

Guillaume VALLIN
M2-SPV-AP
2015-2016



Une cartographie génétique saturée du fraisier cultivé octoploïde et son utilisation pour la recherche de régions génomiques associées à la variation quantitative de métabolites secondaires.

Rapport méthodologique non confidentiel



Maitre de stage : **Béatrice Denoyes**

Introduction :

La filière fraisicole en France occupe la 3^{ème} place en chiffre d'affaire au sein de la filière fruit après la pomme et la poire et avant la pêche-nectarine (INTERFEL, comm. pers.). Ce fruit est le 8^{ème} plus consommé par les ménages français (données FranceAgriMer, 2014).

Dans un contexte d'ouverture internationale et d'échanges, la France doit faire face à une très forte concurrence des pays du sud de l'Europe, principalement l'Espagne (1^{er} producteur européen) et le Maroc (La production de fraise à Huelva, Cécile Bernard, 1992). Ces pays produisent un fruit bon marché grâce à leurs grandes zones de production et leurs coûts réduits de main d'œuvres. De plus, ils se placent sur le marché précoce grâce à leurs conditions climatiques favorables. Compte-tenu du transport, les fraises issues de ces zones de production ont une qualité moyenne car elles ont été sélectionnées pour leur fermeté et récoltées souvent avant maturité.

Face à cette situation, le marché français a fait le choix de se démarquer en produisant une fraise de haute qualité organoleptique. La variété Gariguette a ainsi permis la mise en place, au sein de la filière française, d'un marché de fraises de bouche de grandes saveurs et aux propriétés antioxydantes importantes (La production de fraise à Huelva, Cécile Bernard, 1992).

Aujourd'hui, les sélectionneurs français comme Ciref, Marionnet ou Darbonne/Planasa poursuivent cet effort en recherchant une variété produisant des fraises de qualité et dont le rendement permette le retour sur investissement.

Un programme de sélection étant sur 10 ans, l'utilisation de marqueurs dans la sélection s'impose pour raccourcir et optimiser le processus de création variétale. Le génome du fraisier cultivé présente la particularité génétique d'être octoploïde ($2n = 8x = 56$) (Darrow 1966), ce qui rend plus complexe le développement de marqueurs moléculaires liés à des caractères d'intérêt agronomiques. Cependant, des améliorations techniques ont été apportées ces dernières années dans le génotypage haut débit chez cette espèce (Bassil et al. 2015 ; Sanchez-Sevilla et al. 2015), ce qui permet de rechercher efficacement les zones du génome intervenant dans les critères clés du fraisier dont la qualité du fruit (Lerceteau-Köhler et al. 2012). Une fois ces zones identifiées grâce à une approche QTL (Quantitative Trait Loci), les marqueurs sélectionnés les encadrant doivent être validés chez des ressources génétiques puis utilisés en sélection pour créer de nouvelles variétés de haute qualité organoleptique.

En 2015, l'INRA s'est associée au Ciref (Création Variétale Fraises Fruits Rouges), qui représente un groupement de producteur tournée vers la sélection, ainsi que Rennes Agrocampus-ouest et le GIS Fruits pour présenter un sujet d'étude centré sur la qualité du fruit et l'identification de marqueurs liés à un fruit riche en antioxydants. L'objectif est, à terme, de développer des marqueurs permettant une sélection assistée plus fiable pour améliorer les caractères organoleptiques des futures variétés. Ce projet était rendu possible par le génotypage de la population de cartographie de référence à l'aide de la puce Affymetrix de 90K SNP (Bassil et al. 2015) et par la disposition d'un grand jeu de données de métabolites secondaires quantifiés dans un projet trilatéral ERA-CAPS (FraGenomics, 2009-2012).

La méthodologie utilisée pour la construction des cartes de liaison génétiques à l'aide des logiciels R et JoinMap sera détaillée ainsi que l'approche QTL à l'aide du package 'qtl' de R (Karl W Broman and Hao Wu, Tools for Analyzing QTL Experiments). Les résultats des QTL associés à la teneur en antioxydants seront présentés puis discutés en fin de mon rapport.

Le premier objectif de ce stage était de réaliser une nouvelle carte génétique dense de la population de référence française du fraisier cultivé en s'appuyant sur les marqueurs SNP générés par une puce Affymetrix de 90K SNP (Bassil et al. 2015).

Le deuxième objectif était l'analyse exploratoire des données de métabolites secondaires, principalement appartenant à la voie des polyphénols. Ces données avaient été générées dans le projet FraGenomics (ERANET France, Espagne, Allemagne).

