QTL entre deux marqueurs). Dans la littérature, on retrouve certains travaux utilisant un seuil de signification par génome entier (Szalma *et al.*, 2002; Davey *et al.*, 2006; Werij *et al.*, 2007) ou par chromosome (Margawati *et al.*, 2009). D'autres encore choisissent un seuil arbitraire (Tozlu *et al.*, 1999a ; Tozlu *et al.*, 1999b; Lespinasse *et al.*, 2000; Li *et al.*, 2000). Ces deux derniers choix semblent cependant imprudents dans la mesure où ils augmentent la possibilité de prendre en compte de faux QTL. Précisons également que d'autres méthodes de calcul sont disponibles pour calculer les seuils de détection, comme les tableaux de simulations développés par Van Ooijen (1999) (Van Ooijen, 1999) mais dont l'utilisation est moins fréquente ou couplée avec les tests de permutation de Churchill et Doerge (1994) (Atienza *et al.*, 2003).

Deuxièmement, l'utilisation du test K-W permet de confirmer que les QTL détectés au niveau des marqueurs significativement liés au phénotype ne correspondent pas à des artéfacts dus à de la distorsion de ségrégation ou à l'absence de distribution gaussienne des valeurs phénotypiques (Costantini *et al.*, 2008), même si dans notre cas, la distribution est normale pour chaque caractère (cf. Chapitre I).

Les QTL dépassant le seuil_{GW} et liés à des marqueurs ayant révélé une forte association ($p < 0,005$) aux valeurs phénotypiques par le test de K-W sont alors considérés comme particulièrement robustes (tableau 13). C'est ainsi le cas des QTL ACHLO_{LG2}, PCHLO_{LG2}, ACHLO_{LG9}, PCAFT_{LG9} et PCHLO_{LG9}. Notons qu'Atienza *et al.* utilisent une méthode plus restrictive que la nôtre pour éviter les faux positifs: les deux seuils de signification ainsi que les résultats de K-W sont pris en compte mais ils ne parlent que de QTL potentiels si le seuil_{GW} n'est pas franchi et ne prennent pas en compte les QTL non liés à des marqueurs mis en exergue par le test de K-W (Atienza *et al.*, 2003). Cette dernière précaution n'a pas été retenue pour notre étude car les marqueurs associés aux QTL détectés lors de l'analyse MQM n'ont pas pu être décelés par le test de K-W, celui-ci ne prenant pas en compte l'effet des co-facteurs. Dans notre cas, les QTL ACHLO_{LG1}, PCAFT_{LG1}, PCAFT_{LG2}, PCAFT_{LG7}, PCHLO_{LG4}, AAR_{LG1-I} et AAR_{LG1-II} se sont révélés après l'étape MQM. Pour les QTL AAR_{LG2} et ACAFT_{LG9}, déjà présents lors de l'IM, même si les marqueurs « sous » le QTL ne sont pas significativement liés au phénotype pour un seuil de probabilité $p < 0,005$, le test de K-W a tout de même révélé un lien pour une probabilité moins forte (tableau 1). De plus, le QTL AAR_{LG2} a un LOD score maximal supérieur au seuil_{GW}. Par contre, les QTL PCAFT_{LG8} et PCHLO_{LG6} ne sont pas associés à des marqueurs significativement liés au phénotype. Cependant, leur détection ne peut pas être complètement rejeté, leur LOD score maximal étant également supérieur au seuil_{GW} (5,25 et 5,37 >> 3,9).

Dans le chapitre II, nous avons évoqué des distorsions de ségrégation, particulièrement importantes sur le LG7, or un QTL a été détecté sur ce groupe de liaison, le PCAFT_{LG7}. Ce dernier co-localise avec le gène candidat *AAE* et se situe dans la zone de distorsion, même si le marqueur *AAE* n'a pas lui-même présenté d'effectifs distordus. Le LOD score maximal (2,7) est en dessous du seuil_{GW} et aucune liaison significative entre le génotypage et le phénotypage n'a été détecté (test de K-W). Il faut donc rester prudent quant à la détection d'un QTL dans cette région chromosomique. Sur d'autres groupes de liaison, certains QTL sont également associés à des marqueurs liés à des loci présentant de la distorsion : *tgta365* et *CAD* associés aux QTL AAR_{LG1} et PCAFT_{LG1}, et *C4H* associé au QTL

ACHLO_{LG1}. Ces marqueurs n'ont pas été associés significativement aux valeurs phénotypiques (test de K-W), cependant, comme précisé dans le paragraphe précédent, ces QTL ont été détectés après l'analyse MQM. Seul ACHLO_{LG1} atteint le seuil_{GW}.

III.2. Co-localisations

III.2.A. Co-localisation des différents QTL des caractères étudiés

Les analyses MQM effectuées révèlent une co-localisation de plusieurs QTL. C'est ainsi le cas des QTL PCAFT_{LG1}, et AAR_{LG1-I} (position entre 14 et 20 cM), des QTL PCAFT_{LG2}, ACHLO_{LG2}, PCHLO_{LG2} et AAR_{LG2} (26-28 cM), des QTL ACAFT_{LG9} et ACHLO_{LG9} (55-58 cM) et enfin PCAFT_{LG9} et PCHLO_{LG9} (66-68 cM). Les QTL associés à l'activité antiradicalaire co-localisent deux fois (sur le LG1 et le LG2) avec des QTL PCAFT et également deux fois avec un QTL PCHLO (LG2 et LG6). On retrouve également un QTL AAR dans la même région chromosomique qu'un QTL correspondant au caractère ACHLO alors qu'aucun QTL correspondant aux caractères ACAFT et ACHIC n'a été détecté dans cette région. Cela est relativement surprenant car, dans le chapitre I, lorsque nous évoquions les différentes corrélations entre les caractères, aucune corrélation significative n'avait été détectée entre AAR et ACHLO alors qu'entre AAR et ACAFT, ainsi qu'entre AAR et ACHIC, des corrélations significatives - bien qu'assez faibles ($r = 0,35$ et $r = 0,41$, respectivement) - avaient été révélées. Pour AAR_{LG2} comme pour ACHLO_{LG2}, c'est l'allèle porté par K59 qui a un effet positif sur les valeurs phénotypiques.

Les caractères ACHLO et ACAFT ont un QTL à la position commune, sur le LG9, suggérant des bases génétiques communes entre ces deux molécules, pour lesquelles une corrélation de $r = 0,55$ avait été trouvée (chapitre I). Cependant, la corrélation entre ACHLO et ACHIC était encore plus importante puisqu'elle atteignait un coefficient de corrélation « r » égale à 0,64. Il a donc été surprenant de ne pas détecter de loci communs pour les caractères ACAFT, ACHIC et ACHLO, la biosynthèse de ces trois composés devant être régulée par les gènes codant pour les enzymes impliquées dans la formation de l'acide caféique, leur molécule commune. Il est encore plus surprenant qu'aucun QTL en commun n'est été révélé pour les acides caftarique et chicorique (ACAFT et ACHIC), ceux-ci étant formés à partir des mêmes molécules (esters d'acide caféique et d'acide tartarique) et pour lesquelles une corrélation de $r = 0,77$ avait été trouvée dans le chapitre I. Deux hypothèses pourraient

expliquer cette absence de bases génétiques. La première suggérerait des voies de biosynthèse secondaires distinctes pour ces deux composés : un gène codant une enzyme capable d'estérifier deux molécules d'acide caféique avec une molécule d'acide tartarique et un autre gène capable d'estérifier une seule molécule d'acide caféique avec une molécule d'acide tartarique, pour l'obtention de l'acide chicorique et de l'acide caftarique, respectivement. L'autre hypothèse suppose que des bases génétiques communes existent pour la formation de ces deux molécules mais qu'aucun QTL commun n'a été détecté car il se situerait dans une région non balayées par les marqueurs moléculaires de la carte « oxychic ». En effet, dans le chapitre précédent, nous avons remarqué des zones peu denses en marqueur: entre PPX1L3R3 et *C4H* (LG1), entre FDCB131 et EU06A09 (LG2), entre ESTE02C09 et tacg306, entre atgc275 et CSY (LG3), entre *MYB12* et EU01H12 (LG4), entre tacg191 et EU05C06 (LG6), ou encore FDCB98b/EU07B09 et OC2060 du LG8.

Sur le LG9, une autre région chromosomique semble contrôler la régulation des molécules d'acide caftarique et chlorogénique (PCAFT et PCHLO). Ce QTL est adjacent à celui qui est impliqué dans le contrôle de la quantité d'acide caftarique et d'acide chicorique mais ces deux régions ne se chevauchent pas, même avec un intervalle de confiance de - 2 LOD. D'ailleurs, les QTL contrôlant les quantités et les proportions par rapport à l'ensemble des acides phénoliques ne co-localisent pas toujours. On retrouve par exemple deux QTL communs pour PCHLO et ACHLO (sur le LG2 et le LG9) et un seul pour PCAFT et ACAFT (LG9). Le reste des QTL PCHLO et PCAFT se retrouvent sur les LG1, LG2, LG4, LG6, LG7 et LG8. Par ailleurs, les QTL de PCHLO et PCAFT co-localisent pour deux QTL (LG2, LG9) sur 5 détectés, supposant ici des bases génétiques communes quant à la régulation des hydroxycinnamates. Pour les QTL PCAFT et PCHLO en commun sur le LG9, le R^2 de $PCHLO_{LG9}$ est égale à 20,4% alors qu'il est seulement de 7% pour $PCAFT_{LG9}$. Cette différence supposerait que cette région chromosomique soit plus impliquée dans la régulation de l'acide chlorogénique que dans celle de l'acide caftarique. Dans le tableau 13, on note que le parent K28 porte l'allèle impliqué dans l'effet du QTL au marqueur le plus proche et que cet effet est antagoniste pour PCHLO (K28↘) et PCAFT (K28↗). Cette tendance n'est pas sans rappeler la corrélation négative (bien que très faible mais tout de même significative) entre ces deux caractères, souligné dans le chapitre I. Ces éléments laissent supposer que ce marqueur est lié à un gène codant une protéine (enzyme ou facteur de transcription) « charnière » entre la biosynthèse d'acide caftarique et d'acide chlorogénique et donc, à

priori, non une protéine impliquée dans les premières étapes de la voie de biosynthèse des phénylpropanoïdes.

Nous pouvons également remarquer que le nombre de QTL détecté pour PCHLO est supérieur à celui détecté pour ACHLO et qu'il en est de même pour les QTL de PCAFT et ACAFT. Cette différence suggérerait que la régulation des molécules étudiées par rapport à l'ensemble des acides phénols est un caractère plus complexe que leur simple production.

A notre connaissance, aucun article relatant la détection de QTL liés à la concentration en acide chicorique ou caftarique n'a été publié pour l'instant. Par contre, chez différentes espèces, plusieurs analyses QTL ont été effectuées pour la concentration en chlorogénate (Szalma *et al.*, 2002; Werij *et al.*, 2007), la capacité antioxydante et la concentration en polyphénols totaux (Rousseaux *et al.*, 2005; Jin *et al.*, 2009; Frary *et al.*, 2010).

Les analyses QTL associées à la concentration en acide chlorogénique à partir de chaume de maïs pour trois populations F2 ont révélés 4 à 6 QTL selon la population (Szalma *et al.*, 2002). Les R^2 correspondants à ces QTL varient de 2,9 à 42,7% alors que le maximum, parmi les trois QTL ACHLO détectés dans notre travail, est de 27,2% (ACHLO_{LG9}). Par ailleurs, pour une population de 249 haploïdes doublés (pomme de terre), l'étude de la variation phénotypique de trois formes d'acide chlorogénique (acide 3-, 4-, 5-O-caféoylquinique) n'a révélé qu'un seul QTL ($R^2 = 21\%$, LOD score 3,8) et uniquement pour la forme 5-, qui correspond à celle que nous étudions (Werij *et al.*, 2007).

L'étude de Jin *et al.* (2009) a révélé, trois QTL correspondant à l'activité antioxydante (grain de riz, 117 haploïdes doublés) expliquant 32,76% de la variation phénotypique, valeur proche de la notre (31,1% avec quatre QTL AAR). Leur analyse n'a pas détecté de QTL commun entre l'activité antiradicalaire et la concentration en composés phénoliques totaux malgré une corrélation correcte entre ces deux caractères ($r = 0,446$). C'est également le cas des caractères AAR et ACAFT ou ACHIC qui n'ont aucun QTL commun bien qu'une corrélation positive significative ait été révélée.

Les travaux de Frary *et al.* (2010) et de Rousseaux *et al.* (2005) concernant des QTL associés à l'activité antiradicalaire sur des lignées d'introgession de tomates sont intéressants dans la mesure où ils confrontent les résultats des analyses QTL dans des conditions de stress salin ou sur plusieurs saisons et milieu, respectivement. Dans le premier cas, 35 QTL ont été détectés pour la capacité antioxydante dont onze en commun pour les deux conditions (stress salin/conditions contrôlées) ; pour la quantité de phénols totaux, 32 QTL ont été révélés et seulement cinq sont communs aux deux conditions. Dans la publication de Rousseaux *et al.*,

l'analyse QTL faite sur trois années a permis d'identifier en tout cinq QTL liés à l'activité antioxydante la première année. Seulement trois de ces QTL ont été retrouvés sur deux des trois autres saisons, les QTL restants étant spécifiques à l'année du prélèvement. Pareillement, neuf QTL ont été détectés la première année pour les phénols totaux, et seulement deux d'entre eux ont été retrouvés l'année suivante. Il serait ainsi intéressant de comparer les vingt QTL trouvés au cours de ces travaux de thèse à ceux que l'on pourrait révéler dans d'autres environnements, avec des répétitions annuelles ou sous des conditions de stress spécifiques. Cela nous permettrait de savoir s'ils existent des loci généralistes (exprimés dans plusieurs types de conditions), et des QTL spécialistes, liés à l'induction par des stress, donc éventuellement constitués de gènes s'exprimant dans ces conditions spécifiques.

III.2.B. Co-localisation des QTL avec des gènes candidats

Sur les vingt QTL détectés, huit co-localisent avec sept gènes candidats : PCAFT_{LG1} et AAR_{LG1-I} avec *CAD* (cinnamyl-alcohol-dehydrogenase), ACHLO_{LG1} avec *C4H* (cinnamate-4-hydroxylase), ACHIC_{LG3} avec *NAD* (NADPH-cytochrome P450 reductase), PCAFT_{LG7} avec *AAE* (beta-xylosidase, acyl activating enzyme), PCAFT_{LG8} avec *4CL* (4-coumarate-CoA-ligase), ACHIC_{LG8} avec *C3H* (p-coumarate 3-hydroxylase) et PCAFT_{LG9} avec *HMT* (= OC0603, homocystéine S-méthyltransferase). Tous ces gènes candidats sont impliqués dans la voie de biosynthèse des phénylpropanoïdes (figure 22). Aucun d'entre eux n'est un facteur de transcription.

Les QTL PCAFT_{LG1}, PCAFT_{LG7} et PCAFT_{LG9} co-localisent chacun avec un gène candidat ne codant pas pour une enzyme intervenant en amont de la synthèse de l'acide caféique ou tartarique, précurseur de l'acide caftarique. Au contraire, l'enzyme CAD catalyse la formation de l'alcool sinapylique et de l'alcool coniférylique, précurseur d'unités monolignols comme le guaicyl et le syringyl, composants de la lignine. HMT est une enzyme impliquée dans la biosynthèse de la S-adénosyl-L-méthionine, substrat des enzymes OMT (*O*-méthyltransferase) et CCoAOMT (cafeoyl-CoA 3-*O*-méthyltransferase), catalysant la formation de féruloyl CoA et de synapaldéhyde, respectivement, précurseurs également des unités monolignols. AAE est une enzyme catalysant la biosynthèse du coumarinate, cette voie se dirigeant vers la synthèse des coumarines (figure 22). Ces deux voies (vers les lignines et vers les coumarines) sont des voies concurrentes à la formation de l'acide caftarique, synthétisé à

partir de l'acide caféique et de l'acide tartarique. Les R^2 correspondant à ces QTL sont assez faibles (8,2% pour le QTL associé à *CAD*, 3,8% pour le QTL associé à *AAE* et 7 % pour le QTL associé à *HMT*). On peut interpréter ces résultats comme représentatifs des différentes voies possibles pour les précurseurs de l'acide caftarique, régulant ainsi sa production :

- l'acide cinnamique, molécule au carrefour de la synthèse des acides phénols, des coumarines et des flavonoïdes.
- l'acide caféique, précurseurs de dérivés comme les hydroxycinnamates ou de dérivés caféoyl aboutissant à la biosynthèse de lignines.

Pour le caractère PCAFT, deux autres QTL ont été trouvés : PCAFT_{LG2} ($R^2 = 10,4\%$), PCAFT_{LG8} ($R^2 = 15\%$). Le QTL PCAFT_{LG8} est associé au R^2 le plus important parmi les cinq QTL détectés et co-localise avec le gène candidat *4CL*. Ce gène code pour l'enzyme 4-coumarate-CoA-ligase, enzyme dont un des substrats est l'acide caféique (figure 22). Cette implication plus directe avec un des précurseurs de l'acide caftarique corrobore avec le fait que le QTL associé est celui qui montre la part d'explication de la variation phénotypique la plus élevée (15%) pour ce caractère.

Aucun gène candidat ne se situe dans les intervalles de confiance des QTL PCHLO. Par contre, ACHLO_{LG1} co-localise avec le gène candidat *C4H*. Ce gène code pour la seconde enzyme impliquée dans le métabolisme des phénylpropanoïdes : la cinnamate-4-hydroxylase, membre de la famille des enzymes de type cytochrome P450 et ayant pour substrat l'acide cinnamique, précurseurs des hydroxycinnamates. Aucun QTL détecté pour ACHLO ne co-localise avec *HCT*, cartographié sur le LG5, pourtant présumé excellent gène candidat impliqué dans la variation des quantités d'acide chlorogénique (Hoffmann *et al.*, 2003; Hoffmann *et al.*, 2004). Ce gène ne semble donc pas directement impliqué dans la variation des teneurs en acide chlorogénique dans notre descendance.

Le même discours peut être tenu pour les gènes candidats *CHS* et *CHI* qui codent pour des enzymes clé de l'embranchement entre les dérivés caféoyl et la voie de biosynthèse des flavonoïdes (cf. figure 3, chapitre II). Même si *CHS* et *CHI* n'ont pas pu être inclus dans la carte « oxychic », leur localisation approximative a été rajoutée entre les marqueurs B77 et ataa456 (LG1) pour *CHI* et entre agga214 et *CCR* (LG7) pour *CHS*. A proximité de *CHI*, plusieurs QTL ont été détectés : PCAFT_{LG1}, AAR_{LG1-I}, AAR_{LG1-II}. Sur le LG7, c'est le QTL PCAFT_{LG7} qui serait adjacent à la position de *CHS* telle que nous l'avons définie. La définition de nouvelles amorces correspondant à ces gènes candidats permettrait de les inclure

de façon correcte à la carte et de confirmer ou non la localisation des QTL détectés avec *CHS* et *CHI*.

Deux autres gènes candidats codant pour une enzyme de type cytochrome P450 ont été cartographiés : *C3H* (*p*-coumarate 3 hydroxylase) et le gène candidat baptisé *NAD* sur la carte « oxychic » pour NADPH-cytochrome P450 réductase. *C3H* est compris dans l'intervalle de confiance du QTL $ACHIC_{LG8}$. Cette enzyme est impliquée directement dans la formation de l'acide caféique, précurseur de l'acide chicorique mais aussi dans la formation d'autres dérivés caféoyl comme l'acide chlorogénique (figure 22). Ce QTL reste pourtant un QTL à effet mineur avec un R^2 de 6,8%.

NAD correspond au gène *AR2* chez *Arabidopsis thaliana* et dont l'implication dans le début de la voie de biosynthèse des phénylpropanoïdes a été mise en valeur par Mizutani et Ohta en 1998 (Mizutani & Ohta, 1998). Ce gène candidat co-localise avec $ACHIC_{LG3}$ ($R^2 = 7,1\%$).