Le troisième objectif était d'identifier des Quantitative Traits Loci associés à la variation des concentrations en métabolites secondaires dans le fruit en s'appuyant sur la nouvelle carte génétique. Cette analyse était complétée par une analyse des eQTL disponibles sur un nombre limité de descendants (21).

Ces analyses permettront une recherche de gènes d'intérêts influant de manière directe (enzyme) ou indirecte (facteur de transcription) dans la voie des polyphénols. De plus, ce stage permet d'identifier des marqueurs associés à des gènes d'intérêts qui seront utilisés en sélection pour augmenter la teneur en antioxydants.

Methodologie :

1) Le matériel végétal

La population en ségrégation, Capitola x CF1116

Le matériel utilisé est une population en ségrégation pseudo-F₁ provenant du croisement entre la variété 'Capitola' (CA75.121-101 x Parker, University of California, Davis, USA) et la lignée du Ciref 'CF1116' ('Pajaro' x ['Earlyglow' x 'Chandler'], référence Ciref, France). Les parents ont été choisis pour leurs caractères contrastés (Lerceteau-Khöhler et al. 2012). En résumé, Capitola présente des caractères intéressants en termes de production (nombreux fruits à gros calibre) alors que CF1116 a des fruits de haute qualité organoleptiques mais moins nombreux.

2) Données phénotypiques à disposition

Les données obtenues par GC-MS

Les quantités de 21 métabolites secondaires ont été mesurées par rapport à un témoin (quantités relatives) chez 160 descendants de la population Capitola X CF1116 répartis sur deux années : 80 descendants en 2010 et 149 descendants en 2011. Trois répétitions techniques ont été réalisées chaque année.

Les mesures ont été réalisées par GC-MS. Cette technique qui combine les caractéristiques de la chromatographie en phase gazeuse (GC) et du spectromètre de masse (MS) permet l'identification et la quantification de différents composés dans des échantillons. La GC utilise une colonne de capillaire qui possède des caractères physiques (taille, diamètre) et chimiques (propriétés de la phase gazeuse) spécifiques. Elle permet la séparation des molécules suivant leurs affinités avec la phase stationnaire le long de la colonne. Les substances sont rejetées par la colonne à des temps différents (temps de rétention) vers le spectromètre de masse qui les capture, les accélère et les ionise. Les différents fragments ionisés des molécules peuvent ainsi être détectés et caractérisés suivant leur temps de rétention, leur masse et leur charge.

Les données eQTL

Une puce a été réalisée à partir d'oligomères de 60 nucléotides et représentait 34921 gènes (Bombarely et al. 2010; Medina-Puche et al. 2014). L'ARN des fruits de 20 descendants a été extrait. Pour l'hybridation microarray, 200 ng d'ARN total de chaque échantillon a été marqué à l'aide du kit 'Low Input Quick Amp Labeling Kit' (Agilent Technologies). Le scan des puces hybridées a permis d'attribuer une valeur d'expression de chaque gène pour chacun des échantillons.

L'analyse des données de la puce montre que 41% des données correspondent à des gènes localisés et caractérisés sur le génome de référence diploïde Fvba2.0 (*Fragaria vesca* ssp *bracteata*), 13% sont localisés mais leur fonction n'est pas encore définie, 45% sont des gènes non localisés et non caractérisés et les 1% restant sont des contigs ou des fragments génomique

3) La construction des cartes de liaison génétique

La puce SNPs 'Affymetrix® Axiom® IStraw90®'

La puce 'Affymetrix® Axiom® IStraw90®' a été développée entre 2010 et 2013 dans le cadre du projet RosBREED pour améliorer la recherche de marqueurs pour la sélection de caractères agronomiques chez les rosacées (Bassil et al, 2015). Cette puce regroupe 138 000 marqueurs SNPs qui représentent 90.000 localisations sur le génome de *F vesca* Hawaii v1.1, génome qui était disponible au moment de la construction de cette puce. A l'aide de cette puce, 178 individus de la population Capitola X CF1116 ont été génotypés.

La sélection des marqueurs SNPs

Choix des marqueurs

Les ratios observés au sein de la population en ségrégation ont été étudiés en tenant compte de l'octoploïde du fraisier cultivé. Ils ont été comparés à l'ensemble des ratios théoriques attendus chez un octoploïde à l'aide du test du Chi2 et au seuil de 1 %.

Les marqueurs polymorphes ségrégeant 1:1 dans la descendance et présents chez l'un des deux parents (configurations AB : BB ou AA : AB) ont été retenus pour la construction des cartes de liaison génétiques mâle ou femelle. Ceux qui présentaient des ratios de ségrégation distordus (<1%) n'ont pas été retenus dans cette analyse. De plus, les SNP présentant plus de 20 données manquantes ont été écartés. Parmi les différents types de SNP, seuls les marqueurs de type PolyHighResolution et NoMinorHom qui sont les plus informatifs et les plus « sûrs » ont été sélectionnés. On peut remarquer que lorsque l'on choisit ce type de SNP, le nombre de NA est au maximum de 3%. Environ 27 600 marqueurs SNP ont été gardés pour l'étape suivante, ils sont de la forme AA : AB, AB : BB ou AB : AB.

La carte de liaison génétique étant utilisée par la suite pour la détection de QTL, nous n'avons pas intégrés les marqueurs présents chez les deux parents et ségrégeant 3:1 (configuration AB : AB). En effet, leur positionnement est moins précis sur une carte issue d'une population de type pseudo-F₁.

En résumé, nous avons retenu un total de 27616 SNP (24% de SNP PolyHighResolution et 76% de SNP NoMinorHom). Parmi ces marqueurs, 13037 et 9117 permettaient de construire les cartes femelle et mâle respectivement. Afin de classer rapidement les groupes de liaison dans les groupes d'homéologie et de relier cette nouvelle carte aux anciennes cartes déjà construites (Lerceteau-Kohler et al. 2003 ; 2012), 607 marqueurs AFLP et SSRs dont la répartition sur l'ensemble des cartes disponibles étaient bien connus ont été intégrés.

Similarité des marqueurs avec le logiciel JoinMap et attribution des groupes de liaison

Dans un premier temps, les marqueurs présentant une ségrégation identique sur l'ensemble de la population (distance génétique 0 cM) ont été identifiés grâce au logiciel JoinMap4.0 (Van Ooijen and Voorrips 2001). Ces marqueurs n'ont pas été intégrés dans la suite de l'analyse.

Pour l'attribution des marqueurs à un groupe de liaison, le logiciel R a été choisi afin de permettre une puissance suffisante pour les analyses. Le package 'OneMap' de R permet, grâce à la fonction « group », de rassembler les marqueurs en des groupes de liaison potentiels.

Les SNP ont donc été analysés par type de ségrégation et par groupes de 1000 marqueurs avec les anciens marqueurs localisés. Les paramètres choisis (minimum de LOD = 4 et maximum de recombinaison = 0,2) pour cette analyse d'attribution ont été stringents afin de disposer de cartes fiables. Les marqueurs SSR et AFLP ont été utilisés comme « guides » pour attribuer les marqueurs SNP à un groupe de liaison. Ainsi, chaque groupe formés par le package est attribué à un groupe de liaison suivants les marqueurs SSR et AFLP qu'il contient.

Cartographie génétiques mâle et femelle à l'aide de JoinMap

La cartographie de liaison génétique a été effectuée à l'aide du logiciel JoinMap 4.0 (Van Ooijen 2006). Les options suivantes ont été déclarées : population de type 'CP', fonction de cartographie Kosambi, LOD seuil > 10, fréquence de recombinaison de 0,35, et un seuil de saut (jump) de 5,0. Les groupes d'homéologie ont été analysé un par un. Les cartes mâles et femelles contenant respectivement les marqueurs de type AB : BB et AA : AB ont été analysés séparément.

En se basant sur une première carte de liaison génétique, les erreurs sur les individus génotypés par la puce Affymetrix ont été repérées en comparant la ségrégation des marqueurs SNP à celle des marqueurs obtenus précédemment avec les marqueurs SSR et AFLP. Ces individus ont été par la suite retirés de l'analyse et de nouvelles cartes de liaison génétique ont été obtenues.

4) Détection de QTL

Simple Interval Mapping

La détection de QTL a été effectué dans un premier temps par Simple Interval Mapping (SIM) en utilisant le package 'qtl' du logiciel R. La population utilisée contient les 165 individus issus du croisement Capitola × CF1116 cartographiés sur les cartes de liaison génétique mâle et femelle obtenues précédemment. Les données de phénotypage regroupaient les 21 métabolites secondaires mesurés en 2010 et 2011. Les analyses ont été conduites séparément sur les cartes mâle et femelle et sur les années 2010 et 2011.