Comme pour $PCAF_{LG1}$, le gène candidat *CAD* est compris dans l'intervalle de confiance d' AAR_{LG1-I} ($R^2 = 7,6\%$). L'enzyme CAD (cinnamyl alcohol dehydrogenase) catalyse les premières étapes vers la formation des monolignols, précurseurs des lignanes et lignines. Dans un premier temps, cette co-localisation entre un QTL associé au caractère AAR et *CAD* pourrait faire attribuer à des monolignols, lignanes ou lignines une part importante de l'activité antiradicalaire trouvée dans nos échantillons. Cependant, les monolignols ne sont pas accumulés en grande quantité chez les végétaux car ce sont des composés peu stables et toxiques pour la plante. Par contre, les monolignols glucosides semblent être la forme de stockage et de transport des monolignols (Whetten & Sederoff, 1995). A notre connaissance, aucun papier relatant l'activité antiradicalaire des alcools monolignols ou des monolignols glucosides n'a été publié. D'un autre côté, plusieurs publications démontrent la capacité antioxydante des lignanes (Fauré *et al.*, 1990, Kitts *et al.*, 1999) et des lignines (Dizhbite *et al.*, 2004; Ugartondo *et al.*, 2008). Les lignines sont des composants de la paroi cellulaire et l'extraction passive (sans broyage donc sans cassure de la paroi) subit par nos échantillons permet d'exclure l'implication des lignines dans le pouvoir antioxydant détecté. Concernant les lignanes, la publication de Milder *et al.* (2005) répertorie la chicorée parmi les plantes ayant de faible taux de lignanes (Milder *et al.*, 2005). L'extraction des lignanes selon notre méthode (mélange d'eau et d'éthanol) ne peut pas être exclue selon l'article de Zhang *et al.* (2007) qui relate une extraction des lignanes de ce type à partir de graine de lin. Cependant,

les auteurs de cette publication ont testé des températures d'extraction entre 20 et 40°C et non à 4°C comme ce fut le cas pour nos échantillons (Zhang *et al.*, 2007). Etant donné le peu de lignanes détectés chez la chicorée et le doute quant à leur extraction par notre protocole, la co-localisation entre *CAD* et *AAR_{LG1-1}* mettrait plutôt en évidence la compétition entre la formation des alcools monolignol et l'accumulation d'acides phénols, précurseurs des premières et molécules aux propriétés antioxydantes largement reconnues.

Tous les QTL détectés lors de notre étude ne sont pas associés à des gènes candidats. Un des premiers facteurs limitant est le manque de séquences disponibles chez *Cichorium intybus*. La création d'amorces dégénérées à partir d'*Arabidopsis* ou d'autres astéracées apparaît comme une première solution à ce manque, notamment pour le gène *HQT* (hydroxycinnamoyl-CoA-quinase) directement impliqué dans la formation de l'acide chlorogénique (Niggeweg *et al.*, 2004; Lepelley *et al.*, 2007) et isolé, caractérisé et cartographié dans la carte génétique de l'artichaut (Comino *et al.*, 2009). Par ailleurs, parmi les gènes codant pour des enzymes impliquées dans la biosynthèse des hydroxycinnamates étudiés, certains n'ont pas présenté de polymorphisme avec les amorces définies autour des introns et n'ont donc pas pu être cartographiés. Il serait donc intéressant de définir d'autres amorces pour les séquences correspondant aux gènes codant pour les enzymes ADT, PAL, SAM, CCoAOMT, MLS ou les facteurs de transcription MYB4, PAP1, PAP2 (cf. introduction bibliographique et tableau 3, chapitre II). D'autre part, les QTL non associés à un gène candidat peuvent également impliquer d'autres facteurs génétiques, comme des facteurs de transcription non encore identifiés et impliqués dans la régulation et/ou production d'acide chlorogénique, chicorique et caftarique. Notons en plus que pour ces deux dernières molécules, l'enzyme qui permet leur formation à partir de l'acide caféique et de l'acide tartarique n'est pas encore connue.

III.2.C. Co-localisation avec des QTL du poids secs résiduels

Les valeurs phénotypiques correspondant aux caractères ACAFT, ACHLO, ACHIC et AAR ont été divisées par le poids comme expliqué dans le Matériel et Méthode du chapitre I. En effet, nous avons choisi de standardiser nos mesures par la surface foliaire prélevée mais aussi par le poids sec résiduel (PSR) des six rondelles, proportionnel au poids frais. Cette étape avait pour but d'éviter d'attribuer des variations phénotypiques aux caractères étudiés provenant d'une différence de poids prélevé au départ.

Une analyse IM avec les données PSR, présentée dans l'annexe 5, nous a permis de détecter quatre QTL : PSR_{LG3} ($R^2 = 9,9\%$), PSR_{LG5-I} ($R^2 = 24,2\%$), PSR_{LG5-II} ($R^2 = 19,1\%$), PSR_{LG6} ($R^2 = 11\%$). Seul PSR_{LG6} est supérieur au seuil_{GW}. Si l'on compare avec l'analyse IM des caractères ACAFT, ACHLO, ACHIC et AAR, on constate que (i) le QTL le PSR_{LG6} co-localise avec celui retrouvé pour AAR ; (ii) les courbes PSR et ACHIC sont similaires notamment sur le LG8.

Cette similarité dans les courbes IM nous interroge sur le fait d'avoir créé ou pas une variabilité en divisant les données phénotypiques par le PSR. En procédant à cette correction, nous avons admis que l'activité antiradicalaire ou la quantité en acide caftarique, chlorogénique ou chicorique était répartie de façon homogène dans l'ensemble des cellules végétales constituant les rondelles de chicorée prélevées, donc proportionnelle au poids (plus de poids = plus de cellules = plus de molécules). Cependant, si les composés étudiés ne s'accumulent que dans les cellules épidermiques et que la variation du PSR était due aux nombres de cellules parenchymateuses, la prise en compte du poids n'est pas la meilleure façon de standardiser nos données. Dans ces conditions, l'implication des QTL AAR_{LG3} et $ACHIC_{LG8}$ dans le contrôle de l'activité antiradicalaire et du métabolisme de l'acide chicorique sont à considérer avec précaution.

III.2.D. Travaux parallèles à partir du même croisement

A partir du même croisement K28xK59, des loci associés à d'autres caractères ont été étudiés au sein du laboratoire : le locus lié à l'auto-incompatibilité est situé sur le LG2 (à 0,5 cM de l'extrémité du groupe de liaison), les loci liés à la stérilité mâle nucléaire (LG5, 23 cM) et les loci liés à l'embryogénèse somatique (LG1, LG2, LG3, LG4, LG6 et LG8). Certains

QTL liés à l'embryogénèse somatique et certains QTL détectés lors de nos travaux se chevauchent, sur les LG1, LG2, LG6 et LG8 suggérant des régions chromosomiques pléiotropes.

III.3. QTL à effet interactif

A l'instar de Szalma *et al.* (2002), nous avons choisi de tester les interactions deux à deux entre tous les marqueurs de la carte « oxychic » et non pas entre les marqueurs uniquement associés aux QTL (Kliebenstein *et al.*, 2001; Barrière *et al.*, 2008). Cela nous a permis de détecter des phénomènes d'épistasie (cf. tableau 11) parmi des marqueurs ne correspondant pas à ceux qui sont significativement liés aux QTL détectés, mis à part le marqueur sw1E08.2 du LG4. Ils n'ont pas d'effet sur le phénotype individuellement et n'ont ainsi pas pu être détectés par une analyse QTL de type IM dont le principe est de ne tester que les effets sur le phénotype, dans un intervalle entre deux loci. Dans le cas d'épistasie, quand l'interaction d'un marqueur M1 (génétiquement lié au locus 1) avec un marqueur M2 (génétiquement lié au locus 2) est significative, cela indique que l'effet sur le phénotype de chacun de ces loci 1 et 2 dépend du génotype à l'autre locus, comme l'ont montré les graphes de la figure 21.

Le seuil de probabilité choisi pour valider ces interactions est déjà relativement rigoureux : $p < 0,0001$. Notons cependant que l'interaction entre FDCB40 et ttat336 (LG8, PCHLO) a une signification très élevée avec une valeur de p égale à $5,36.10^{-8}$. Cette interaction est la seule qui est inférieure au seuil de signification de Bonferroni (seuil très conservatif), qui se calcule en divisant le seuil statistique 0,05 par le nombre d'interactions ($4,99.10^{-6}$ pour 10011 interactions).

Dans un article de synthèse publié en 2009 (Kliebenstein, 2009), les phénomènes d'épistasie sont classés selon trois catégories : (i) les épistasies fonctionnelles, (ii) les épistasies dues aux familles de gènes et (iii) les épistasies dites « diffuses ».

Les épistasies fonctionnelles sont évoquées lorsque les gènes sont connectés par une voie de biosynthèse, comme c'est le cas pour notre étude. Mis à part l'implication de plusieurs gènes codant différentes enzymes pour la régulation/formation d'un composé, ce type d'épistasie peut également renvoyer à l'interaction entre facteurs de transcription et gènes codant des enzymes, les premiers régulant les taux de transcrits des dernières (Kliebenstein, 2009).

Les épistasies intra famille de gènes pourraient également être associées à notre cas, bien que les gènes candidats codant des enzymes de la famille des cytochromes P450 et cartographiés (C3H, C4H, NAD) n'aient pas montré d'interaction significative avec d'autres marqueurs répartis sur les neufs groupes de liaison. Ce type d'épistasie est rencontré par exemple lorsqu'une réaction enzymatique est contrôlée par plusieurs gènes appartenant à la même famille et que chacun de ces gènes est impliqué dans une part de l'activité enzymatique (Kliebenstein, 2009).

L'épistasie diffuse se révèle, quant à elle, dans le cas d'un réseau métabolique complexe. Ce type de réseau avec de nombreuses connections entre les différentes voies de biosynthèse est donc assez résistant aux altérations éparses. Cependant, certaines combinaisons de mutations indépendantes vont affecter des phénotypes autres que ceux directement touchés par la mutation. Contrairement à l'épistasie fonctionnelle qui pourrait s'appliquer à notre cas, l'épistasie diffuse est souvent difficile à détecter car elle génère des phénotypes extrêmes, souvent traités comme des données statistiquement aberrantes (« outliers ») et éliminées de l'analyse (Kliebenstein, 2009).

IV. Conclusion

Vingt QTL ont été détectés pour les caractères ayant montré une variabilité significative suite au prélèvement foliaire sur une population F1' de 201 individus répétés en trois, quatre ou cinq blocs (954 chicorées). Parmi les QTL révélés par l'analyse MQM, $ACHLO_{LG9}$ et $PCHLO_{LG9}$ peuvent être considérés comme des QTL majeurs et AAR_{LG6} , $PCAFT_{LG7}$, $ACHIC_{LG8}$ comme des QTL mineurs. La détection des QTL dépend de plusieurs facteurs incluant la taille de la population, la densité en marqueurs de la carte génétique, la précision des valeurs phénotypiques, l'environnement et la méthode du choix du seuil de détection. La prise en compte de conditions restrictives sur ce dernier point permet de classer les QTL détectés en trois groupes, des plus robustes au moins robustes: (i) $ACHLO_{LG2}$, $PCHLO_{LG2}$, $ACHLO_{LG9}$, $PCAFT_{LG9}$ et $PCHLO_{LG9}$; (ii) $ACHLO_{LG1}$, $PCAFT_{LG2}$, AAR_{LG2} , $ACHIC_{LG3}$, $PCHLO_{LG4}$, AAR_{LG6} , $PCHLO_{LG6}$, $PCAFT_{LG8}$, $ACHIC_{LG8}$, $PCHLO_{LG8}$; (iii) $PCAFT_{LG1}$, AAR_{LG1-I} , AAR_{LG1-II} , $PCAFT_{LG7}$, $ACAFT_{LG9}$.

Huit QTL co-localisent avec sept gènes candidats et ceux-ci sont tous impliqués dans la voie de biosynthèse des phénylpropanoïdes (*CAD*, *C4H*, *NAD*, *AAE*, *4CL*, *C3H*, *HMT*). Aucun d'entre eux n'est un facteur de transcription. Cependant, des interactions entre

marqueurs ont été détectés hors marqueurs reliés aux QTL suggérant des phénomènes d'épistasie entre gènes non cartographiés sur la carte « oxychic » ou non encore identifiés comme ayant un rôle dans la biosynthèse de l'acide caftarique, chicorique et caféique (figure 23).

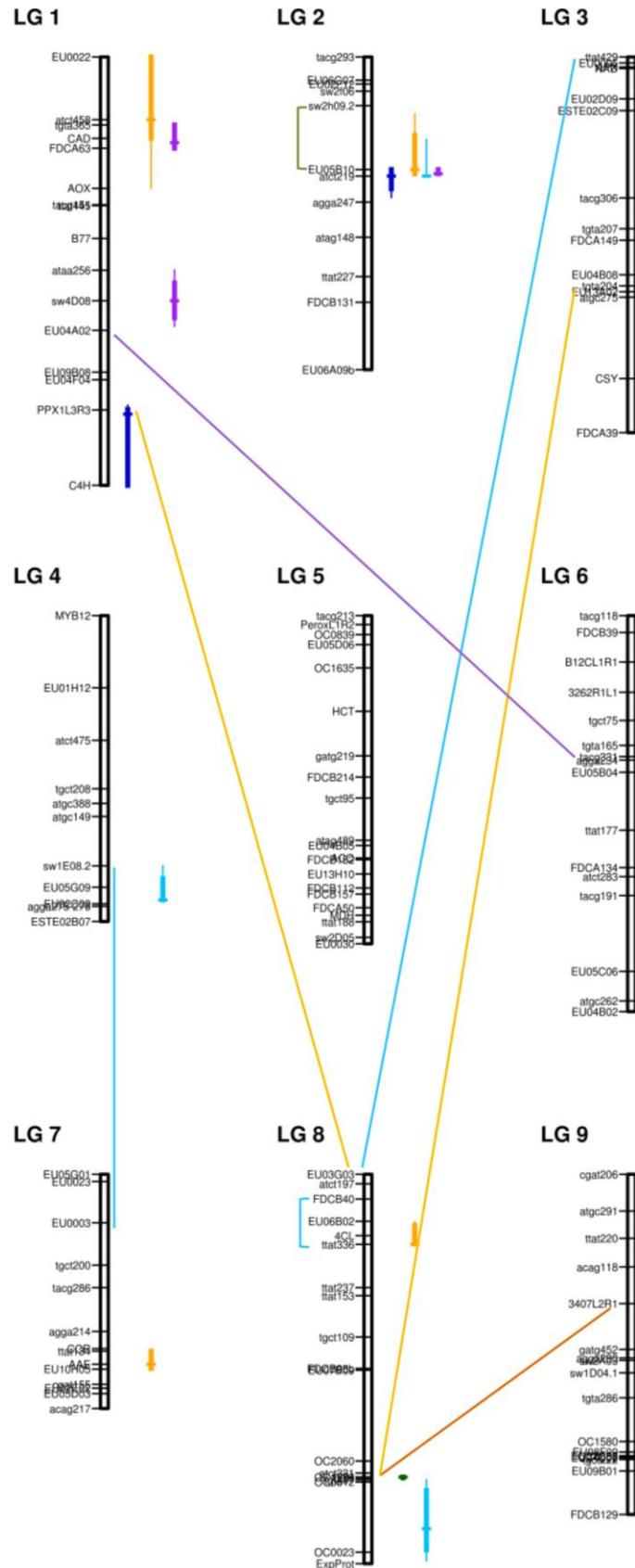


Figure 23. Figure récapitulative des 20 QTL à effet propre et des 8 QTL interactifs détectés. En orange foncé : QTL correspondant au caractère ACAFT ; en bleu roi : QTL ACHLO ; en vert : QTL ACHIC ; en orange clair : QTL PCAFT ; en bleu turquoise : QTL PCHLO ; en violet : QTL AAR. Pour les QTL à effet propre, le LOD score maximal est désigné par un trait horizontal, les intervalles de confiance à - 1 LOD et à - 2 LOD sont représentés par un trait vertical épais et fin, respectivement.

CONCLUSION & PERSPECTIVES

Le principe d'une analyse QTL est de relier statistiquement des données quantitatives (dont la variabilité est significative) aux génotypes des marqueurs composant une carte génétique. Dans notre cas, l'outil carte génétique était disponible (Cadalen *et al.*, 2010) et donc particulièrement intéressant pour étudier les régions du génome impliquées dans la biosynthèse des acides phénols détectés dans les feuilles de chicorées industrielles, encore non valorisées comme source d'antioxydants naturels.

Au commencement de cette étude, trois pics ont été identifiés à la CLHP puis par spectrométrie de masse comme étant l'acide caftarique, l'acide chlorogénique et l'acide chicorique. Ces résultats révélant l'acide chicorique comme le composé phénolique majeur, quelque soit la plante analysé, son âge ou la saison de prélèvement, concordent avec les résultats précédemment publiés sur la chicorée (Rees & Harborne, 1985; Mulinacci *et al.*, 2001; Llorach *et al.*, 2004; Innocenti *et al.*, 2005; Heimler *et al.*, 2007).

En fonction des identifications précédentes, un protocole original et économique a été établi pour extraire ces molécules. Il est en effet modulable selon les conditions de prélèvement (serre ou en champs), du nombre de plantes à analyser et du nombre de personnes impliquées. Le stockage en microplaque permet un gain de place important. Une méthode CLHP rapide et résolutif ainsi que la miniaturisation du test DPPH sont venus compléter cette méthode d'extraction, toujours dans l'optique de réduire les temps et les coûts d'analyse.

Le phénotypage s'est ensuite concentré sur les trois hydroxycinnamates détectés par CLHP et l'activité antiradicalaire mesurée via le test chimique DPPH ; une descendance de 201 individus (F1', K59*K28) a été consacrée à ce projet (trois à cinq répétitions disposées en bloc). Neuf caractères ont été pris en compte : l'aire des pics correspondant à l'acide caftarique / chlorogénique / chicorique (ACAFT / ACHLO / ACHIC), la somme des aires des pics correspondant aux trois molécules d'intérêt (ATOT), la proportion de l'aire du pic correspondant à l'acide caftarique / chlorogénique / chicorique par rapport à l'aire totale des pics du chromatogramme (PCAFT / PCHLO / PCHIC), la somme des proportions des aires des pics des trois molécules d'intérêt, par rapport à l'aire totale des pics du chromatogramme (PTOT) et enfin, l'activité antiradicalaire (AAR).

PCHIC et PTOT n'ont pas montré de variabilité phénotypique significative entre les individus étudiée ($p < 0,001$) alors que les sept autres caractères ont suggéré que cette population est bien adaptée à l'analyse QTL, comme les caractères quantitatifs choisis à la base du projet de cartographie (en relation avec l'embryogenèse somatique). De la

transgression a été observée pour chaque donnée (valeurs inférieures ou supérieures à celles de K59 et K28) ainsi qu'un effet bloc important et supérieur à l'effet individu. Cet effet de l'environnement a donc été pris en compte lors du calcul des moyennes ajustées (LSmeans) servant de valeurs phénotypiques pour les analyses QTL. Les héritabilités obtenues sont relativement faibles avec un maximum de 0,531 pour AAR, et un minimum de 0,296 pour ATOT.

En parallèle, nous avons construit une carte génétique spécialement dédiée à ce projet baptisé « oxychic ». Cette carte inclût 142 marqueurs dont 16 gènes candidats à priori impliqués dans la biosynthèse des hydroxycinnamates détectés dans les feuilles de chicorée : des gènes codant des facteurs de transcription et des enzymes de la voie de phénylpropanoïdes et de la dégradation de l'acide ascorbique. Parmi ces marqueurs, 36% présentent un polymorphisme SSR, 22 % un polymorphisme SSCP, 3% un polymorphisme HRM et 39% un polymorphisme AFLP. Les neuf groupes de liaison ont en moyenne 15,8 marqueurs et un intervalle moyen entre deux marqueurs de 6,08 cM. Le LG7 porte plusieurs marqueurs reflétant des loci en distorsion de ségrégation d'origine gamétique.

A partir de la carte génétique « oxychic » et des valeurs phénotypiques collectées, vingt QTL ont été détectés via la cartographie d'intervalle (IM), puis en fixant les marqueurs associés aux QTL comme co-facteurs (MQM). Pour les six caractères étudiés, un QTL a été détecté pour ACAFT ($R^2 = 16,4\%$), trois pour ACHLO ($R^2 = 50\%$), deux pour ACHIC ($R^2 = 13,9\%$), cinq pour PCAFT ($R^2 = 44,4\%$) et PCHLO ($R^2 = 60,7\%$) et quatre pour AAR ($R^2 = 31,1\%$). La prise en compte de deux seuils de détection (seuil_{LG} et seuil_{GW}) et les résultats du test de Kruskal-Wallis ont permis de classer les QTL en trois groupes, des plus robustes au moins robustes:

- ACHLO_{LG2}, PCHLO_{LG2}, ACHLO_{LG9}, PCAFT_{LG9} et PCHLO_{LG9} ;
- ACHLO_{LG1}, PCAFT_{LG2}, AAR_{LG2}, ACHIC_{LG3}, PCHLO_{LG4}, AAR_{LG6}, PCHLO_{LG6}, PCAFT_{LG8}, ACHIC_{LG8}, PCHLO_{LG8} ;
- PCAFT_{LG1}, AAR_{LG1-I}, AAR_{LG1-II}, PCAFT_{LG7}, ACAFT_{LG9}.