Le LOD seuil, permettant de déclarer un QTL comme significatif, a été calculé pour chaque variable en effectuant 1000 permutations à l'aide de la fonction 'perm'. Les moyennes des LOD seuil pour la totalité des 21 variables ont été mesurées afin de faciliter la lecture des résultats.

Elles sont égales à 2.8 et 3 respectivement pour $\alpha=0.10$ et $\alpha=0.05$ sur les deux cartes génétiques. La proportion de la variation phénotypique expliquée pour chaque QTL a été estimée par le coefficient R^2 au niveau du pic du QTL.

Choix des covariables et Composite Interval Mapping

Le choix du nombre de covariables dépend du nombre de QTL significatifs obtenus en SIM. Les marqueurs utilisés comme covariables sont situés au niveau des pics des QTLs à l'aide de la fonction « cim » du package 'qtl'. Une fenêtre de 10 cM a été choisie arbitrairement autour de ces marqueurs. Les LOD seuils restent identiques à l'analyse en SIM effectuée précédemment.

L'analyse en CIM permet d'affiner la localisation du QTL dans le but d'une recherche de gène d'intérêt. Les marqueurs SNP sont localisés sur la carte physique Fvba2.0 qui comprend les 34921 gènes utilisés pour l'acquisition des données d'expressions eQTL.

Une analyse de détection de eQTL est effectuée avec les données d'expression des gènes situés au niveau du QTL pour rechercher une éventuelle colocalisation entre le QTL du caractère et un gène localisé sur la carte physique.

Gènes candidats et eQTL

Pour un nombre restreint de caractères, nous avons conduit l'analyse jusqu'à la recherche du gène candidat le plus pertinent. Dans un premier temps, en s'appuyant sur la localisation des marqueurs SNP sur la carte physique de la référence Fvba.2, il a été possible d'identifier tous les gènes présents dans la région du QTL. Cependant, même avec une taille restreinte du QTL (5cM par exemple), plusieurs centaines de gènes pouvaient être présents sur la zone du QTL. Afin de conduire une démarche d'analyse d'eQTL sur un nombre plus restreints de gènes candidats, nous avons retenu ceux qui étaient le plus pertinents selon le caractère étudié.

Dans un deuxième temps, nous avons conduit une analyse de détection d'eQTL en nous appuyant sur les données d'expression des gènes retenus situés au niveau du QTL. L'objectif était d'identifier un eQTL significatif statistiquement et qui co-localise avec le QTL du caractère étudié.

Dans un troisième temps, nous avons recherché si le gène identifié était un bon candidat compte tenu des données de la voie de biosynthèse des polyphénols

Résultats

1) Les cartes génétiques mâles et femelles

Les cartes de liaison génétiques femelle et mâles ont été construites avec respectivement 5216 et 4879 marqueurs tous types confondus (SSR, AFLP et SNP). Les marqueurs sont cartographiés sur 28 groupes de liaison pour la carte femelle et 32 pour la carte mâle. Chez cette dernière carte, la présence de quatre groupes de liaison supplémentaires par rapport aux 28 attendus pourrait être due à des petites parties de chromosomes non liées aux autres groupes ou à des groupes de liaison coupés en deux.

Parmi ces quatre groupes de liaison, les groupes M41 et M44 pourraient être intégrés dans un des groupes de liaison appartenant au groupe d'homéologie 6. En effet, la position des marqueurs SNP de ces deux groupes sur le chromosome 6 de la référence Fvb_a2.0, de 37 million à 37,7 million de pb pour M41 et 36,1 à 38,8 million de pb pour M44, suggère que un de ces deux groupes pourrait être rattachés au groupe M6c qui ne fait que 23 cM et qui inclus des marqueurs de 14 à 23,7 million de paires de base sur ce chromosome. L'autre groupe serait rattaché au M6d qui inclut des marqueurs de 1,4 à 25,1 million de paires de base.

La scission du groupe M3c en deux sous-groupes, M3c1 et M3c2, est probablement due à la présence d'un gap entre ces deux parties. De façon surprenante, le groupe M1 ne contient exclusivement que des marqueurs AFLP (23 marqueurs). Aucun marqueur SNP n'est venu s'ajouter à ce groupe qui préexistait dans l'ancienne carte décrite par Lerceteau Kholler et al. (2003). Le fait qu'aucun marqueur SNP n'ait pu être ajouté à ce groupe M1 rend son existence hypothétique et la liaison entre ses marqueurs pourrait être due au hasard.