Parmi eux, ACHLO_{LG9} et PCHLO_{LG9} peuvent être considéré comme des QTL majeurs et AAR_{LG6}, PCAFT_{LG7}, ACHIC_{LG8} comme des QTL à effet mineur. Plusieurs QTL co-localisent entre eux : PCAFT_{LG1} avec AAR_{LG1-I}, PCAFT_{LG2} avec ACHLO_{LG2}, PCHLO_{LG2} et AAR_{LG2}, ACAFT_{LG9} avec ACHLO_{LG9} et PCAFT_{LG9} et PCHLO_{LG9} (66-68 cM). Huit QTL co-localisent avec sept gènes candidats : PCAFT_{LG1} et AAR_{LG1-I} avec *CAD*, ACHLO_{LG1} avec *C4H*, ACHIC_{LG3} avec *NAD*, PCAFT_{LG7} avec *AAE*, PCAFT_{LG8} avec *4CL*, ACHIC_{LG8} avec *C3H* et PCAFT_{LG9} avec *HMT*. Ces co-localisations renforcent l'implication supposée des

QTL détectés dans la régulation ou la formation des trois hydroxycinnamates identifiés. Ces gènes candidats sont impliqués dans la voie de biosynthèse des phénylpropanoïdes mais aucun d'entre eux n'est un facteur de transcription. En dehors des QTL détectés, des phénomènes d'épistasie ont été décelés pour les caractères ACAFT, ACHIC, PCAFT, PCHLO et AAR, suggérant l'implication de gènes non cartographiés sur la carte « oxychic » ou non encore identifiés comme ayant un rôle dans la biosynthèse de l'acide caftarique, chicorique et caféique ou l'activité antiradicalaire.

Les échantillons de la descendance F1' analysés en CLHP peuvent être caractérisés pour leur contenu en sesquiterpènes lactones (~ 210 nm). Si une variabilité phénotypique significative est détectée au sein de la population, une nouvelle analyse QTL peut être envisagée pour ces caractères. Il serait par exemple intéressant de constater que des QTL pour le caractère AAR co-localisent avec des QTL impliqués dans la biosynthèse des sesquiterpènes lactones, comme pourrait le laisser sous entendre la publication de Kim *et al.* (2001) qui attribue une -faible- activité antioxydante à ce type de molécules chez *Lactuca scariola*. En 2010, neuf sesquiterpènes lactones ont été testées au DPPH et seule la cichorioside a révélé une activité antiradicalaire (Han *et al.*, 2010). Cette molécule est la forme glucoside du sesquiterpène lactone 11 β -13 dihydrolactucine, déjà identifiée chez la chicorée (Seto *et al.*, 1988). Ces informations concernant le contrôle génétique des sesquiterpènes lactones est également très intéressant pour le secteur industriel qui sélectionne depuis des années les chicorées présentant le moins d'amertume pour satisfaire les consommateurs (Hance *et al.*, 2007). Par ailleurs, le protocole mis en place pourrait également intéresser le domaine industriel et s'étendre à d'autres légumes-feuilles que la chicorée.

Afin de compléter les résultats obtenus au cours de cette thèse, il serait intéressant de cartographier des gènes candidats responsables de la dégradation des molécules d'intérêt comme les gènes codant des polyphénols oxydases (Nüsslein *et al.*, 2000; Martinez & Duvnjak, 2006). Concernant les gènes candidats pour lesquels la recherche de polymorphisme a échoué (*ADT*, *CCoAOMT*, *MYB4*, *PAL*, *SAM*, *PAP1*, *PAP2*, *MLS*), la définition d'autres amorces devra être envisagée. Pour augmenter les chances de les situer sur la carte « oxychic » construite à partir du croisement K28xK59, il serait judicieux de chercher le polymorphisme également sur les autres croisements utilisés pour la carte consensus (Cadalen *et al.*, 2010). Ainsi, si du polymorphisme est détecté sur le croisement Rubis 118 ou BR, il

sera possible de connaître la région où se situe le gène candidat chez K28xK59 en fonction des marqueurs « ancrés ».

Pour le gène candidat *HQT*, dont il n'y avait pas de séquence correspondante chez *Cichorium intybus*, une alternative pour le cartographier sur la carte « oxychic » existe. En alignant les séquences correspondant à ce gène chez d'autres espèces, notamment l'artichaut dont la séquence *HQT* est disponible (Comino *et al.*, 2007), il est possible de produire des amorces dégénérées. Le produit d'amplification des amorces dégénérées pourra alors être cloné dans un plasmide puis séquencé afin de définir de nouvelles amorces spécifiques à la chicorée. La détection de polymorphisme permettra alors de l'inclure à la carte « oxychic ».

Cependant, un outil disponible au laboratoire SADV offre de nouvelles perspectives au travail présenté dans ce manuscrit. La banque BAC (Bactériel Artificial Chromosome) de chicorée (Gonthier *et al.*, 2010) peut effectivement permettre d'accroître le nombre de marqueurs informatifs pour les deux parents sur le LG9 afin de vérifier le bon ordre des marqueurs, comme cela a été développé dans le paragraphe Discussion du Chapitre III.

La deuxième perspective offerte par cet outil concerne la détection de QTL à effet majeur mais ne co-localisant pas avec un gène candidat, comme $ACHLO_{LG9}$ ($R^2 = 27,2\%$), $PCHLO_{LG9}$ ($R^2 = 20,4\%$). L'analyse QTL reste une façon grossière de localiser des loci impliqués dans la variation d'un caractère quantitatif. En plus, la distance entre le marqueur associé au QTL et le QTL peut être très importante. Pour localiser le QTL de manière plus précise, l'analyse QTL inversée ou RQM (Reverse QTL Mapping, Wye *et al.*, 2000) peut être envisagée. En effet, la première étape de cette méthode est l'identification du QTL majeur impliqué dans le caractère d'intérêt à partir d'une population d'environ 200 individus, comme la nôtre. Les marqueurs flanquants des QTL détectés sont alors utilisés pour génotyper une descendance beaucoup plus large (2000 individus par exemple) produite à partir des mêmes parents. Les individus recombinants pour les marqueurs flanquants le QTL majeur mais non recombinants pour les autres QTL détectés sont alors sélectionnés. Ils sont ensuite génotypés pour un maximum de marqueurs cartographiés dans cette région de recombinaison. Pour saturer la zone en question, la technique AFLP peut être utilisée. Les individus recombinants sont aussi phénotypés et ces nouvelles données de phénotypage sont à associer statistiquement aux nouvelles données de génotypage pour préciser d'avantage la localisation du QTL.

Ensuite, les marqueurs placés à quelques cM du locus seraient utilisés comme sonde : ils permettraient de détecter les clones positifs de la banque BAC, c'est-à-dire les clones possédant un insert correspondant à la région bordant le locus recherché. Les extrémités des

inserts pourraient alors être séquencées. Si ces séquences s'alignent avec un gène impliqué dans la voie de biosynthèse de l'acide chlorogénique (comme *HQT* par exemple, bon candidat à l'explication des QTL majeurs détectés), nous pourrions procéder au clonage positionnel. Autrement dit, il faudrait trouver des régions de chevauchement pour les extrémités des inserts séquencées pour situer le gène d'intérêt (sens et distance) sur la carte génétique. L'ultime validation de l'implication du gène dans le contrôle du caractère ciblé serait la restauration du gène chez un mutant ou l'étude de sa sous-expression. Le QTL ACHLO_{LG9} semble être le meilleur candidat pour ce type de démarche : le R^2 est fort, le LOD score maximal dépasse le seuil_{GW} et il avait déjà été mis en valeur par le test Kruskal-Wallis.

D'un point de vue fondamental, ces nouvelles données sont d'autant plus originales que relativement peu d'espèces accumulent l'acide chicorique et l'acide caftarique. Ces travaux sont donc un premier pas vers la compréhension du contrôle génétique du métabolisme de ces molécules peu étudiées et qui font partie du panel de composés qui jouent un rôle important dans le développement de la plante, face aux stress environnementaux comme les ultra-violets ou les blessures.

Pour le secteur industriel, les résultats apportés par cette étude pourraient intéresser les sélectionneurs qui souhaiteraient associer les allèles favorables au contenu polyphénolique et à la capacité antioxydante tout en conservant les caractéristiques agronomiques acceptables des chicorées commerciales. Soit le gène impliqué est directement introgressé par rétrocroisement d'un individu donneur vers un individu receveur, soit les marqueurs flanquants le QTL permettront de créer ce changement phénotypique.

REFERENCES
BIBLIOGRAPHIQUES

- Abe H, Nakano M, Nakatsuka A, Nakayama M, Koshioka M, Yamagishi M. 2002.** Genetic analysis of floral anthocyanin pigmentation traits in Asiatic hybrid lily using molecular linkage maps. *Theoretical and Applied Genetics* **105**(8): 1175-1182.
- Adlercreutz H, Mazur W. 1997.** Phyto-oestrogens and Western diseases. *Annals of Medicine* **29**(2): 95-120.
- Alonso-Blanco C, Bentsink L, Hanhart CJ, Blankestijn-de Vries H, Koornneef M. 2003.** Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. *Genetics* **164**: 711-729.
- Areshchenkova T, Ganai MW. 1999.** Long tomato microsatellites are predominantly associated with centromeric regions. *Genome* **42**: 536-544.
- Atienza SG, Satovic Z, Petersen KK, Dolstra O, Martín A. 2003.** Identification of QTLs influencing agronomic traits in *Miscanthus sinensis* Anderss. I. Total height, flag-leaf height and stem diameter. *Theoretical and Applied Genetics* **107**(1): 123-129.
- Aziz A, Sauve R. 2008.** Genetic mapping of *Echinacea purpurea* via individual pollen DNA fingerprinting. *Molecular breeding* **21**(2): 227-232.
- Barrière Y, Thomas J, Denoue D. 2008.** QTL mapping for lignin content, lignin monomeric composition, *p*-hydroxycinnamate content, and cell wall digestibility in the maize recombinant inbred line progeny F838 × F286. *Plant Science* **175**(4): 585-595.
- Baum BR, Mechanda S, Livesey JF, Binns SE, Arnason JT. 2001.** Predicting quantitative phytochemical markers in single *Echinacea* plants or clones from their DNA fingerprints. *Phytochemistry* **56**(6): 543-549.
- Benov L. 2001.** How superoxide radical damages the cell. *Protoplasma* **217**(1): 33-36.
- Biedenkapp H, Borgmeyer U, Sippel AE, Klempnauer K-H. 1988.** Viral *myb* oncogene encodes a sequence-specific DNA-binding activity. *Nature* **335**: 835-837.
- Blervacq A, Dubois T, Dubois J, Vasseur J. 1995.** First divisions of somatic embryogenic cells in *Cichorium* hybrid "474". *Protoplasma* **186**: 163-168.
- Borevitz JO, Xia Y, Blount J, Dixon RA, Lamb C. 2000.** Activation tagging identifies a conserved MYB regulator of phenylpropanoid biosynthesis. *Plant Cell* **12**(12): 2383-2394.
- Brummer EC, Bouton JH, Kochert G. 1993.** Development of an RFLP map in diploid alfalfa. *Theoretical and Applied Genetics* **86**(2): 329-332.
- Bunea A, Andjelkovic M, Socaciu C, Bobis O, Neacsu M, Verhé R, Camp JV. 2008.** Total and individual carotenoids and phenolic acids content in fresh, refrigerated and processed spinach (*Spinacia oleracea* L.). *Food Chemistry* **108**(2): 649-656.
- Burke JM, Tang S, Knapp SJ, Rieseberg LH. 2002.** Genetic Analysis of Sunflower Domestication. *Genetics* **161**(3): 1257-1267.
- Cadalen T, Mörchen M, Blassiau C, Clabaut A, Sheer I, Hilbert J-L, Hendriks T, Quillet M-C. 2010.** Development of SSR markers and construction of a consensus genetic map for chicory (*Cichorium intybus* L.). *Molecular Breeding* **25**(4): 699-722.
- Cantuti-Castelvetri I, Shukitt-Hale B, Joseph JA. 2000.** Neurobehavioral aspects of antioxidants in aging. *International Journal of Developmental Neuroscience* **18**: 367-381.
- Chalker-Scott L. 1999.** Environmental significance of anthocyanins in plant stress responses. *Photochemistry and Photobiology* **70** (1): 1-9.
- Chalker-Scott L, Fuchigami LH 1989.** The role of phenolic compounds in plant stress responses. In: P. H. Li ed. *In Low Temperature Stress Physiology in Crops*. Boca Raton, FL. : CRC Press, 68-76. .
- Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li PH, Liu J, Hatfield J, Main D, Farrar K, Henderson C, Barnett**

- L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo M, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA. 2002. An integrated physical and genetic map of the rice genome. *Plant Cell* **14**(3): 537-545.
- Chkhikvishvili ID, Kharebava GI. 2001. Chicoric and chlorogenic acids in plant species from Georgia. *Applied Biochemistry and Microbiology* **37**(2): 188-191.
- Christie PJ, Alfenito MR, Walbot V. 1994. Impact of low-temperature stress on general phenylpropanoid and anthocyanin pathways: enhancement of transcript abundance and anthocyanin pigmentation in maize seedlings. *Planta* **194**(4): 541-549.
- Churchill GA, Doerge RW. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* **138**: 963 - 971.
- Clé C, Hill LM, Niggeweg R, Martin CR, Guisez Y, Prinsen E, Jansen MAK. 2008. Modulation of chlorogenic acid biosynthesis in *Solanum lycopersicum*; consequences for phenolic accumulation and UV-tolerance. *Phytochemistry* **69**: 2149-2156.
- Clifford MN. 2000. Chlorogenic acids and other cinnamates - nature, occurrence, dietary burden, absorption and metabolism. *Journal of the Science of Food and Agriculture* **80**(7): 1033-1043.
- Clifford MN, Scalbert A. 2000. Ellagitannins - nature, occurrence and dietary burden. *Journal of the Science of Food and Agriculture* **80**(7): 1118-1125.
- Coinu R, Carta S, Urgeghe PP, Mulinacci N, Pinelli P, Franconi F, Romani A. 2007. Dose-effect study on the antioxidant properties of leaves and outer bracts of extracts obtained from Violetto di Toscana artichoke. *Food chemistry* **101**(2): 524-531.
- Comino C, Hehn A, Moglia A, Menin B, Bourgaud F, Lanteri S, Portis E. 2009. The isolation and mapping of a novel hydroxycinnamoyltransferase in the globe artichoke chlorogenic acid pathway. *BMC Plant Biology* **9**(1): 30.
- Comino C, Lanteri S, Portis E, Acquadro A, Romani A, Hehn A, Larbat R, Bourgaud F. 2007. Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in *Cynara cardunculus* L. *BMC Plant Biology* **7**(14): doi:10.1186/1471-2229-1187-1114.
- Costantini L, Battilana J, Lamaj F, Fanizza G, Grando MS. 2008. Berry and phenology-related traits in grapevine (*Vitis vinifera* L.): from Quantitative Trait Loci to underlying genes. *BMC Plant Biology* **8**(1): 38.
- Darvasi A, Weinreb A, Minke V, Weller JI, Soller M. 1993. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134**(3): 943-951.
- Darvill AG, Albersheim P. 1984. Phytoalexins and their Elicitors - A Defense Against Microbial Infection in Plants. *Annual Review of Plant Physiology* **35**(1): 243-275.
- Davey MW, Kenis K, Keulemans J. 2006. Genetic control of fruit vitamin C contents. *Plant Physiology* **142**: 343-351.
- De Paolis A, Pignone D, Morgese A, Sonnante G. 2008. Characterization and differential expression analysis of artichoke phenylalanine ammonia-lyase-coding sequences. *Physiologia Plantarum* **132**(1): 33-43.
- 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. *Molecular breeding* **3**: 415-425.
- DeBolt S, Cook DR, Ford CM. 2006. L-tartaric acid synthesis from vitamin C in higher plants. *Proceedings of the National Academy of Sciences of the United States of America* **103**(14): 5608-5613.

- Dechaine JM, C. BJ, A. CM, J. SG, R. B, J. KS, M. BJ. 2009.** Fitness effects and genetic architecture of plant-herbivore interactions in sunflower crop-wild hybrids. *New Phytologist* **184**(4): 828-841.
- Deluc L, Barrieu F, Marchive C, Lauvergeat V, Decendit A, Richard T, Carde J-P, Merillon J-M, Hamdi S. 2006.** Characterization of a Grapevine R2R3-MYB Transcription Factor That Regulates the Phenylpropanoid Pathway. *Plant Physiology* **140**(2): 499-511.
- Desprez B, Delesalle L, Dhellemmes C, Desprez M. 1994.** Genetics and breeding of industrial chicory. *Comptes Rendus de l'Académie d'Agriculture de France* **80**: 47-62.
- Dixon R. 2005.** Engineering of plant natural product pathways. *Current Opinion in Plant Biology* **8**(3): 329 - 336.
- Dizhbite T, Telysheva G, Jurkjane V, Viesturs U. 2004.** Characterization of the radical scavenging activity of lignins-natural antioxidants. *Bioresource Technology* **95**(3): 309-317.
- Dong Y, Tsuzuki E, Kamiunten H, Lin D, Terao H, Matsuo M, Cheng S. 2005.** Molecular genetic analysis of QTLs for ferulic acid content in dried straw of rice (*Oryza sativa* L.). *Biochemical Genetics* **43**(1): 25-34.
- Dong Y, Xu J, Xiao K, Zhang Y, Zhang J, Luo L, Matsuo M. 2008.** Genomic regions associated with the degree of red coloration in pericarp of rice (*Oryza sativa* L.). *Journal of Cereal Science* **48**(2): 556-560.
- 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**: 669-675.
- Dubois T, Guedira M, Dubois J, Vasseur J. 1991.** Direct somatic embryogenesis in leaves of *Cichorium*. An histological and SEM study of early stages. *Protoplasma* **162**(120-127).
- Ebrahimi A, Maury P, Berger M, Calmon A, Grieu P, Sarrafi A. 2009.** QTL mapping of protein content and seed characteristics under water-stress conditions in sunflower. *Genome* **52**: 419-430.
- Eenink AH. 1981.** Compatibility and incompatibility in Witloof chicory. *Euphytica* **30**(1): 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**(3): 773-786.
- Fauré M, Lissi E, Torres R, Videla LA. 1990.** Antioxidant activities of lignans and flavonoids. *Phytochemistry* **29**(12): 3773-3775.
- Fontaine AS, Briand M, Barrière Y. 2003.** Genetic variation QTL mapping of para-coumaric and ferulic acid contents in maize stover at silage harvest. *Maydica* **48**: 75-84.
- Foster JG, Clapham WM, Belesky DP, Labreveux M, Hall MH, Sanderson MA. 2006.** Influence of cultivation site on sesquiterpene lactone composition of forage chicory (*Cichorium intybus* L.). *Journal of agricultural and food chemistry* **54**(5): 1772-1778.
- Frary A, Gol D, Keles D, Okmen B, Pinar H, Sigva H, Yemenicioglu A, Doganlar S. 2010.** Salt tolerance in *Solanum pennellii*: antioxidant response and related QTL. *BMC Plant Biology* **10**(1): 58.
- Galland R, Randoux B, Vasseur J, Hilbert J-L. 2001.** A glutathione S-transférase cDNA identified by mRNA differential display is upregulated during somatic embryogenesis in *Cichorium*. *Biochimica et Biophysica Acta* **1522**: 212-216.

- Gonda TJ, Sheiness DK, Bishop JM. 1982.** Transcripts from the cellular homologs of retroviral oncogenes: distribution among chicken tissues. *Molecular and Cellular Biology* **2**(6): 617-624.
- Gonthier L, Bellec A, Blassiau C, Prat E, Helmstetter N, Rambaud C, Huss B, Hendriks T, Bergès H, Quillet M-C. 2010.** Construction and characterization of two BAC libraries representing a deep-coverage of the genome of chicory (*Cichorium intybus* L., Asteraceae). *BMC Research Notes* **submitted**.
- Green MA, Fry SC. 2005.** Vitamin C degradation in plant cells via enzymatic hydrolysis of 4-O-oxalyl-L-threonate. *Nature* **433**: 83-87.
- Griffith M, Huner NPA, Espelle KF, Kolattukudy PE. 1985.** Lipid polymers accumulate in the epidermis and mesophyll sheath cell walls during low temperature development of winter rye leaves. *Protoplasma* **125**: 53-57.
- Guedira M, Dubois-Tylski T, Vasseur J, Dubois J. 1989.** Embryogenèse somatique directe à partir de cellules d'anthères de *Cichorium* (Asteraceae). *Canadian Journal of Botany* **67**: 970-976.
- Guedira M, Dubois T, Dubois J, Vasseur J. 1990.** Influence du saccharose sur le rendement, la morphologie et le développement des embryons somatiques sur *Cichorium*. *Compte Rendus de l'Académie des Sciences de Paris* **310**(453-459).
- Guo J, Wang M-H. 2009.** Characterization of the phenylalanine ammonia-lyase gene (*SIPAL5*) from tomato (*Solanum lycopersicum* L.). *Molecular Biology Reports* **36**(6): 1579-1585.
- Hall TA. 1999.** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95-98.
- Halliwell B. 1978.** Superoxide-dependent formation of hydroxyl radicals in the presence of iron chelates. *FEBS Letters* **92**: 321-326.
- Han Y-F, Cao G-X, Gao X-J, Xia M. 2010.** Isolation and characterisation of the sesquiterpene lactones from *Lactuca sativa* L. var. *anagustata*. *Food Chemistry* **120**(4): 1083-1088.
- Hance P, Martin Y, Vasseur J, Hilbert J-L, Trotin F. 2007.** Quantification of chicory root bitterness by an ELISA for 11 β ,13-dihydroxylactucin. *Food Chemistry* **105**: 742-748.
- Hayashi T, Awata T. 2004.** Efficient method for analysis of QTL using F1 progenies in an outcrossing species. *Genetica* **122**(2): 173-183.
- Heimler D, Isolani L, Vignolini O, Tombelli S, A. R. 2007.** Polyphenol content and antioxidative activity in some species of freshly consumed salads. *Journal of Agricultural and Food Chemistry* **55**: 1724-1729.
- Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, Pollet B, Legrand M. 2004.** Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinic acid hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* **16**(6): 1446 - 1465.
- Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M. 2003.** Purification, cloning, and properties of an acyltransferase controlling shikimate and quinic acid ester intermediates in phenylpropanoid metabolism. *Journal of Biological Chemistry* **278**(1): 95 - 103.
- Hohlfeld M, Veit M, Strack D. 1996.** Hydroxycinnamoyltransferases involved in the accumulation of caffeic acid esters in gametophytes and sporophytes of *Equisetum arvense*. *Plant Physiology* **111**(4): 1153-1159.
- Holton TA, Cornish EC. 1995.** Genetics and Biochemistry of Anthocyanin Biosynthesis. *Plant Cell* **7**(7): 1071-1083.
- Huang D-W, Shen S-C, Wu JS-B. 2009.** Effects of caffeic acid and cinnamic acid on glucose uptake in insulin-resistant mouse hepatocytes. *Journal of Agricultural and Food Chemistry* **57**(17): 7687-7692.

- Hudec J, Burdova M, Kobida L, Komora L, Macho V, Kogan G, Turianica I, Kochanova R, Lozek O, Haban M, Chlebo P. 2007.** Antioxidant capacity changes and phenolic profile of *Echinacea purpurea*, nettle (*Urtica dioica* L.), and dandelion (*Taraxacum officinale*) after application of polyamine and phenolic biosynthesis regulators. *Journal of Agricultural and Food Chemistry* **55**(14): 5689-5696.
- Innocenti M, Gallori S, Giaccherini C, Ieri F, Vincieri FF, Mulinacci N. 2005.** Evaluation of the phenolic content in the aerial parts of different varieties of *Cichorium intybus* L. *Journal of agricultural and food chemistry* **53**: 6497-6502.
- Ishikawa Y, Kitamura M. 2000.** Bioflavonoid quercetin inhibits mitosis and apoptosis of glomerular cells *in vitro* and *in vivo*. *Biochemical and Biophysical Research Communications* **279**(2): 629-634.
- Jansen RC, Stam P. 1994.** High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **158**: 925-926.
- Jeuken M, van Wijk R, Peleman J, Lindhout P. 2001.** An integrated interspecific AFLP map of lettuce (*Lactuca*) based on two *L. sativa* × *L. saligna* F2 populations. *Theoretical and Applied Genetics* **103**(4): 638-647.
- Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, Tonelli C, Weisshaar B, Martin C. 2000.** Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. *The EMBO Journal* **19**(22): 6150-6161.
- Jin L, Xiao P, Lu Y, Shao Y, Shen Y, Bao J. 2009.** Quantitative Trait Loci for Brown Rice Color, Phenolics, Flavonoid Contents, and Antioxidant Capacity in Rice Grain. *Cereal Chemistry* **86**(6): 609-615.
- Johnson IT, Williamson G, Musk SRR. 1994** Anticarcinogenic factors in plant foods: a new class of nutrients? *Nutrition Research Reviews* **7**: 175-204.
- Joseph JA, Shukitt-Hale B, Denisova NA, Bielinski D, Martin A, McEwen JJ, Bickford PC. 1999.** Reversals of age-related declines in neuronal signal transduction, cognitive, and motor behavioral deficits with blueberry, spinach, or strawberry dietary supplementation. *The Journal of Neuroscience* **19**(18): 8114-8121.
- Ju YH, Allred CD, Allred KF, Karko KL, Doerge DR, Helferich WG. 2001.** Physiological concentrations of dietary genistein dose-dependently stimulate growth of estrogen-dependent human breast cancer (MCF-7) tumors implanted in athymic nude mice. *The Journal of Nutrition* **131**(11): 2957-2962.
- Kader F, Rovel B, Girardin M, Metche M. 1997** Mechanism of browning in fresh highbush blueberry fruit, partial purification and characterization of blueberry polyphenol oxidase. *Journal of the Science of Food and Agriculture* **73**: 513-516.
- Kandeler R, Ullrich WR. 2009.** Symbolism of plants: examples from European-Mediterranean culture presented with biology and history of art: NOVEMBER: Chicory. *Journal of Experimental Botany* **60**(14): 3973-3974.
- 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**(4): 1435-1446.
- Kiers A. 2000.** *Endive, Chicory and their wild relatives. A systematic and phylogenetic study of Cichorium (Asteraceae)*. Universiteit Leiden Leiden.
- Kim DK. 2001.** Antioxidative components from aerial parts of *Lactuca scariola* L. *Archives of Pharmacal Research* **24**: 427-430.
- Kirst M, Myburg AA, De Leon JPG, Kirst ME, Scott J, Sederoff R. 2004.** Coordinated genetic regulation of growth and lignin revealed by quantitative trait locus analysis of cDNA microarray data in an interspecific backcross of *Eucalyptus*. *Plant Physiology* **135**(4): 2368-2378.

- Kitts DD, Yuan YV, Wijewickreme AN, Thompson LU. 1999.** Antioxidant activity of the flaxseed lignin secoisolariciresinol diglycoside and its mammalian lignan metabolites enterodiol and enterolactone. *Molecular and Cellular Biochemistry* **201**(1): 91-100.
- Klempnauer K-H, Gonda TJ, Bishop JM. 1982.** Nucleotide sequence of the retroviral leukemia gene *v-myb* and its cellular progenitor *c-myb*: The architecture of a transduced oncogene. *Cell* **31**(2, Part 1): 453-463.
- Kliebenstein DJ. 2009.** Advancing genetic theory and application by metabolic quantitative trait loci analysis. *Plant Cell* **21**(6): 1637-1646.
- Kliebenstein DJ, Kroymann J, Brown P, Figuth A, Pedersen D, Gershenzon J, Mitchell-Olds T. 2001.** Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiology* **126**(2): 811-825.
- Koopman WJM. 2002.** *Zooming in on the lettuce genome: species relationships in Lactuca s.l., inferred from chromosomal and molecular characters.*, Nederland University of Wageningen.
- Kuntz S, Wenzel U, Daniel H. 1999.** Comparative analysis of the effects of flavonoids on proliferation, cytotoxicity, and apoptosis in human colon cancer cell lines. *European Journal of Nutrition* **38**(3): 133-142.
- Kyndt JA, Meyer TE, Cusanovich MA, Van Beeumen JJ. 2002.** Characterization of a bacterial tyrosine ammonia lyase, a biosynthetic enzyme for the photoactive yellow protein. *FEBS Letters* **512**(1-3): 240-244.
- Lander ES, Botstein D. 1989.** Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**: 185 - 199.
- Landry BS, Kesseli RV, Farrara B, Michelmore RW. 1987.** A Genetic Map of Lettuce (*Lactuca sativa* L.) With Restriction Fragment Length Polymorphism, Isozyme, Disease Resistance and Morphological Markers. *Genetics* **116**(2): 331-337.
- Lanteri S, Acquadro A, Comino C, Mauro R, Mauromicale G, Portis E. 2006.** A first linkage map of globe artichoke (*Cynara cardunculus* var. *scolymus* L.) based on AFLP, S-SAP, M-AFLP and microsatellite markers. *Theoretical and Applied Genetics* **112**(8): 1532-1542.
- Lee D, Douglas CJ. 1996.** Two divergent members of Tobacco 4-coumarate: CoEnzyme A ligase (*4CL*) gene family. *Plant Physiology* **112**: 193-205.
- Lee J, Scagel CF. 2010.** Chicoric acid levels in commercial basil (*Ocimum basilicum*) and *Echinacea purpurea* products. *Journal of Functional Foods* **2**: 77-84.
- 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**(1): 27.
- Lehmann EL. 1975.** *Nonparametrics*. New York: McGraw-Hill.
- Lepelley M, Cheminade G, Tremillon N, Simkin A, Caillet V, McCarthy J. 2007.** Chlorogenic acid synthesis in coffee: an analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in *C. canephora*. *Plant Science* **172**(5): 978 - 996.
- Lespinasse D, Grivet L, Troispoux V, Rodier-Goud M, Pinard F, Seguin M. 2000.** Identification of QTLs involved in the resistance to South American leaf blight (*Microcyclus ulei*) in the rubber tree. *Theoretical and Applied Genetics* **100**(6): 975-984.
- Li H, Liu H, Han Y, Wu X, Teng W, Liu G, Li W. 2010.** Identification of QTL underlying vitamin E contents in soybean seed among multiple environments. *Theoretical and Applied Genetics* **120**: 1405-1413.

- Li H, Ye G, Wang J. 2007.** A Modified Algorithm for the Improvement of Composite Interval Mapping. *Genetics* **175**(1): 361-374.
- Li JX, Yu SB, Xu CG, Tan YF, Gao YJ, Li XH, Zhang Q. 2000.** Analyzing quantitative trait loci for yield using a vegetatively replicated F2 population from a cross between the parents of an elite rice hybrid. *Theoretical and Applied Genetics* **101**: 248-254.
- Limami A, Roux L, J. L, Y. R. 1993.** Dynamics of nitrogen compounds in the chicory (*Cichorium intybus* L.) tuberised tap root during the growing season and cold storage period. *Journal of Plant Physiology* **141**(3): 263-268.
- Llorach R, Martínez-Sánchez A, Tomás-Barberán FA, Gil MI, Ferreres F. 2008.** Characterisation of polyphenols and antioxidant properties of five lettuce varieties and escarole. *Food Chemistry* **108**(3): 1028-1038.
- Llorach R, Tomàs-Barberà FA, Ferreres F. 2004.** Lettuce and chicory byproducts as a source of antioxidant phenolic extracts. *Journal of Agricultural and Food Chemistry* **52**: 5109-5116.
- Loewus FA. 1999.** Biosynthesis and metabolism of ascorbic acid in plants and of analogs of ascorbic acid in fungi. *Phytochemistry* **52**: 193-210.
- Lorieux M, Perrier X, Goffinet B, Gonzalez DdL, Lanaud C. 1995a.** Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations. *Theoretical and Applied Genetics* **90**: 73-80.
- Lorieux M, Perrier X, Goffinet B, Lanaud C, Gonzalez DdL. 1995b.** Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F2 populations. *Theoretical and Applied Genetics* **80**(81-89).
- Lu H, Romero-Severson J, Bernardo R. 2002.** Chromosomal regions associated with segregation distortion in maize. *Theoretical and Applied Genetics* **105**(4): 622-628.
- Luo J, Butelli E, Hill L, Parr A, Niggeweg R, Bailey P, Weisshaar B, Martin C. 2008.** AtMYB12 regulates caffeoyl quinic acid and flavonol synthesis in tomato: expression in fruit results in very high levels of both types of polyphenol. *The Plant Journal* **56**(2): 316-326.
- Lynch M, Walsh B. 1998.** *Genetics and Analysis of Quantitative Traits*. : Sunderland, MA, Sinauer Associates, Inc.
- Manach C, Scalbert A, Morand C, Rémésy C, Jiménez L. 2004.** Polyphenols: food sources and bioavailability. *American Journal of Clinical Nutrition* **79**(5): 727-747.
- Manach C, Williamson G, Morand C, Scalbert A, Remesy C. 2005.** Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies. *American Journal of Clinical Nutrition* **81**(1): 230-242.
- Mano Y, Omori F, Muraki M, Takamizo T. 2005.** QTL mapping of adventitious root formation under flooding conditions in tropical maize (*Zea mays* L.) seedlings. *Breeding Science* **55**(3): 343-347.
- Margawati ET, Fullard KJ, Raadsma HW, Subandriyo, Indriawati. 2009.** Identification of quantitative trait loci for growth traits on chromosome 18 backcross populations of Indonesian thin tail (ITT) sheep. *Journal of Biotechnology Research in Tropical Region* **2**(2).
- Marschner H 1991.** Root-induced changes in the availability of micronutrients in the rhizosphere. In: Y. Waisel, A. EshelU. Kafkafi eds. *In Plant Roots, the Hidden half*: Marcel Dekker, Inc., New York, 503-528.
- Martinez E, Duvnjak Z. 2006.** Enzymatic degradation of chlorogenic acid using a polyphenol oxidase preparation from the white-rot fungus *Trametes versicolor* ATCC 42530. *Process Biochemistry* **41**(8): 1835-1841.

- Massion P, Preiser J-C, Balligand J-L. 2002.** Les espèces réactives de l'azote : bénéfiques ou délétères? Reactive nitrogen species: deleterious or not? *Nutrition Clinique et Métabolisme* **16**(4): 248-252.
- Maston GA, Evans SK, Green MR. 2006.** Transcriptional regulatory elements in the human genome. *Annual Review of Genomics and Human Genetics* **7**(1): 29-59.
- Mazza G, Miniati E. 1993.** *Anthocyanins in fruits, vegetables, and grains* CRC Press, Boca Raton.
- McClure TT. 1960.** Chlorogenic acid accumulation and wound healing in sweet potato roots. *American Journal of Botany* **47**: 277-280.
- McCord JM, Day EDJ. 1978.** Superoxide-dependent production of hydroxyl radical catalyzed by iron - EDTA complex. *FEBS Letters* **86**(1): 139-142.
- Meyer D, Stasse-Wolthuis M. 2009.** The bifidogenic effect of inulin and oligofructose and its consequences for gut health. *European Journal of Clinical Nutrition* **63**(11): 1277-1289.
- Milala J, Grzelak K, Krol B, Juskiewicz J, Zdunczyk Z. 2009.** Composition and properties of chicory extracts rich in fructans and polyphenols. *Polish Journal of Food and Nutrition Sciences* **59**(1): 35-43.
- Milder IEJ, Arts ICW, Van de Putte B, Venema DP, Hollman PCH. 2005.** Lignan contents of Dutch plant foods: a database including lariciresinol, pinoresinol, secoisolariciresinol, and matairesinol. *British Journal of Nutrition* **93**: 393-402.
- Mizutani M, Ohta D. 1998.** Two Isoforms of NADPH:Cytochrome P450 Reductase in *Arabidopsis thaliana*. Gene Structure, Heterologous Expression in Insect Cells, and Differential Regulation. *Plant Physiology* **116**(1): 357-367.
- Moglia A, Comino C, Portis E, Acquadro A, De Vos R, Beekwilder J, Lanteri S. 2009.** Isolation and mapping of a C3'H gene (CYP98A49) from globe artichoke, and its expression upon UV-C stress. *Plant Cell Reports* **28**(6): 963-974.
- Mulinacci N, Innocenti M, Gallori S, Romani A, la Marca G, Vincieri FF. 2001.** Optimization of the chromatographic determination of polyphenols in the aerial parts of *Cichorium intybus* L. *Chromatographia* **54**: 455-461.
- Mutanda T, Wilhelmi B, Whiteley C. 2009.** Controlled Production of Fructose by an Exoinulinase from *Aspergillus Ficum*. *Applied Biochemistry and Biotechnology* **159**(1): 65-77.
- Nicolas JJ, Richard-Forget FC, Goupy PM, Amiot MJ, Aubert SY. 1994.** Enzymatic browning reactions in apple and apple products. *Critical Reviews in Food Science and Nutrition* **34**(2): 109-157.
- Niggeweg R, Michael AJ, Martin C. 2004.** Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nature biotechnology* **22**: 746 - 754.
- Nuissier G, Rezzonico B, Grignon-Dubois M. 2010.** Chicoric acid from *Syringodium filiforme*. *Food Chemistry* **120**(3): 783-788.
- Nüsslein B, Kurzmann M, Bauer R, Kreis W. 2000.** Enzymatic degradation of cichoric acid in *Echinacea purpura* preparations. *Journal of Natural Products* **63**: 1615-1618.
- Oetting WS, Lee HK, Flanders DJ. 1995.** Linkage analysis with multiplexed short tandem repeat polymorphism using infrared fluorescence and M13 tailed primers. *Genomics* **30**: 450-458.
- Ohnishi M, Matuo T, Tsuno T, Hosoda A, Nomura E, Taniguchi H, Sasaki H, Morishita H. 2004.** Antioxidant activity and hypoglycemic effect of ferulic acid in STZ-induced diabetic mice and KK-Ay mice. *BioFactors* **21**(1): 315-319.
- Olthof MR, Hollman PCH, Katan MB. 2001.** Chlorogenic acid and caffeic acid are absorbed in humans. *The journal of nutrition* **131**(1): 66-71.

- Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. 1989.** Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Genetics* **86**: 2766-2770.
- Papetti A, Daglia M, Aceti C, Sordelli B, Spini V, Carazzone C, Gazzani G. 2008.** Hydroxycinnamic acid derivatives occurring in *Cichorium endivia* vegetables. *Journal of Pharmaceutical and Biomedical Analysis* **48**(2): 472-476.
- Paxton JD. 1981.** Phytoalexins - A Working Redefinition. *Journal of Phytopathology* **101**(2): 106-109.
- Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H. 1987.** The regulatory c1 locus of *Zea mays* encodes a protein with homology to *myb* proto-oncogene products and with structural similarities to transcriptional activators. *The EMBO Journal* **6**(12): 3553-3558.
- Pflieger S, Lefebvre V, Causse M. 2001.** The candidate gene approach in plant genetics: a review. *Molecular Breeding* **7**(4): 275-291.
- Portis E, Mauromicale G, Mauro R, Acquadro A, Scaglione D, Lanteri S. 2009.** Construction of a reference molecular linkage map of globe artichoke (*Cynara cardunculus* var. *scolymus*). *Theoretical and Applied Genetics* **120**(1): 59-70.
- Price SF, Breen PJ, Valladao M, Watson BT. 1995.** Cluster Sun Exposure and Quercetin in Pinot noir Grapes and Wine. *American Journal of Enology and Viticulture* **46**(2): 187-194.
- 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*. *Theoretical and Applied Genetics* **91**(8): 1195-1202.
- R Development Core Team 2009.** R: A Language and Environment for Statistical Computing. In. Vienna, Austria: R Foundation for Statistical Computing.
- Radtke J, Linseisen J, Wolfram G. 1998.** Phenolic acid intake of adults in a Bavarian subgroup of the national food consumption survey. *Zeitschrift für Ernährungswiss* **37**(2): 190-197.
- Ramsay L, Macaulay M, Cardle L, Morgante M, degli Ivanissevich S, Maestri E, Powell W, Waugh R. 1999.** Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *The Plant Journal* **17**(4): 415-425.
- Rees SB, Harborne JB. 1985.** The role of sesquiterpene lactones and phenolics in the chemical defence of the chicory plant. *Phytochemistry* **24**: 2225-2231.
- Rick CM, Smith PG. 1953.** Novel variation in tomato species hybrids. *Am Nat* **87**: 359-375.
- Rieseberg LH, Archer MA, Wayne RK. 1999.** Transgressive segregation, adaptation and speciation. *Heredity* **83**(4): 363-372.
- Robards K, Prenzler PD, Tucker G, Swatsitang P, Glover W. 1999.** Phenolic compounds and their role in oxidative processes in fruits. *Food chemistry* **66**: 401-436.
- Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV. 2003.** Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Current Biology* **13**(17): 1512-1517.
- Romani A, Pinelli P, Galardi C, Sani G, Cimato A, Heimler D. 2002.** Polyphenols in greenhouse and open-air-grown lettuce. *Food chemistry* **79**: 337-342.
- Rommens CM, Craig MR, H. Y, Navarre DA, Ye J, Krucker M, Swords K. 2008.** Engineered native pathways for high kaempferol and caffeoylquinic acid production in potato. *Plant Biotechnology Journal* **6**(9): 870-886.
- Ronicke S, Hahn V, Vogler A, Friedt W. 2005.** Quantitative Trait Loci Analysis of Resistance to *Sclerotinia sclerotiorum* in Sunflower. *Phytopathology* **95**(7): 834-839.

- Rossetto M, Lante A, Vanzani P, Spettoli P, Scarpa M, Rigo A. 2005.** Red chicories as potent scavengers of highly reactive radicals: a study on their phenolic composition and peroxy radical trapping capacity and efficiency. *Journal of Agricultural and Food Chemistry* **53**: 8169-8175.
- Rousseaux C, Jones CM, Adams D, Chetelat R, Bennett A, Powell A. 2005.** QTL analysis of fruit antioxidants in tomato using *Lycopersicon pennellii* introgression lines. *Theoretical and Applied Genetics* **111**: 1396-1408.
- Rozen S, Skaletsky H. 1999.** Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics Methods and Protocols*: 365 - 386.
- Russo P, Tedesco I, Russo M, Russo GL, Venezia A, Cicala C. 2001.** Effects of de-alcoholated red wine and its phenolic fractions on platelet aggregation. *Nutrition, Metabolism and Cardiovascular Diseases* **11**(1): 25-29
- Scalbert A, Johnson IT, Saltmarsh M. 2005.** Polyphenols: antioxidants and beyond. *American Journal of Clinical Nutrition* **81**(1): 215-217.
- Scalbert A, Williamson G. 2000.** Dietary intake and bioavailability of polyphenols. *The Journal of Nutrition* **130**: 2073S-2085S.
- Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, Werck-Reichhart D. 2001.** CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *The Journal of Biological Chemistry* **276**: 36566 - 36574.
- Schütz K, Kammerer DR, Carle R, Schieber A. 2005.** Characterization of phenolic acids and flavonoids in dandelion (*Taraxacum officinale* WEB. ex WIGG.) root and herb by high-performance liquid chromatography/electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **19**(2): 179-186.
- Seto M, Miyase T, Umehara K, Ueno A, Hirano Y, Otani N. 1988.** Sesquiterpene lactones from *Cichorium endivia* L. and *C. intybus* L. and cytotoxic activity. *Chemical and Pharmaceutical Bulletin* **36**(7): 2423-2429.
- Shahidi F, Chandrasekara A. 2010.** Hydroxycinnamates and their in vitro and in vivo antioxidant activities *Phytochemistry Reviews* **9**(1): 147-170.
- Sibov ST, De Souza CLJ, Garcia AAF, F. GA, Silva AR, Mangolin CA, Benchimol LL, De Souza AP. 2003.** Molecular mapping in tropical maize (*Zea mays* L.) using microsatellite markers. 1. Map construction and localization of loci showing distorted segregation. *Hereditas* **139**(2): 96-106.
- Simon M, Loudet O, Durand S, Berard A, Brunel D, Sennesal F-X, Durand-Tardif M, Pelletier G, Camilleri C. 2008.** Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. *Genetics* **178**(4): 2253-2264.
- Singh M, Arseneault M, Sanderson T, Murthy V, Ramassamy C. 2008.** Challenges for research on polyphenols from foods in Alzheimer's disease: bioavailability, metabolism, and cellular and molecular mechanisms. *Journal of Agricultural and Food Chemistry* **56**(13): 4855-4873.
- Solecka D. 1997.** Role of phenylpropanoid compounds in plant responses to different stress factors. *Acta Physiologiae Plantarum* **19**(3): 257-268.
- Song X-L, Sun X-Z, Zhang T-Z. 2006.** Segregation distortion and its effect on genetic mapping in plants. *Chinese Journal of Agricultural Biotechnology* **3**(03): 163-169.
- Sosulski F, Krygier K, Hogge L. 1982.** Free, esterified, and insoluble-bound phenolic acids. 3. Composition of phenolic acids in cereal and potato flours. *Journal of Agricultural and Food Chemistry* **30**(2): 337-340.
- Sourdille P, Singh S, Cadalen T, Brown-Guedira GL, Gay G, Qi L, Gill BS, Dufour P, Murigneux A, Bernard M. 2004.** Microsatellite-based deletion bin system for the

REFERENCES

- establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). *Functional & Integrative Genomics* **4**(1): 12-25.
- Stapleton AE. 1992.** Ultraviolet radiation and plants: burning questions. *Plant Cell* **4**(11): 1353-1358.
- Stevens R, Buret M, Duffe P, Garchery C, Baldet P, Rothan C, Causse M. 2007.** Candidate genes and quantitative trait loci affecting fruit ascorbic acid content in three tomato populations. *Plant Physiology* **143**(4): 1943-1953.
- Syed N, Sørensen A, Antonise R, van de Wiel C, van der Linden C, van 't Westende W, Hooftman D, den Nijs H, Flavell A. 2006.** A detailed linkage map of lettuce based on SSAP, AFLP and NBS markers. *Theoretical and Applied Genetics* **112**(3): 517-527.
- Szalma SJ, Snook ME, Bushman BS, Houchins KE, McMullen MID. 2002.** Duplicate loci as QTL: the role of chalcone synthase loci in flavone and phenylpropanoid biosynthesis in maize. *Crop Science* **42**(5): 1679-1687.
- Sztejnberg A, Azaizia H, Chet I. 1983.** The Possible Role of Phenolic Compounds in Resistance of Horticultural Crops to *Dematophora necatrix* Hartig. *Journal of Phytopathology* **107**(4): 318-326.
- Tabet N, Mantle D, Walker Z, Orrell M. 2002.** Endogenous antioxidant activities in relation to concurrent vitamins A, C, and E intake in dementia. *International Psychogeriatrics* **14**(1): 7-15.
- Tabor HK, Risch NJ, Myers RM. 2002.** Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics* **3**(5): 391-397.
- Takayama S, Isogai A. 2005.** Self-incompatibility in plants. *Annual Review of Plant Biology* **56**(1): 467-489.
- Talalay P, De Long MJ, Prochaska HJ. 1988.** Identification of a common chemical signal regulating the induction of enzymes that protect against chemical carcinogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **85**(21): 8261-8265.
- Tamagnone L, Merida A, Parr A, Mackay S, Culianez-Macia FA, Roberts K, Martin C. 1998.** The AmMYB308 and AmMYB330 transcription factors from *Antirrhinum* regulate phenylpropanoid and lignin biosynthesis in transgenic tobacco. *Plant Cell* **10**(2): 135-154.
- Tautz D. 1989.** Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**(16): 6463-6471.
- Tozlu I, Guy CL, Moore GA. 1999a** QTL analyses of Na⁺ and Cl⁻ accumulation related traits in an intergeneric BC1 progeny of *Citrus* and *Poncirus* under saline and nonsaline environments *Genome* **42**: 692-705
- Tozlu I, Guy CL, Moore GA. 1999b.** QTL analyses of morphological traits in an intergeneric BC1 progeny of *Citrus* and *Poncirus* under saline and non-saline environments. *Genome* **42**: 1020-1029
- 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.). *Theoretical and Applied Genetics* **115**(6): 735-746.
- Turrens JF. 2003.** Mitochondrial formation of reactive oxygen species. *The Journal of Physiology* **552**(2): 335-344.
- Ugartondo V, Mitjans M, Vinardell MP. 2008.** Comparative antioxidant and cytotoxic effects of lignins from different sources. *Bioresource Technology* **99**(14): 6683-6687.

- Van Acker SABE, Van Den Berg D-J, Tromp MNJL, Griffioen DH, Van Bennekom WP, Van Der Vijgh WJF, Bast A. 1996.** Structural aspects of antioxidant activity of flavonoids. *Free Radical Biology and Medicine* **20**(3): 331-342.
- Van Dam RM, Feskens EJ. 2002** Coffee consumption and risk of type 2 diabetes mellitus. *Lancet* **360**: 1477-1478.
- Van Ooijen J, Voorrips RE 2001.** JoinMap(R) 3.0, Software for the calculation of genetic linkage maps. In. *Plant Research International*. Wageningen, The Netherlands
- Van Ooijen JW. 1999.** LOD significance thresholds for QTL analyses in experimental populations of diploid species. *Heredity* **83**: 613-624.
- Van Ooijen JW, Boer MP, Jansen RC, Maliepaard C 2004.** MapQTL(R) 5.0 Software for the mapping of quantitative trait loci in experimental populations. In K. BV. *Wageningen, The Netherlands*.
- Van Stallen N, Vandenbussche B, Londers E, Noten V, De Proft M. 2005.** QTL analysis of production and taste characteristics of Chicory (*Cichorium intybus* var. *foliosum*). *Plant Breeding* **124**(1): 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 Breeding* **122**: 521-525.
- Vitrac X, Monti J-P, Vercauteren J, Deffieux G, Mérillon J-M. 2002.** Direct liquid chromatographic analysis of resveratrol derivatives and flavanones in wines with absorbance and fluorescence detection. *Analytica Chimica Acta* **458**(1): 103-110.
- Vos P, Hogers R, Bleeker M, Reijmans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M. 1995.** AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407 - 4414.
- Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lübberstedt T, Xia XC, Stam P, Zabeau M, Kuiper M. 1999.** Two high-density AFLP® linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. *Theoretical and Applied Genetics* **99**(6): 921-935.
- Vuylstekker C, Leleu O, Rambour S. 1997.** Influence of BAP and NAA on the expression of nitrate reductase in excised chicory roots. *Journal of Experimental Botany* **48**(310): 1079-1085.
- Wang S, Basten CJ, Zeng Z-B 2005b.** Windows QTL Cartographer 2.5. In. Department of statistics, North Carolina State University, Raleigh, NC.
- Watts KT, Mijts BN, Lee PC, Manning AJ, Schmidt-Dannert C. 2006.** Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family. *Chemistry & Biology* **13**(12): 1317-1326.
- Weisz GM, Kammerer DR, Carle R. 2009.** Identification and quantification of phenolic compounds from sunflower (*Helianthus annuus* L.) kernels and shells by HPLC-DAD/ESI-MSⁿ. *Food chemistry* **115** 758-765.
- Werij J, Kloosterman B, Celis-Gamboa C, de Vos C, America T, Visser R, Bachem C. 2007.** Unravelling enzymatic discoloration in potato through a combined approach of candidate genes, QTL, and expression analysis. *Theoretical and Applied Genetics* **115**(2): 245-252.
- Whetten R, Sederoff R. 1995.** Lignin Biosynthesis. *The Plant Cell* **7**: 1001-1013.
- Wijeratne SSK, Abou-Zaid MM, Shahidi F. 2005.** Antioxidant polyphenols in almond and its coproducts. *Journal of Agricultural and Food Chemistry* **54**(2): 312-318.
- Wojtaszek P, Stobiecki M, Gulewicz K. 1993.** Role of nitrogen and plant growth regulators in the exudation and accumulation of isoflavonoids by roots of intact white lupin plants. *Journal of Plant Physiology* **142**: 689-694.

- Wye C, Rouppe van der Voort J, Peleman J 2000.** Reverse QTL mapping (RQM), an efficient approach for high resolution mapping of QTL. In. *7th Conference on Plant and Animal Genome*. San Diego, California, USA.
- Yamaki S. 1984.** Isolation of Vacuoles from Immature Apple Fruit Flesh and Compartmentation of Sugars, Organic Acids, Phenolic Compounds and Amino Acids. *Plant Cell Physiology* **25**(1): 151-166.
- Yamakoshi J, Kataoka S, Koga T, Ariga T. 1999.** Proanthocyanidin-rich extract from grape seeds attenuates the development of aortic atherosclerosis in cholesterol-fed rabbits. *Atherosclerosis* **142**: 139-149.
- Yang CS, Landau JM, Huang M-T, Newmark HL. 2001.** Inhibition of carcinogenesis by dietary polyphenolic compounds. *Annual Review of Nutrition* **21**(1): 381-406.
- Zang L, Cosma G, Gardner H, Castranova V, Vallyathan V. 2003.** Effect of chlorogenic acid on hydroxyl radical. *Molecular and Cellular Biochemistry* **247**(1-2): 205 - 210.
- Zeng Z-B. 1994.** Precision mapping of quantitative trait loci. *Genetics* **136**: 1129-1138.
- Zhang Z-S, Li D, Wang L-J, Ozkan N, Chen XD, Mao Z-H, Yang H-Z. 2007.** Optimization of ethanol-water extraction of lignans from flaxseed. *Separation and Purification Technology* **57**(1): 17-24.
- Zhong R, Lii WHM, Negrel J, Ye Z-H. 1998.** Dual methylation pathways in lignin biosynthesis. *Plant Cell* **10**(12): 2033-2046.
- Zhou WC, Kolb FL, Bai GH, Domier LL, Boze LK, Smith NJ. 2003.** Validation of a major QTL for scab resistance with SSR markers and use of marker-assisted selection in wheat. *Plant Breeding* **122**: 40-46.
- Zhu QY, Huang Y, Tsang D, Chen Z-Y. 1999.** Regeneration of alpha-tocopherol in human low-density lipoprotein by green tea catechin. *Journal of Agricultural and Food Chemistry* **47**(5): 2020-2025.
- Zou W, Zeng Z-B. 2008.** Statistical methods for mapping multiple QTL. *International Journal of Plant Genomics*.

ANNEXES

A 'fresh' protocol for the analysis of hydroxycinnamic acids in leaf tissue of chicory (*Cichorium intybus* L., Asteraceae)

Meriem Bahri, Philippe Hance, Sébastien Grec, Marie-Christine Quillet, Jean-Louis Hilbert, Theo Hendriks*.

Université Lille Nord de France, Lille1, UMR INRA 1281 "Stress Abiotiques et Différenciation des Végétaux Cultivés", ERT 1067, GIS GENOCHIC, Bâtiment SN2, 3ème étage, F-59650 Villeneuve d'Ascq Cedex, France.

*Author to whom correspondence should be addressed (phone: 0033320434021; fax: 0033320337244; e-mail theo.hendriks@univ-lille1.fr).

ABSTRACT

A 'fresh' protocol is presented for facile and reliable estimation of soluble hydroxycinnamate levels in *Cichorium intybus* L. leaf tissue in large scale experiments. Samples were standardized by punching 6 discs per leaf, and hydroxycinnamates were extracted by submerging the discs in 80% ethanol with 5% acetic acid for at least 48h in darkness at 4°C. Residual dry mass of the discs was used for a posteriori corrections of compound levels. Chlorophyll was eliminated by chloroform, and the aqueous phases were transferred to microplates, dried, and dissolved in 50% methanol for HPLC analysis and storage. A HPLC program of 8 min was developed for analysis of the extracts. Comparisons with extractions of liquid nitrogen powders indicated that the 'fresh' extraction method was reliable. No degradation of the major hydroxycinnamates - caftaric, chlorogenic, and chicoric acid – was observed, not even during maceration at ambient temperatures, nor after storage for 1 year.

Keywords: Chicory leaf, hydroxycinnamates, caftaric acid, chlorogenic acid, chicoric acid, extraction method, HPLC.

INTRODUCTION

Plants produce a wide array of secondary metabolites that play important roles in their interactions with their biotic and abiotic environments, e.g. by impairing attacks by pathogens and herbivores, by attracting pollinators, or by shielding UV radiation (1). Hydroxycinnamates or phenylpropanoids (C6-C3 compounds), i.e. ferulic, coumaric, caffeic, and sinapic acid, and their derivatives, are among the most widely distributed plant secondary metabolites, and are precursors for the synthesis of many other molecules such as flavonoids, tannins, and lignin. They are particularly abundant in cereals, legumes, oilseeds, fruits, vegetables, and beverages (2, 3), and their occurrence, as well as their antioxidant activities, have been studied in relation to their proposed health benefits (3, 4).

The biosynthesis of hydroxycinnamates from phenylalanine via the phenylpropanoid pathway, and its genetic control, has been well studied in species like *Arabidopsis*, tomato and potato (5-7). In contrast, despite their importance in terms of the number of species present in the wild and used by man (8), knowledge on the genetic control of this metabolic pathway in Asteraceae species is still limited, especially concerning the biosynthesis of hydroxycinnamates specific for this plant family, like caftaric and chicoric acid (e.g. (9), (10, 11)).

Chicory (*Cichorium intybus* L.) is an Asteraceae (Compositae) species belonging to the Cichoriinae, a subclade of the Cichorieae (Lactuceae) (12). Appreciated as a medical plant for centuries in several civilizations (13), chicory is nowadays cultivated for numerous food and feed applications (see 14, for references). Hydroxycinnamic acids and flavanoids have been studied in chicory, particularly in leaf chicories (15-20). To unravel the genetic control of the metabolism of these molecules in chicory, we plan to identify quantitative trait loci (QTL) for the level of hydroxycinnamates in leaf tissue. A genetic map for chicory containing over 400 molecular markers is available (14), and initial results indicated the existence of quantitative variation for the level of hydroxycinnamates and antioxidant activity in one of the progenies used for map construction. Though this concerns a progeny of root chicory, the results may be of interest in breeding programmes for all *C. intybus* cultigroups. Furthermore, the foliage of root chicories is often left in the field after the harvest of roots, whereas it may be a source of natural antioxidants, similarly as suggested for lettuce and leaf chicories (16).

Reliable methods of extraction and analysis of polyphenolic compounds in chicory have been published, but most of them are labour intensive, time consuming, and require relative large quantities of material (16, 17, 19). These methods are not well adapted when rapid sampling

of large numbers of plants in the field or in a greenhouse is required, for instance to prevent developmental (and/or environmental) differences between the first and last plants sampled. Here, we describe a ‘fresh’ protocol that allows facile and reliable estimation of soluble hydroxycinnamate levels in chicory leaf tissue, and that is particular amenable for large scale experiments, like those required for QTL analyses.

MATERIALS AND METHODS

Plant material. The studied plants were K59 and K28, two industrial chicory genotypes selected from the improved Hungarian landrace population ‘Koospol’ (Florimond-Desprez, Cappelle-en-Pévèle, France), and plants of the F1 progeny K59×K28 (14). The original plants had been cloned by cutting, and the clones were grown in an unheated glasshouse under natural light conditions. All the studied plants were in vegetative state.

Chemicals. All solvents used were of HPLC grade quality from VWR (West Chester, PA, USA). Authentic standards of caffeic, caftaric (*O*-caffeoyltartaric acid), and chlorogenic acid (5-*O*-caffeoylquinic acid) were purchased from Sigma-Aldrich (St-Louis, MO, USA). Chicoric acid (di-*O*-caffeoyltartaric acid) was supplied by Chromadex, Inc. (Santa Ana, CA, USA).

Extraction of phenolic acids. Per plant two leaves of comparable physiological age, i.e. at similar, but opposite, positions on the plant, were sampled. From each leaf three pairs of discs (diameter 0.9 cm) from the upper, middle, and lower part, respectively, were collected in a 1,7 mL reaction tube (MµlTI®, Sorenson Bioscience Inc, Salt Lake City, UT, USA), by using its lid as a card puncher. The discs of each pair were taken on the both sides of the midvein.

‘Fresh’ method. Directly after sampling, the discs were submerged in 500 µL of 80% ethanol and 5% acetic acid, and allowed to macerate for at least 48h at 4°C in darkness. Thereafter the discs were removed and dried to determine the residual dry weight, used for *a posteriori* corrections of quantitative data.

‘Conventional’ method. Directly after sampling, tubes containing the discs were plunged in liquid nitrogen, the discs crushed in the tube with a close-fitting pestle, and the obtained powder was macerated for 1 hour at 4°C in darkness with 500 µL of 80% ethanol and 5% acetic acid. After centrifugation (2 min, 11180 g, 4°C), the supernatant was transferred to a new tube.

For both methods, 100 μL water and 500 μL chloroform were added to each extract, and after vortexing and centrifugation (2 min, 11180 g, 4°C), 150 μL of the chlorophyll-free aqueous phase were transferred to a well of a 96-well microplate (0.2 mL, Thermo-Fast® 96, Low Profile, Thermo Scientific, UK), dried under vacuum, and dissolved in 100 μL 50% methanol. The microplates were sealed (Thermo scientific, UK) and stored at -20°C for future analyses.

HPLC analysis of hydroxycinnamates. HPLC analyses were accomplished on a Prominence Shimadzu System (Palaiseau, France) equipped with a photodiode array UV-vis detector SPD-M20A (Shimadzu System). Standards were used to identify peaks by retention times and UV-vis spectra.

Initial analyses were carried out using a program of 36 minutes on a 150 x 4,6 mm, 5 μm LiChrospher RP18 column (Phenomenex, USA), with a 4 mm \times 3 mm guard-column. 6 μL of extract are injected and eluted with water (solvent A) and acetonitrile/2% acetic acid (solvent B). This method was inspired on previously described protocols (16, 21). The column temperature was 25°C, and the flow rate was maintained at 1 mL/min. A multi-step linear gradient was applied to elude the sample (**Table 1**). The column was equilibrated for 5 minutes between each injection.

A shorter multistep linear gradient was optimized to reduce the time of analysis (**Table 1**). The 8 minutes program was achieved on a 75 x 4 mm, 4 μm LiChrospher RP18 column (Phenomenex, USA), with the guard-column described previously. A 6 μL sample of each extract was analyzed. Solvents were A: water/1% acetic acid, B: acetonitrile/2% acetic acid. Flow rate, 1.8 mL/min ; oven temperature, 55°C. Gradient steps are summarized in Table 1. The column was equilibrated for 2.8 min between each injection.

LC/MS analysis. LC/MS analyses were realized by a Quattro II tandem quadrupole mass spectrometer (Waters Micromass, Manchester, UK) fitted with an electrospray ionization (ESI) source after separation by LC (Hewlett-Packard Model HP 1100, Agilent, Palo Alto, USA). Data were acquired with MassLynx 4.0 data software. Hydroxycinnamate standards were dissolved to a concentration of 0.1 mg/mL. The mass spectrometer was operated in negative mode, except for chlorogenic acid. The source was maintained at 120°C. ESI interface was operated with a capillary voltage of 3.5 kV. The cone voltage was set at 30 V. Nitrogen was used as nebulization and drying gas. ESI mass spectra were acquired by scanning MS over the m/z range 240–700.

MS/MS analysis. MS/MS analyses were performed by transmitting the appropriate precursor ion through MS (m/z 311) to the collision cell. The collision gas used was argon with a m/z 311 collision energy of 15 eV.

Test of variability induced by the manipulators. Four inexperienced testers without any relation to the experiment were instructed to pierce six discs from each of, in total, 24 chicory leaves at three defined places (bottom, middle and top of the leaves). The discs were macerated with 80% ethanol and 5% acetic acid for one week, removed and dried. Subsequent protocol steps were followed as described. The four results for each of the 24 leaves were statistically compared by ANOVA.

Statistical Analyses. Statistics were conducted with Systat 12 Version 12.00.08, SPSS Inc., Chicago, USA. Differences between samples were tested using ANOVA.

RESULTS AND DISCUSSION

Comparison between ‘fresh’ and a conventional extraction method, and identification of the major hydroxycinnamic acids

Freezing and crushing several grams of fresh leaves in liquid nitrogen is a classical step found in the literature for polyphenolic extraction, followed by many variations concerning solvent, time and manner of maceration. A critical step in this ‘conventional’ method is to prevent the tissue or powder to get damp by thawing before being submerged in the maceration solution, as this may cause the reactivation of liberated enzymes that may modify or destruct the compounds of interest, like polyphenols oxidase (22). In contrast, the extraction step in the ‘fresh’ protocol is based on combining a rapid fixation of tissue structure and proteins, as in preparations of tissues for histological analyses, and the liberation by diffusion of compounds soluble in the maceration solution, in our case hydroxycinnamates.

To compare the efficacy of the ‘fresh’ method with the ‘conventional’ extraction method, we analysed samples obtained by extracting equal amounts of leaf tissue by the two different methods. As the aim of our research is to study hundreds of plants for genetic analysis, we reduced the fresh material to 6 discs (~70 mg FW) punched from upper, middle and lower part of a leaf with the lid of a 1.5 mL reaction tube. The maceration solution was 80% ethanol and 5% acetic acid, chosen as the most efficient solvent among the panels commonly

proposed to extract polyphenols. Similarly, chloroform turned out to be a judicious solvent to eliminate chlorophylls and to preserve molecules of interest (data not shown). HPLC chromatograms recorded at 280 nm showed similar profiles for the two methods, suggesting that the ‘fresh’ method is as efficient as the more conventional one (**Figure 1**).

In the chromatograms, three main peaks were attributed to caftaric acid, chlorogenic acid and chicoric acid by comparing their retention time and UV spectra with that of standard compounds, and their identity was confirmed by LC/MS analysis. The caftaric acid molecular ion was detected at m/z 311 (MW 312), and MS/MS analysis showed the presence of two main fragment ions, at m/z 149 and 135, corresponding to caffeic acid and its decarboxylation product, respectively. The absence of the ion m/z 623 in MS/MS chromatograms, that would correspond to a dimer of caftaric acid, suggests that it is *cis*-caftaric acid (23). The molecular ion $[M+H]^+$ of chlorogenic acid was detected at m/z 355 (MW 354), and the ion $[M-H]^-$ of chicoric acid at m/z 473 (MW 474). Revealing chicoric acid as the most abundant compound, followed by caftaric and chlorogenic acid (**Figure 1**), these results agree with previous studies on other varieties of *C. intybus* L. (17, 19, 21) or *C. endivia* L. (16).

Quantitative differences between the ‘fresh’ method and the ‘conventional’ extraction method were investigated by comparing the levels of the most abundant compounds obtained from 6 leaves taken from 3 different plants (**Table 2**). Pair-wise comparisons of samples, i.e. between two similar leaves from the same plant (samples 1 and 2, 3 and 4, and 5 and 6), showed reproducible differences in the levels of caftaric, chlorogenic, and chicoric acid extracted by the two methods. This indicated that the ‘fresh’ extraction method is at least as reliable as the more conventional method.

In general the overall levels of hydroxycinnamates extracted with the ‘fresh’ method were higher than those in the conventional method (**Table 2**). One possible cause for these differences may be the time of maceration in the conventional method in comparison to the ‘fresh’ method, i.e. 1 h vs 48 h, respectively. Why this should affect specifically chicoric acid, which showed the most important differences in comparison to caftaric and chlorogenic acid (**Table 2**), is not clear. In this respect, it is interesting to note that in an assay referred to as a validated method to determine caftaric acid, chicoric acid, chlorogenic acid, and echinacoside content in plant materials of *Echinacea* species, samples are prepared by extracting 0,125 g powdered material by shaking (ultrasonic) in 25 mL 70% (v/v) ethanol for 15 min only (Method 106.000 of the Institute for Nutraceutical Advancement, (24, 25). An alternative explanation may be that levels of chicoric acid in the conventional method extracts are reduced by the activity of enzymes released from the tissue upon grinding and maceration. As

shown for *Echinacea purpurea* preparations (22), chicoric acid was not stable under conditions where oxidative degradation of caffeic acid derivatives was inhibited, and it was found that under these conditions an esterase hydrolyzing the ester bonds between tartaric acid and caffeic acid was still active. Our results (**Table 2**) did not allow discriminating between these two possibilities, and further research is required to resolve this difference between the two extraction methods. Nonetheless, extraction by the ‘fresh’ method seemed to prevent degradation of the three selected compounds, most likely as enzymes are precipitated within the tissue and are not released to the maceration solution.

In chromatograms of some samples extracted by the ‘fresh’ method sometimes a very small peak corresponding to the position of caffeic acid was observed, but generally in the material analysed so far caffeic acid levels remained below the level of detection and quantification. Caffeic acid was identified by Rees and Harbone (1985) in leaves of root chicory ‘Magdebourg’, and it was also detected in some salad chicories, like Chioggia, Grumolo, Castelfranco and Verona (18). However, in other cultigroups, caffeic acid was not detected in free form but only esterified with quinic or tartaric acid (16, 17, 19, 21, 23, 26). Characterization of the other minor peaks present in the chromatograms is in progress. In comparison with previous investigations in chicory varieties, the presence of gallic, protocatechuic acid and other caffeoyl quinic and tartaric derivatives can be expected. Flavonoids identified in green chicory varieties are kaempferol, quercetine, luteoline, and apigenine, often bound to a glucose, or a glucuronide (16-21, 23, 26).

Optimization and characteristics of the ‘fresh’ protocol

To further speed up sampling in the ‘fresh’ protocol, discs were directly submerged in maceration solution without determination of fresh mass. Instead, the dry mass of the discs was determined after extraction. As investigated in a separate experiment on 6 different plants, dry mass before extraction represented on average $15,7\% \pm 1,5$ of disc’s fresh mass, whereas this was $12,8\% \pm 0,9$ for the residual dry mass after extraction of the discs. The constancy between fresh mass and the dry mass after extraction thus allows a correction of the hydroxycinnamate levels for the amount of tissue extracted.

To evaluate the influence of temperature on extraction, two temperatures of maceration were tested on 12 samples: 4°C and 22°C. Pair-wise comparison of HPLC profiles showed that they overlapped, and ANOVA tests on peak areas revealed no significant differences ($p > 0.05$)

between the two temperatures. This seemed to confirm a lack of enzymatic degradation of hydroxycinnamate compounds in extracts obtained by the 'fresh' protocol.

As for maceration time, 4 samples from the same leaf were macerated 24, 48, 72 or 96 hours. No qualitative differences were detected between the samples. At the quantitative level, 24 hours of maceration was slightly less efficient than the more prolonged periods of maceration. Taken together, these observations mean that for the 'fresh' method no strict temperature conditions are necessary, and that maceration time can be adapted according to the number of plants to be analysed and to the number of persons assuming the work.

Though maceration takes a relative long time, this is compensated for by its ease, and by the number of samples that can be treated at the same time. Furthermore, the number of disposables used is low, and there is no need for additional transfers and treatments that may require specialized (expensive) equipment when large numbers of sample are to be processed simultaneously. Considering the number of samples needed for quantitative genetic analyses, the storage in micro-well plates is convenient, and the micro-well plate format can also be used for further analyses of the samples, for example the evaluation of antioxidant capacity (e. g. DPPH•, ABTS•⁺ or FRAP assays, (27-29)).

To investigate the stability of samples conserved in 50% methanol, 36 extracts were re-analysed one year after storage at -20°C. Comparisons of HPLC profiles showed that quantitative differences between the two analyses were within the range of standard errors. This confirmed once more the absence of enzymatic degradation, and indicated that samples can safely be stored in 50% methanol for a long time at -20°C before analysis.

Test of variability induced by the manipulators

A possible source of variation in the 'fresh' protocol could be the sampling of the 6 discs per leaf by the experimenter. To test this, four inexperienced persons were instructed to sample 6 discs of 24 leaves, 2 at the top, 2 in the middle, and 2 at the base, and each pair separated by the midvein. Thereafter the discs were extracted as described, the extracts analysed by HPLC, and the results obtained for each of the 24 leaves sampled by the 4 different experimenters were compared. ANOVA test indicated there were no significant differences between the results obtained by the 4 testers ($p > 0.05$, underlined values, **Table 3**). As each leaf was sampled 4 times (i.e. 24 discs per leaf), and thus the positions at the top, middle or lower portion of the leaf at which the discs were taken differed for each of the experimenters, this indicated that the method allows reliable sampling by inexperienced, interchangeable persons.

Improvement of HPLC program

Having established the conditions for the extraction of hydroxycinnamic acids in the 'fresh' protocol, the HPLC method was adapted to reduce the time of analysis. The results obtained with the initial method indicated that the separation on C18 reversed-phase silica gave reliable results upon analysis of the chicory extracts. To achieve faster analysis, separation was carried out on a 75 mm long Lichrospher® RP18 column, rather than the one measuring 150 mm in the previous method, and elution solvent composition and proportion were adjusted (**Table 1**). This resulted in a reduction of the time of HPLC-DAD analysis to 8 min without loss of baseline separation of the components of interest (**Figure 2**).

Pure hydroxycinnamic acids and chicory extracts were used to validate the adapted HPLC analytical method. Sequential analyses of standards (chlorogenic, caffeic, and chicoric acid, cf. insert **Figure 2**), diluted series of chicory leaf extract, and chicory leaf extract spiked with standards, each followed by a sample of extraction solution, were repeated 3 times per day on 6 different days. After the comparison of the UV spectra, no endogenous interference at the retention time of the hydroxycinnamic acids was detected in the blanks following each analysis from sample extract. Overall, no interfering peaks were observed from the matrix extract and peak purity was higher than 90% for any eluted compound. Resolution was higher than 1.5 for all peaks. Finally, robustness of the method was tested by successively decreasing the parameters temperature, flow rate, and acetic acid concentration by 10% and 20%. None of these adjustments resulted in significant differences ($p < 0.05$), and similarity for quantitative values was higher than 95%.

To quantify hydroxycinnamic acids levels in extracts, an external standard method based on peak areas of the eluted compounds was used. Working range concentrations for each compound was estimated from dilution series of a sample extract, and chlorogenic acid was used to construct a five point calibration curve (0.5-150 µg/mL) applying a least-squares regression analysis. A good linear relationship was obtained with a correlation coefficient of 0.997. The LOD (limit of detection: concentration of compound yielding a signal/noise ratio of 3:1) and the LOQ (limit of quantification: concentration of compound yielding a signal/noise ratio of 10:1) obtained, 0.07 µg/mL and 0.24 µg/mL respectively, were suitable for quantitative analyses of hydroxycinnamic acid levels in chicory extracts. Repeatability and reproducibility for retention times and of quantifications were higher than 97%. Accuracy was assessed by analyzing chicory extract before and after spiking with three different

concentrations of standards (0.06, 0.12 and 0.25 µg/mL), and quantities recovered ranged between 96 % and 108%.

The ‘fresh’ protocol in QTL analyses

As indicated in the introduction, the ‘fresh’ protocol was developed to study the genetic control of hydroxycinnate metabolism in chicory leaf tissue. Preliminary results obtained after sampling near 1800 leaves, followed by quantitative analyses using the new HPLC protocol, indicated average concentrations of 246 ± 120 mg/kg FW for caftaric acid, 130 ± 92 mg/kg FW for chlorogenic acid, and 370 ± 202 mg/kg FW for chicoric acid. However, genotype-dependent quantitative differences were found for each of the three compounds, as well as for the total amount of hydroxycinnamates (cf. **Table 2**). This indicated that further analysis by the ‘fresh’ protocol of this progeny, previously used for the construction of a molecular genetic map for chicory (14), will allow identification of QTL for the level of these compounds in chicory leaf tissue. This knowledge will eventually enable breeders to use marker assisted selection to obtain high antioxidant vegetables, or transformed products, that respond to consumer’s expectations.

A further advantage of the ‘fresh’ protocol in QTL analyses is that it provides the possibility to express levels of the hydroxycinnamate compounds per area of leaf tissue and per residual dry mass. As the discs represent the same leaf tissue area (*i.e.* 0.636 cm^2 per disc), differences in discs’ dry mass may correspond to differences in the numbers and/or sizes of the cells in the different tissues present in the sample, *i.e.* upper and lower epidermis, parenchyma, and mesophyll. We will use both estimates of levels of the compounds, *i.e.* amounts expressed per area leaf tissue or per tissue mass, as well as residual dry mass of the discs as separate quantitative characters. This should allow the detection of QTL for the level of one or more of the compounds that are dependent or independent of the mass of leaf tissue extracted. Whereas both types of QTL may be of use for breeding purposes, the latter probably harbour genes that specifically participate in the regulation of the metabolism of the compounds of interest.

The estimation of soluble hydroxycinnamic acid levels in chicory leaf tissue by the protocol presented in this study proved to be facile and reliable, and allows a rapid screening of the phenolic acids’ content in leaves of different chicory cultigroups, and possibly also in other leafy vegetables or pharmaceutical herbs (*e.g.* (30, 31)). As the extraction part of the ‘fresh’

protocol doesn't require liquid nitrogen to preserve the material sampled, it is particularly suited for sampling large numbers of plants in a greenhouse or in the field. Considering the number of samples needed for quantitative genetic analyses, the storage in micro-well plates is convenient, and the micro-well plate format can also be used for further analyses of the samples, for example the evaluation of antiradical capacity (DPPH•, ABTS•⁺ or FRAP). Preliminary results indicated that the 'fresh' protocol is a useful tool in the identification of QTL for the level of hydroxycinnamates in chicory leaf tissue.

ABBREVIATIONS

ABTS, 2,2'-azinobis-(3-ethylbenzothiazoline-6-sulfonic acid); ANOVA, Analysis of variance; DAD, diode array detector; DPPH, 1,1-diphenyl-2-picryl-hydrazyl; ESI, electrospray ionization; FRAP, ferric reducing antioxidant power; FW, fresh weight; HPLC, high-performance liquid chromatography; MW, molecular weight; UV, ultra-violet.

ACKNOWLEDGEMENT

We like to thank Bruno Desprez for providing seeds of the Kospool population from which the plants analysed were derived, students of University of Lille for testing the protocol, Angélique Bourceaux, Cédric Glorieux, Nathalie Faure, and Eric Schmitt from the "Plateforme de culture en environnement contrôlé" for taking care of the plants. Helpful comments on the experiments by Francis Troitin, and critical reading of the manuscript by David Gagneul have also been greatly appreciated.

LITERATURE CITED

1. Dixon, R.; Paiva, N., Stress-induced phenylpropanoid metabolism. *Plant Cell* **1995**, 7, (7), 1085 - 1097.
2. Manach, C.; Scalbert, A.; Morand, C.; Rémésy, C.; Jiménez, L., Polyphenols: food sources and bioavailability. *American Journal of Clinical Nutrition* **2004**, 79, (5), 727-747.
3. Shahidi, F.; Chandrasekara, A., Hydroxycinnamates and their in vitro and in vivo antioxidant activities *Phytochemistry Reviews* **2010**, 9, (1), 147-170.
4. Cheng, J.-C.; Dai, F.; Zhou, B.; Yang, L.; Liu, Z.-L., Antioxidant activity of hydroxycinnamic acid derivatives in human low density lipoprotein: Mechanism and structure–activity relationship *Food Chemistry* **2007**, 104, (1), 132-139.
5. Hoffmann, L.; Besseau, S.; Geoffroy, P.; Ritzenthaler, C.; Meyer, D.; Lapierre, C.; Pollet, B.; Legrand, M., Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinic hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. *Plant Cell* **2004**, 16, (6), 1446 - 1465.
6. Niggeweg, R.; Michael, A. J.; Martin, C., Engineering plants with increased levels of the antioxidant chlorogenic acid. *Nature Biotechnology* **2004**, 22, 746 - 754.
7. Villegas, R. J.; Kojima, M., Purification and characterization of hydroxycinnamoyl D-glucose. Quinic hydroxycinnamoyl transferase in the root of sweet potato, *Ipomoea batatas* Lam. *Journal of Biological Chemistry* **1986**, 261, (19), 8729-9733.
8. Funk, V. A.; Bayer, R. J.; Keeley, S.; Chan, R.; Watson, L.; Gemeinholzer, B.; Schilling, E.; Panero, J. L.; Baldwin, B. G.; Garcia-Jacas, N.; Susanna, A., Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. *Biologiske Skrifter* **2005**, 55, 343-374.
9. Schütz, K.; Kammerer, D. R.; Carle, R.; Schieber, A., Characterization of phenolic acids and flavonoids in dandelion (*Taraxacum officinale* WEB. ex WIGG.) root and herb by high-performance liquid chromatography/electrospray ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **2005**, 19, (2), 179-186.
10. Romani, A.; Pinelli, P.; Galardi, C.; Sani, G.; Cimato, A.; Heimler, D., Polyphenols in greenhouse and open-air-grown lettuce. *Food Chemistry* **2002**, 79, 337-342.
11. Hudec, J.; Burdova, M.; Kobida, L.; Komora, L.; Macho, V.; Kogan, G.; Turianica, I.; Kochanova, R.; Lozek, O.; Haban, M.; Chlebo, P., Antioxidant capacity changes and phenolic profile of *Echinacea purpurea*, nettle (*Urtica dioica* L.), and dandelion (*Taraxacum officinale*) after application of polyamine and phenolic biosynthesis regulators. *Journal of Agricultural and Food Chemistry* **2007**, 55, (14), 5689-5696.

12. Kilian, N.; Gemeinholzer, B.; Lack, H. W., Tribe Cichorieae Lam. & DC. In: V.A. Funk, A. Susanna, T. Stuessy and R. Bayer, Editors, Systematics, Evolution and Biogeography of the Compositae **2009**, IAPT, Vienna, Austria.
13. Kandeler, R.; Ullrich, W. R., Symbolism of plants: examples from European-Mediterranean culture presented with biology and history of art: NOVEMBER: Chicory. *Journal of Experimental Botany* **2009**, 60, (14), 3973-3974.
14. Cadalen, T.; Mörchen, M.; Blassiau, C.; Clabaut, A.; Sheer, I.; Hilbert, J.-L.; Hendriks, T.; Quillet, M.-C., Development of SSR markers and construction of a consensus genetic map for chicory (*Cichorium intybus* L.). *Molecular Breeding* **2010**, 25, (4), 699-722.
15. Papetti, A.; Daglia, M.; Gazzani, G., Anti- and pro-oxidant activity of water soluble compounds in *Cichorium intybus* var. *silvestre* (Treviso red chicory). *Journal of Pharmaceutical and Biomedical Analysis* **2002**, 30, 939-945.
16. Llorach, R.; Tomàs-Barberà, F. A.; Ferreres, F., Lettuce and chicory byproducts as a source of antioxidant phenolic extracts. *Journal of Agricultural and Food Chemistry* **2004**, 52, 5109-5116.
17. Innocenti, M.; Gallori, S.; Giaccherini, C.; Ieri, F.; Vincieri, F. F.; Mulinacci, N., Evaluation of the phenolic content in the aerial parts of different varieties of *Cichorium intybus* L. *Journal of Agricultural and Food Chemistry* **2005**, 53, 6497-6502.
18. Rossetto, M.; Lante, A.; Vanzani, P.; Spettoli, P.; Scarpa, M.; Rigo, A., Red chicories as potent scavengers of highly reactive radicals: a study on their phenolic composition and peroxy radical trapping capacity and efficiency. *Journal of Agricultural and Food Chemistry* **2005**, 53, 8169-8175.
19. Heimler, D.; Isolani, L.; Vignolini, O.; Tombelli, S.; A., R., Polyphenol content and antioxidative activity in some species of freshly consumed salads. *Journal of Agricultural and Food Chemistry* **2007**, 55, 1724-1729.
20. Lavelli, V., Antioxidant activity of minimally processed red chicory (*Cichorium intybus* L.) evaluated in xanthine oxidase-, myeloperoxidase-, and diaphorase-catalyzed reactions. *Journal of Agricultural and Food Chemistry* **2008**, 56, (16), 7194-7200.
21. Mulinacci, N.; Innocenti, M.; Gallori, S.; Romani, A.; la Marca, G.; Vincieri, F. F., Optimization of the chromatographic determination of polyphenols in the aerial parts of *Cichorium intybus* L. *Chromatographia* **2001**, 54, 455-461.
22. Nüsslein, B.; Kurzmann, M.; Bauer, R.; Kreis, W., Enzymatic degradation of cichoric acid in *Echinacea purpurea* preparations. *Journal of Natural Products* **2000**, 63, 1615-1618.

23. Papetti, A.; Daglia, M.; Aceti, C.; Sordelli, B.; Spini, V.; Carazzone, C.; Gazzani, G., Hydroxycinnamic acid derivatives occurring in *Cichorium endivia* vegetables. *Journal of Pharmaceutical and Biomedical Analysis* **2008**, 48, (2), 472-476.
24. Perry, N. B.; Burgess, E. J.; L., G. V., *Echinacea* Standardization: Analytical Methods for Phenolic Compounds and Typical Levels in Medicinal Species. *Journal of Agricultural and Food Chemistry* **2001**, 49, (4), 1702-1706.
25. INA Phenolics in Echinacea by HPLC.
<http://www.nutraceuticalinstitute.com/methods/echinacea.html>.
26. Rees, S. B.; Harborne, J. B., The role of sesquiterpene lactones and phenolics in the chemical defence of the chicory plant. *Phytochemistry* **1985**, 24, 2225-2231.
27. Brand-Williams, W.; Cuvelier, M. E.; Berset, C., Use of a free radical method to evaluate antioxidant activity. *LWT - Food Science and Technology* **1995**, 28, (1), 25-30.
28. Re, R.; Pellegrini, N.; Proteggente, A.; Pannala, A.; Yang, M.; Rice-Evans, C., Antioxidant activity applying an improved ABTS radical cation decolorization assay. *Free Radical Biology and Medicine* **1999**, 26, (9-10), 1231-1237.
29. Benzie, I. F. F.; Strain, J. J., The ferric reducing ability of plasma (FRAP) as a measure of "antioxidant power": the FRAP assay. *Analytical Biochemistry* **1996**, 239, (70-76).
30. Kroll, U.; Cordes, C., Pharmaceutical prerequisites for a multi-target therapy. *Phytomedicine* **2006**, 13, (Supplement 1), 12-19.
31. Baur, S.; Klaiber, R. G.; Koblo, A.; Carle, R., Effect of different washing procedures on phenolic metabolism of shredded, packaged iceberg lettuce during storage. *Journal of Agricultural and Food Chemistry* **2004**, 52, (23), 7017-7025.

NOTE

Meriem Bahri was supported by a doctoral fellowship 50% from the Région Nord-Pas de Calais, France, and 50% from the GIS GENOCHIC.

TABLES

Table 1. Solvent gradients in the ‘long’ and ‘short’ HPLC analyses of hydroxycinnamtes in extracts of chicory leaf tissue. Solvent A consisted of water for the ‘long’ program and water/1% acetic acid for the ‘short’ program.

Long program (36 minutes)		Short program (8 minutes)	
Time (min)	Solvent A (%)	Time (min)	Solvent A (%)
0	100	0	94
2	92	1.4	90
5	87	4.2	84.3
15	85	4.4	81
16	83.5	4.8	0
18	81.5	5.2	94
22	80		
26	0		
29	0		
31	100		

Table 2. Levels of caftaric, chlorogenic and chicoric acid in extracts obtained by the ‘fresh’ method and a conventional method. Similar amounts of tissue sampled from 2 comparable leaves from 3 different chicory plants were extracted by the two methods, and quantities of the hydroxycinnamate compounds in the extracts were determined by HPLC (see text for details). Content are expressed in $\mu\text{g}/\text{cm}^2$ of fresh leaf tissue area.

sample	compound	‘fresh’		‘conventional’		fresh/conventional (%)
		$\mu\text{g}/\text{cm}^2$	%	$\mu\text{g}/\text{cm}^2$	%	
1	caf	7.1	26.2	7.6	33.4	92.6
	chlo	5.8	21.7	5.8	25.2	101.4
	chic	14.1	52.1	9.5	41.4	148.4
	sum	27	100	22.9	100	117.9
2	caf	6.8	25.1	7.6	33.2	89.9
	chlo	6.2	22.9	5.7	24.8	109.7
	chic	14.2	52.1	9.6	42	147.6
	sum	27.2	100	22.8	100	119
3	caf	13.8	19.1	16	28.7	86
	chlo	15.8	21.9	16.6	29.9	94.6
	chic	42.5	59	23	41.4	184.3
	sum	72	100	55.7	100	129.2
4	caf	13.9	18.9	16	28.6	87.2
	chlo	15.7	21.2	16.7	29.9	93.8
	chic	44.2	59.9	23.1	41.5	191.1
	sum	73.8	100	55.8	100	132.2
5	caf	11.3	17.9	7.8	18.7	145.5
	chlo	16.8	26.5	13.9	33.2	121.2
	chic	35.3	55.7	20.1	48.1	176.1
	sum	63.5	100	41.7	100	152.1
6	caf	11.4	17.9	8.8	20.3	128.8
	chlo	16.8	26.4	14.2	32.5	118.8
	chic	35.5	55.7	20.6	47.2	172.3
	sum	63.7	100	43.6	100	146.1

Table 3. ANOVA of caftaric, chlorogenic and chicoric acid contents estimations of 24 chicory leaves extracted by 4 different testers using the ‘fresh’ method.

Variables and source of variation	Degree of freedom	Mean square	F-ratio	<i>P</i> -value
Caftaric acid				
- Tester	3	0.005	2.67	<u>0.054</u>
- Leaf	23	0.096	49.63	<0.001
- Error	68	0.002		
Chlorogenic acid				
- Tester	3	0.004	0.266	<u>0.850</u>
- Leaf	23	0.376	23.11	<0.001
- Error	68	0.016		
Chicoric acid				
- Tester	3	0.005	0.869	<u>0.461</u>
- Leaf	23	0.118	21.75	<0.001
- Error	68	0.005		

FIGURES

Figure 1. Chromatographic profiles of chicory leaves at 280 nm. Dotted line and continuous line represent HPLC profiles obtained for the same fresh leaf tissue area with a conventional liquid nitrogen grinding and the ‘fresh’ method, respectively.

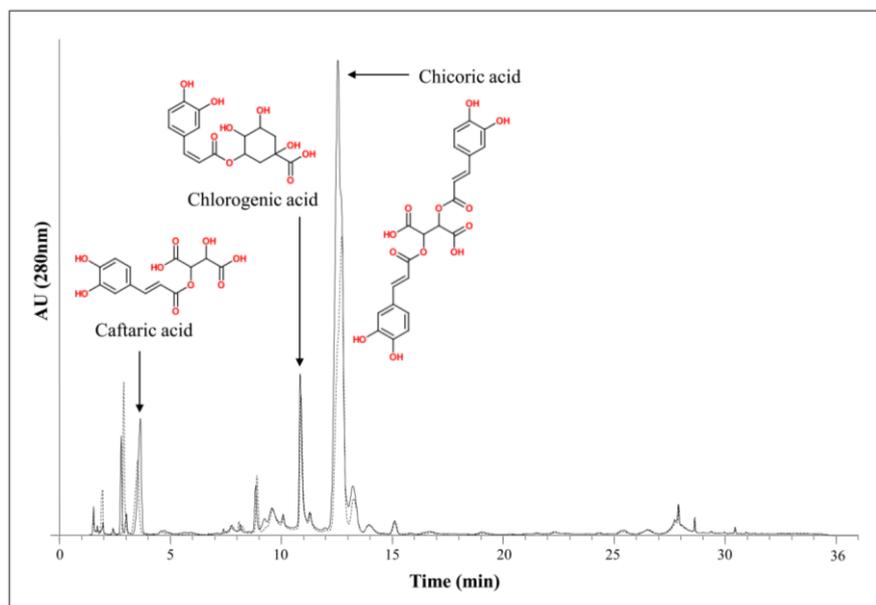
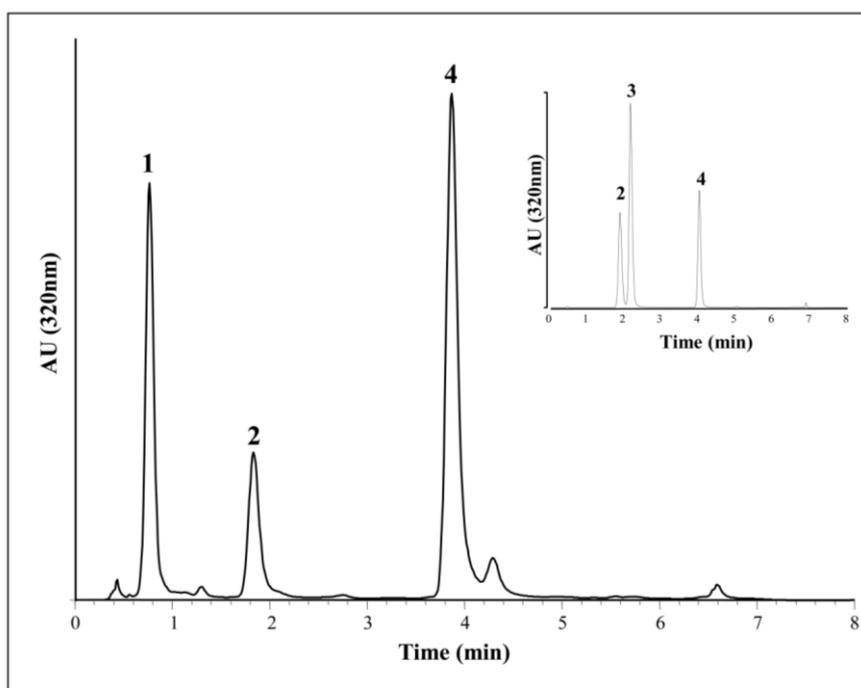
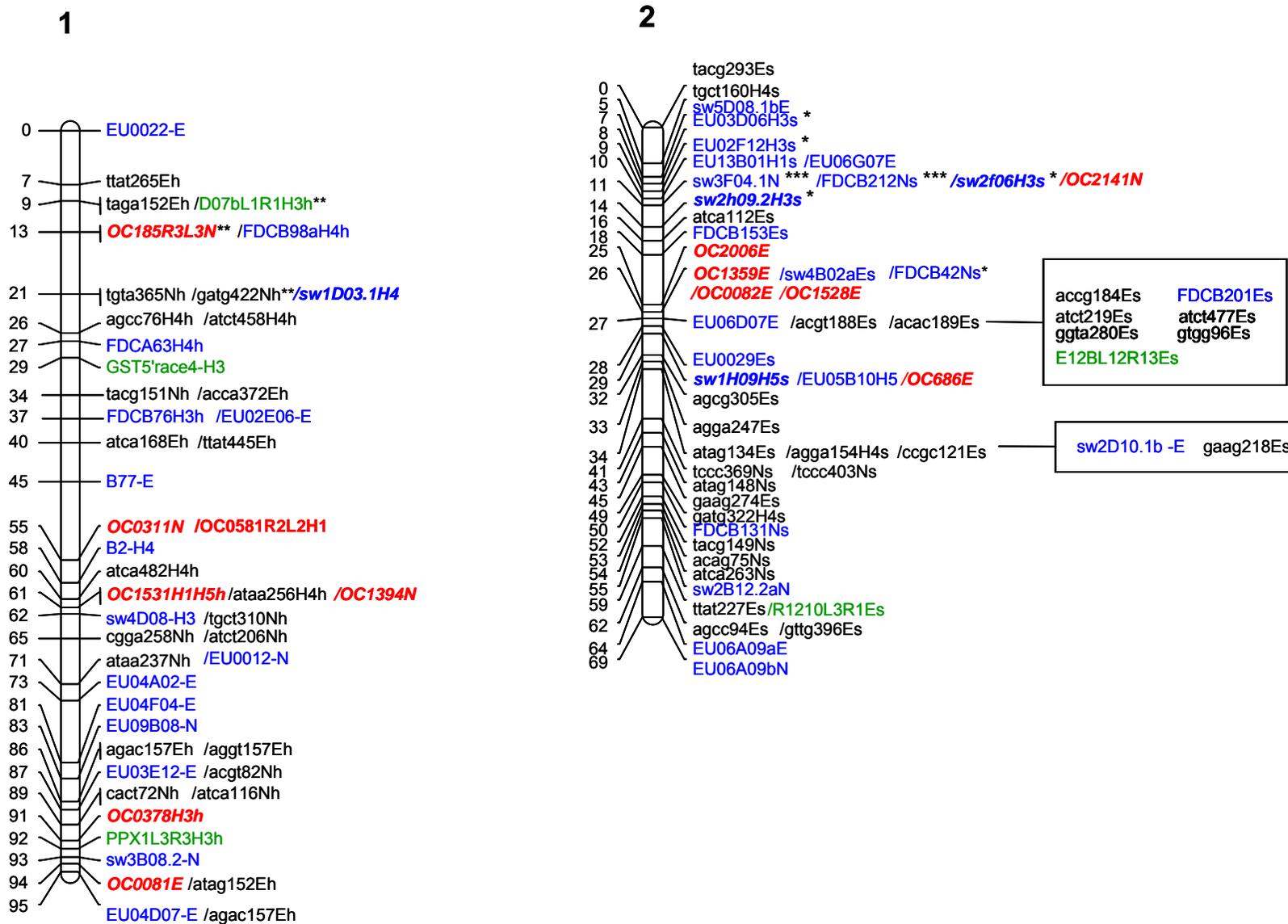
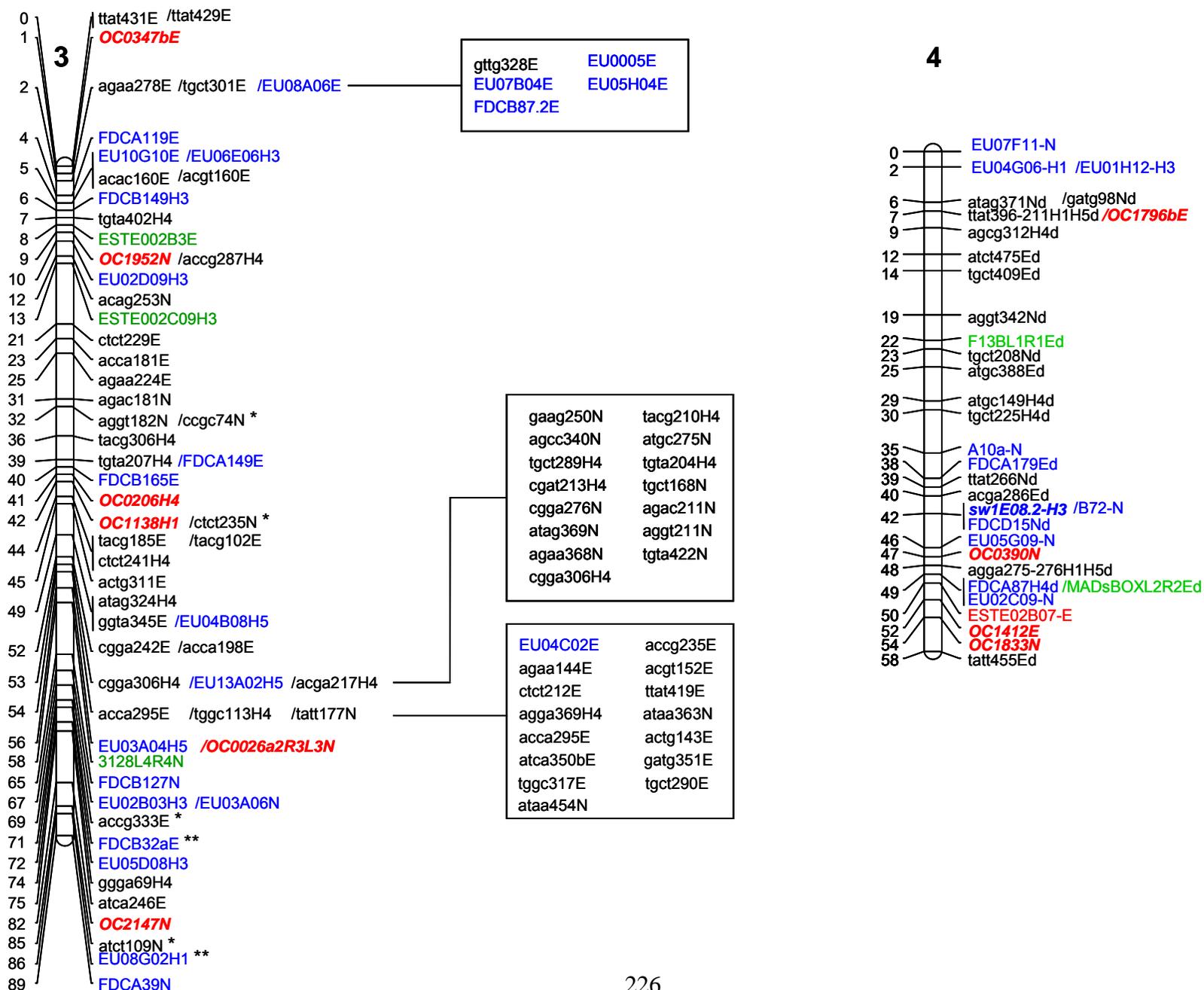


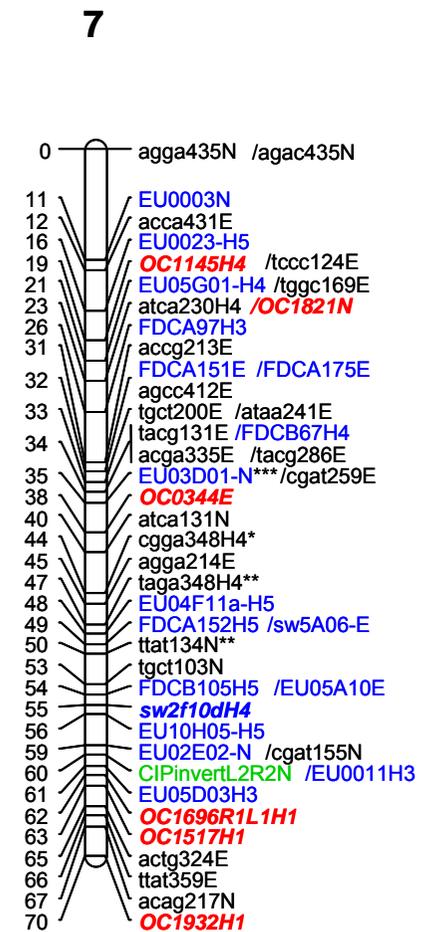
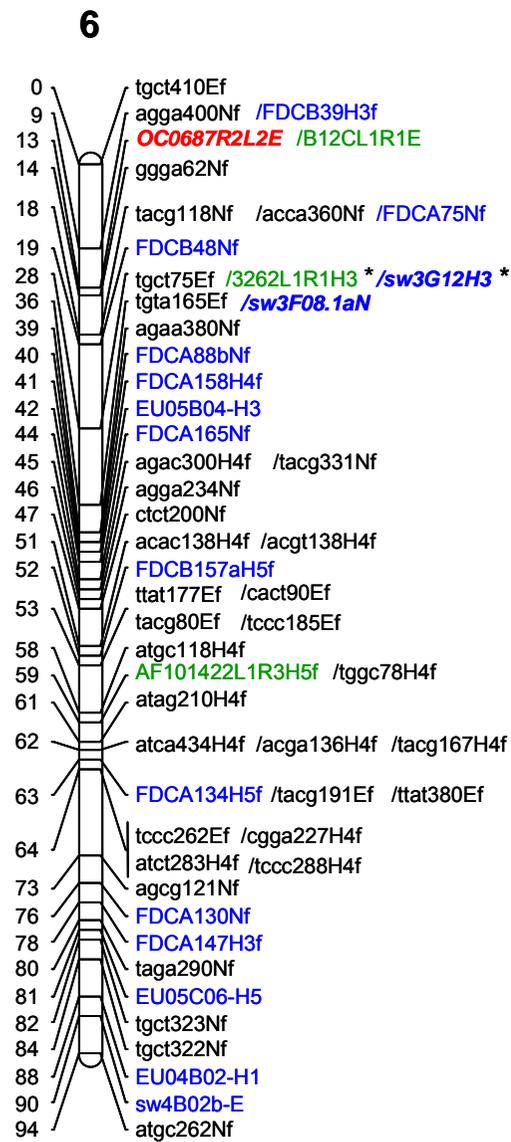
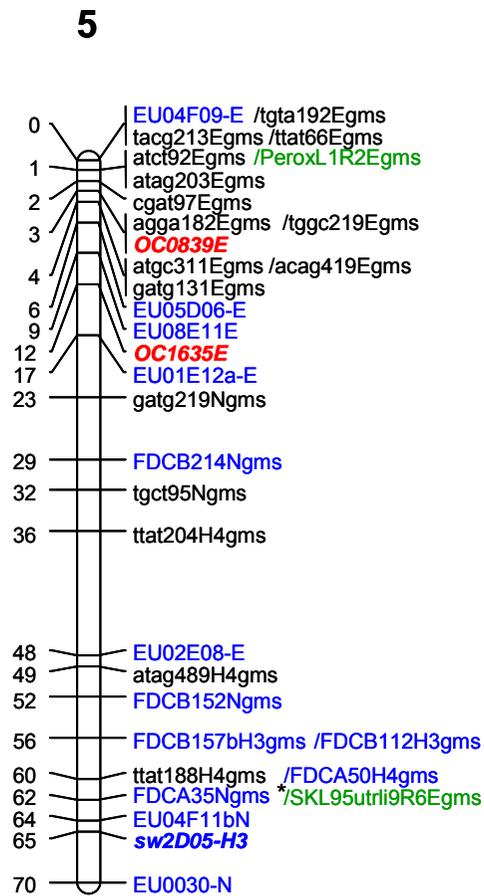
Figure 2. HPLC profile of chicory leaves with the ‘short program’, using the ‘fresh’ method. Peak identifications: (1) caftaric acid, (2), chlorogenic acid, (3) caffeic acid (insert), (4) chicoric acid. Chromatogram was recorded at 320 nm.



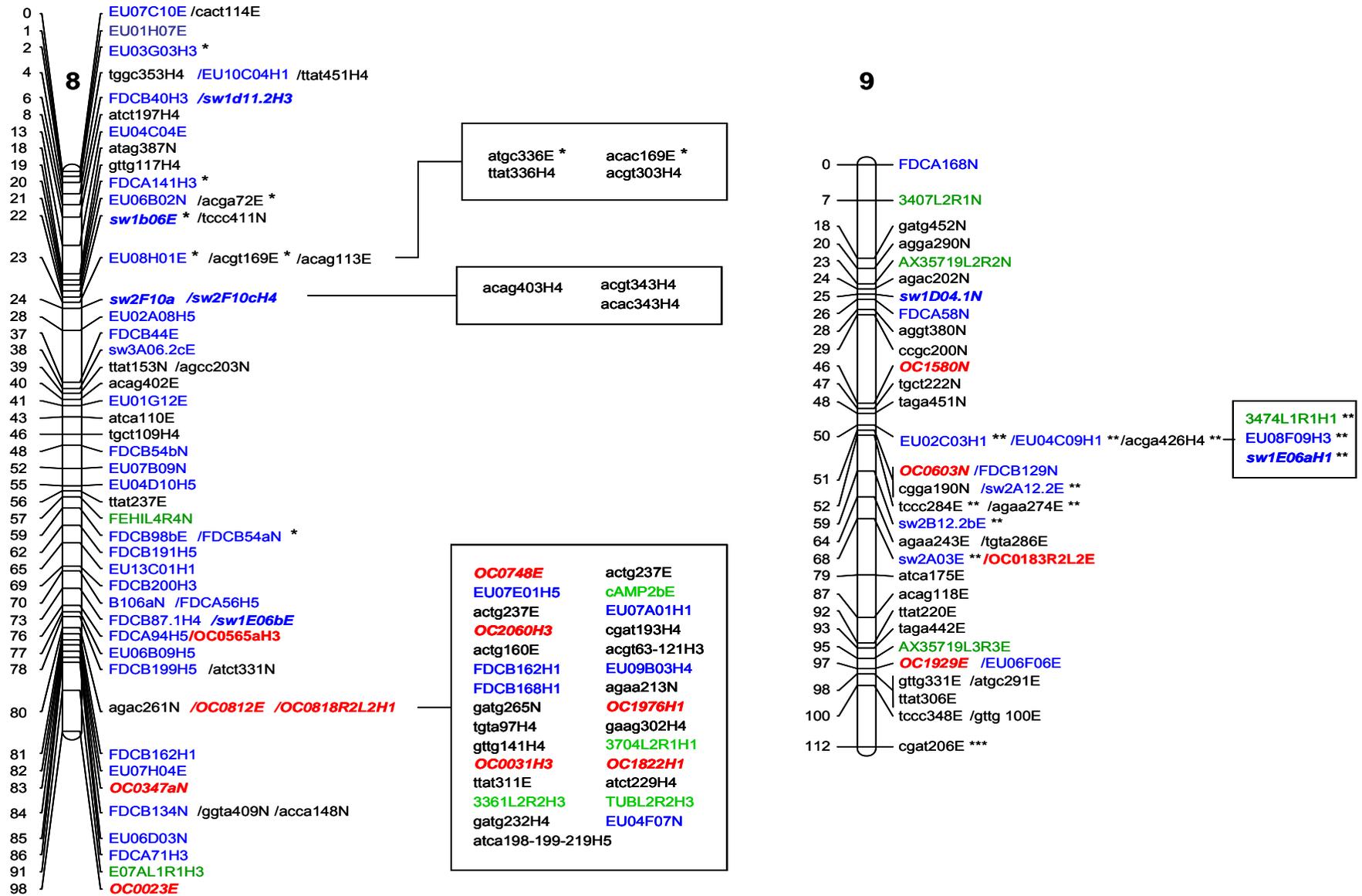


Annexe 2 : Carte de référence K28xK59





Annexe 2 : Carte de référence K28xK59



Annexe 3 : Amorces définies pour la détection du polymorphisme de 19 gènes candidats. Sens de lecture 5' vers 3'.

Nom du gène candidat	amorce gauche	amorce droite
<i>4CL</i>	AGCACTTTTGCTTTCGCATT	CAAATACCACTCGCTTCAATC
<i>AAE</i>	GGAAGAAATGTTGCCTCGAA	TTAACATGCATTAGCCCAACA
<i>ACO</i>	GGCCAAGCAGTGAAGAAATC	TGAAATACGGTGGTTCGTGA
<i>Anthoc</i>	TGGAGAAAGGATGGGAGATG	GAAATCCATTGTTGCCTGCT
<i>AOX</i>	ATGTGCCATAAACCTGGAG	CGAACACTTTTATGCCACCA
<i>APX</i>	GCAGTGTCACATGAAACT	TGCGTCAACAATTTCAATCC
<i>C3H</i>	GGAGGTGCTGAAGGAAAATG	AATGTCTGAAAATCGGTTCAA
<i>C4H</i>	GCGAAGATGATCCGTTGTTT	TCCAATTTCTTCGTGCTTCC
<i>CAD</i>	TGGGAAGCATGAAAGAGGAG	CATGGGGTAATTGGACATTC
<i>CCR</i>	CCAAAGGTTTCGCCTCTAAA	CCACCACAAAAGAAGGGTTG
<i>CHI</i>	AAGACGAAAACCTCCACCA	ACCGCTTGCCCTAATTTCTC
<i>CHS</i>	TTCGTGGCCCTAGTGACTCT	CCAATGGAGAAAACGCTTGT
<i>CSY</i>	TGTCGTATAAACCAAATCCAAGA	TTGCACCACCATGTAGAGGA
<i>ExpProt</i>	ACATGGAGGGCGTATGTTTC	AAAAAGAGTGAGCGTCATCCA
<i>HCT</i>	TACATGAGATCGGCCATTGA	AGTCTAAGCACACCGCAAGC
<i>MDA</i>	TGAAATTGAAGGACGGAAGG	CAGCAGATTTACGGGAATGG
<i>MDH</i>	TCTCCACCCTTCAAATGTCC	CGGTATCCATGTGGCTGATA
<i>MYB12</i>	GCCATTTACCAGGACGAACA	CTGGTGGCGACTTGATGTTA
<i>NAD</i>	ACTGAAGCCGAACGACTCAA	GGGGAGGAGGAAATGGAGTA

Annexe 4 : Résultats du test de Tukey pour les différentes interactions significatives détectés pour les caractères ACHIC, PCAFT, PCHLO et AAR. Chaque colonne indique la valeur phénotypique associée à la combinaison entre les allèles des marqueurs interactifs. Les lettres en dessous des valeurs phénotypique indiquent les combinaisons qui sont significativement différentes des autres ($p < 0,05$).

ACHIC	ee / ac	ee / ad	ee / bc	ee / bd	ef / ac	ef / bc	ef / bd	eg / ac	eg / ad	eg / bd	fg / ad	fg / bc	fg / bd
sw2h09.2 / EU05B10	271757,388 a, b	310955,643 a, b	296303,7 a, b	427953,800 a, b	347123,2 a	249426,47 a, b	306532,067 a, b	240009,45 a, b	257927,734 a, b	310984,933 a, b, c	259890,94 a, b	336176,14 a, b	242593,232 b

PCAFT	h- / ll	h- / lm	kk / ll	kk / lm
tgta204 / OC0812	0,21503077 a, b	0,22123684 b, c	0,23311765 c	0,20832 a

PCAFT	ee / ee	ee / ef	ee / eg	ee / fg	ef / ee	ef / ef	ef / eg	ef / fg	eg / ef	eg / eg	eg / fg	fg / ee	fg / ef	fg / eg	fg / fg
PPX1L3R3 / EU03G03	0,207 a, b, c	0,218 a, b, c	0,1975 b	0,21183333 a, b, c	0,23433333 c	0,21388889 a, b, c	0,23075 a, b, c	0,2102 a, b, c	0,2091 b	0,2252 c	0,22916667 c	0,193 b	0,2172 a, b, c	0,23942857 c	0,216 a, b, c

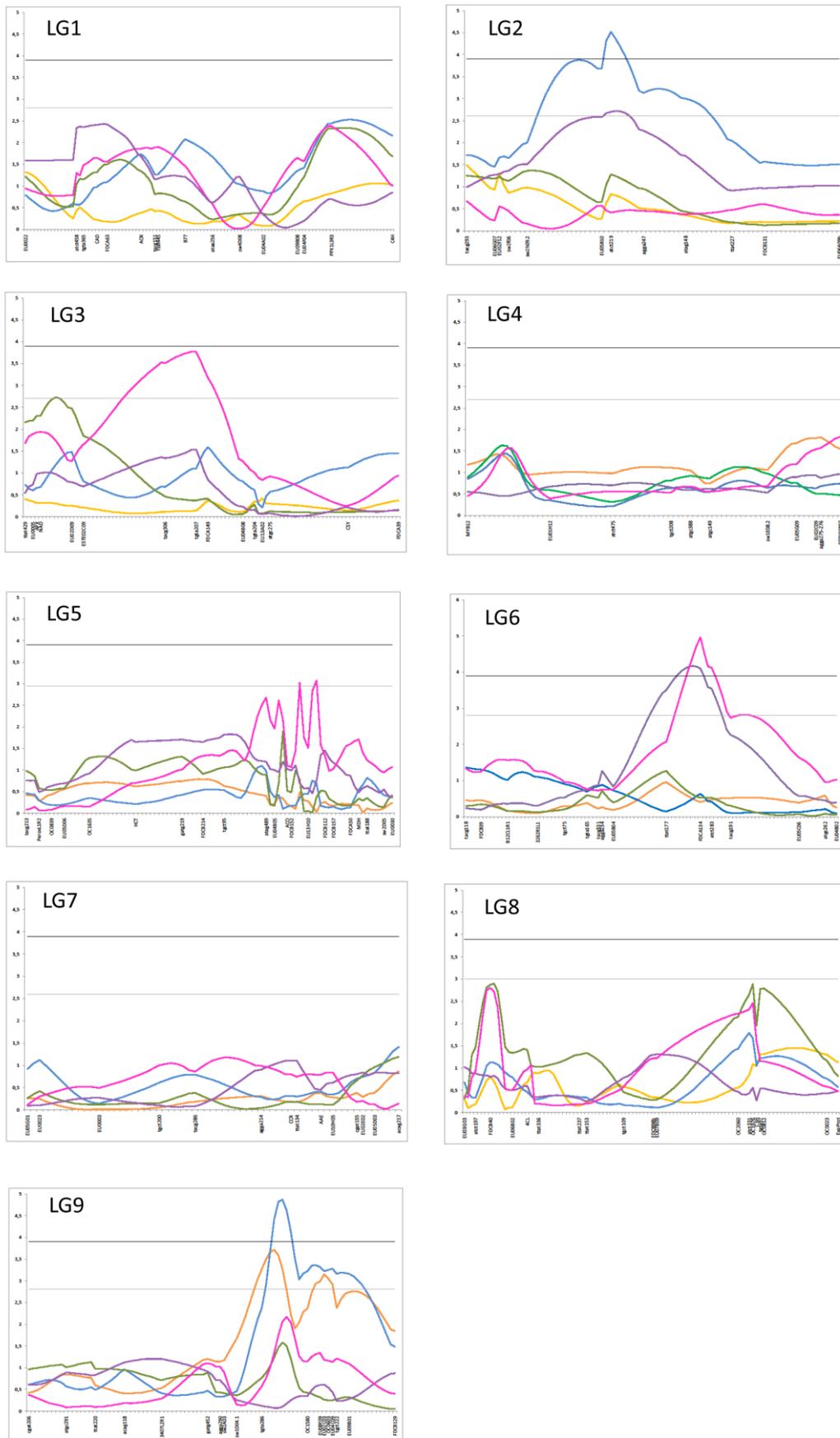
PCHLO	ee / h-	ee / kk	ef / h-	ef / kk	eg / h-	eg / kk	fg / h-	fg / kk
FDCB40 / ttat336	0,15305 a	0,13966667 a	0,1658 a	0,15543243 a	0,14605556 a	0,28500000 b	0,15137500 a	0,13550000 a

PCHLO	ll / ee	ll / ef	ll / eg	ll / fg	lm / ee	lm / ef	lm / eg	lm / fg
ttat429 / EU03G03	0,14888889 a, b, c, d	0,1552 b, d	0,152875 a, b, c, d	0,13181818 a, c	0,1445 a, b, c, d	0,14922222a, b, c, d	0,134875 a, c, d	0,177 b

PCHLO	ee / nn	ee / np	ef / nn	ef / np	eg / nn	eg / np	fg / nn	fg / np
swE08.2 / EU0003	0,14273077 a	0,17490476 b	0,15500000 a, b	0,14846667 a	0,14885714 a	0,15129167 a	0,15239286 a	0,14413333 a

AAR	ll / nn	ll / np	lm / nn	lm / np
EU04A02 / agga234	0,37404348 a	0,361625 b, c	0,35858333 b	0,37320455 a, c

Annexe 5 : Résultats de l'analyse par cartographie d'intervalle (IM). Les marqueurs sont en abscisse. Le LOD score est en ordonnée. Le seuilGW (=3,9) est représenté par un trait noir tandis que les seuilsLGn sont colorés en gris. La courbe rose correspond au caractère PSR. Les autres courbes correspondent aux caractères dont les valeurs ont été corrigées par le PSR : courbe orange = ACAFT ; courbe bleue = ACHLO ; courbe verte = ACHIC ; courbe violette = AAR.



ABSTRACT

Secondary metabolism corresponds to a class of compounds allowing plants to survive in their environment. Among these molecules, polyphenols display widely studied antioxidant properties. The aim of this work was to understand the genetic basic control of biosynthesis of polyphenols detected in chicory leaf tissue, using a QTL (Quantitative Trait Loci) analysis approach based on a F1' progeny (K28xK59).

First, a high-throughput protocol was set up to analyse and quantify the polyphenolic content of chicory leaves. Three phenolic acids were detected by HPLC and mass spectrometry analyses: caftaric acid, chlorogenic acid and chicoric acid. The protocol was combined with an adapted HPLC method and a radical scavenging ability test (DPPH assay). Quantity of each compound and radical scavenging ability showed significant variability in the F1' progeny, allowing us to use it for QTL analyses.

In this context, 201 individuals (repeated in 3 to 5 clones) from the F1' progeny were phenotyped for six traits that showed significant variability: caftaric, chlorogenic and chicoric acid amounts (respectively ACAFT, ACHLO, ACHIC), the total amount for these three molecules (ATOT), the radical scavenging ability (AAR), the ratio between the caftaric acid or chlorogenic acid and the entire phenolic acids detected (PCAFT and PCHLO, respectively).

With the aim of performing QTL mapping, a specific genetic map was established from a previous map already published. Using SSR, AFLP, SSCP, HRM polymorphism, 142 markers covered the nine linkage groups of this map. Among them, 16 candidate genes encoding transcription factors and enzymes belonging to phenylpropanoid pathway and ascorbate degradation pathway were mapped.

Altogether, 20 QTLs were then detected: 1 for ACAFT ($R^2 = 16.4\%$), 3 for ACHLO ($R^2 = 50\%$), 2 for ACHIC ($R^2 = 13.9\%$), 4 for AAR ($R^2 = 31\%$) et 5 for PACFT ($R^2 = 44\%$) and PCHLO ($R^2 = 61\%$). Eight QTLs co-localised with 7 encoding-enzymes genes involved in the phenylpropanoid pathway. Besides, epistatic effects were detected outside QTLs confidence intervals.

This study is the first step toward understanding the genetic control of the main phenolic acids in chicory leaves, particularly caftaric and chicoric acid, only synthesised in few species. For breeders, the identification of molecular markers and specific genes associated with loci controlling antioxidant activity and phenolic acids amounts in chicory leaves open the possibility to apply this knowledge to commercial chicory cultigroups.

Key words: chicory (*Cichorium intybus* L.), QTL, antioxidants, polyphénols, candidate gene.

RESUME

Le métabolisme secondaire regroupe un panel de composés permettant aux plantes de réagir dans un environnement donné. Parmi ces molécules, on retrouve les polyphénols dont le pouvoir antioxydant a été largement démontré. Le but de cette thèse est de comprendre les bases génétiques contrôlant la biosynthèse des molécules polyphénoliques synthétisées dans les feuilles de chicorée via une approche QTL (Quantitative Trait Loci), basée sur une descendance de type F1'(K28xK59).

Dans un premier temps, un protocole d'extraction simple et miniaturisé a été mise en place afin d'analyser et de quantifier le contenu polyphénolique des feuilles de chicorée. Trois acides phénols ont été détectés par CLHP et validés par spectrométrie de masse: l'acide caftarique, chlorogénique et chicorique. Ce protocole a ensuite été complété par une méthode CLHP rapide et adaptée aux molécules d'intérêt et par un test antiradicalaire (DPPH). La variabilité concernant la quantité de ces molécules et du pouvoir antiradicalaire a été constaté sur des échantillons de la descendance K28xK59, justifiant l'utilisation de cette descendance pour une étude génétique de type QTL.

Dans ce cadre, une descendance F1' de 201 individus a été dédiée à ce projet (3 à 5 clones par individus). Lors du phénotypage, six caractères ont témoigné de variations significatives : la quantité d'acide caftarique (ACAFT), d'acide chlorogénique (ACHLO), d'acide chicorique (ACHIC), la quantité de ces trois molécules réunies (ATOT), l'activité antiradicalaire (AAR) ainsi que le ratio de la quantité d'acide caftarique et d'acide chlorogénique par rapport à l'ensemble des acides phénols détectés en CLHP, soit PCAFT et PCHLO, respectivement.

A partir d'une carte de référence déjà publiée, une carte génétique spécifique au projet a été construite. Sur les neuf groupes de liaison de la chicorée, 142 marqueurs moléculaires de type SSR, AFLP, SSCP et HRM ont été répartis. Parmi eux sont cartographiés seize gènes candidats codant des facteurs de transcription et des enzymes impliquées dans la voie de biosynthèse des phénylpropanoïdes et de dégradation de l'acide ascorbique, celle-ci étant étroitement liée à la formation des acides caftarique et chicorique.

Vingt QTL ont alors été détectés : un QTL pour ACAFT ($R^2 = 16,4\%$), trois pour ACHLO ($R^2 = 50\%$), deux pour ACHIC ($R^2 = 13,9\%$), quatre pour AAR ($R^2 = 31\%$) et cinq pour PACFT ($R^2 = 44\%$) et PCHLO ($R^2 = 61\%$). Parmi eux, huit QTL co-localisent avec 7 gènes candidats codant des enzymes impliquées dans la voie de biosynthèse des phénylpropanoïdes. Par ailleurs, des effets épistatiques ont été constatés parmi les marqueurs de la carte, hors intervalle de confiance des QTL.

Les résultats de cette étude nous ont permis de fournir pour la première fois des éléments sur le contrôle génétique des principaux acides phénols détectés chez la chicorée, notamment l'acide chicorique et caftarique qui ne sont accumulés que chez quelques espèces. Pour les sélectionneurs, l'identification de marqueurs moléculaires et de gènes spécifiques reliés aux loci contrôlant l'activité antiradicalaire et la quantité d'acides phénols dans les feuilles de *Cichorium intybus* L. est d'autant plus intéressante que de nombreux cultigrupes de chicorées « salades » sont commercialisées.

Mots-clés : chicorée (*Cichorium intybus* L.), QTL, antioxydants, polyphénols, gènes candidats.