Les tailles totales des cartes femelle et mâle sont respectivement de 4135 cM et 3929 cM. Elles contiennent des groupes de liaison allant de 19 à 249 cM (femelle) et de 6 à 235 cM (mâle) respectivement. Les marqueurs sont répartis de façon homogène sur les groupes de liaison, excepté sur certains groupes qui présentent des gaps importants (e.g. 37 cM sur le F4b). Ces gaps pourraient être dus à l'homozygotie résiduelle des parents Capitola et CF1116 ou à des portions du génome mal couvertes par les marqueurs.

La répartition des marqueurs en couplage et en répulsion est significativement identique à celle attendue de 50% de marqueurs en couplage et 50% de marqueurs en répulsion excepté pour huit groupes de liaison sur la carte femelle et huit sur la carte mâle. La présence exclusive, ou quasi exclusive, de marqueurs en couplage chez certains groupes suggèrent, qu'à la méiose, le chromosome ne s'apparie pas de façon spécifique avec toujours le même chromosome homologue. Ceci serait le résultat d'un comportement polysomique résiduel chez le fraisier cultivé. L'existence de ces groupes de liaison en couplage seul est renforcée par la présence des mêmes groupes chez les cartes femelles et mâles comme par exemple pour le groupe 7b qui présente respectivement 0% et 4% de marqueurs en répulsion chez les cartes femelle et mâle.

Tableau 5 : Caractéristiques des cartes de liaison génétique femelle (gauche) et mâle (droite) cartographiés avec le logiciel JoinMap 4.0

Groupe de liaison	Nb. Marqueurs	Taille du LG	Couplage	Répulsion	Groupe de liaison	Nb. Marqueurs	Taille du LG	Couplage	Répulsion
F1a	223	162,423	63%	37%	M1a	198	130,614	100%	0%
F1b	279	158,672	77%	23%	M1b	42	28,609	48%	52%
F1c	112	86,393	100%	0%	M1c	106	82,526	50%	50%
F1d	143	110,872	59%	41%	M1d	53	54,009	96%	4%
F2a	345	227,932	26%	74%	M2a	207	167,074	58%	42%
F2b	188	189,109	100%	0%	M2b	186	136,648	75%	25%
F2c	57	58,285	46%	54%	M2c	223	141,882	100%	0%
F2d	47	20,153	100%	0%	M2d	60	69,27	87%	13%
F3a	128	145,804	18%	82%	M3a	290	198,294	100%	0%
F3b	282	185,897	42%	58%	M3b	143	165,152	85%	15%
F3c	129	103,933	81%	19%	M3c1	131	76,283	58%	42%
F3d	178	167,432	57%	43%	M3c2	129	61,892	24%	76%
F4a	157	131,831	49%	51%	M3d	232	154,774	85%	15%
F4b	108	128,95	73%	27%	M4a	218	192,232	81%	19%
F4c	166	142,538	60%	40%	M4b	122	99,821	100%	0%
F4d	116	143,863	31%	69%	M4c	124	174,777	98%	2%
F5a	274	223,159	15%	85%	M4d	108	145,081	42%	58%
F5b	147	146,919	73%	27%	M5a	265	217,012	67%	33%
F5c	169	137,192	50%	50%	M5b	277	206,742	39%	61%
F5d	221	165,443	100%	0%	M5c	223	128,092	43%	57%
F6a	417	241,222	29%	71%	M5d	30	34,115	72%	28%
F6b	364	249,226	40%	60%	M6a	329	235,457	73%	27%
F6c	145	145,183	76%	24%	M6b	245	227,902	37%	63%
F6d	141	125,903	30%	70%	M6c	24	23,108	23%	77%
F7a	200	161,043	80%	21%	M6d	241	194,925	25%	75%
F7b	194	205,998	100%	0%	M7a	133	71,336	71%	29%
F7c	220	150,739	48%	52%	M7b	136	77,326	96%	4%
F7d	66	18,984	26%	74%	M7c	124	121,35	80%	20%
Total (28 LGs)	5216	4135,098	55%	45%	M7d	201	142,013	33%	67%
					M_AFLP	23	142,057	100%	0%
					M41	11	6,413	53%	47%
					M44	45	22,409	59%	41%
					Total (32 LGs)	4879	3929,195	65%	35%

2) Détection de QTLs liés à la teneur en métabolites secondaires

QTLs observés

En 2011, au moins un QTL a été identifié pour chacun des métabolites alors qu'en 2010, où le nombre d'individus était plus faible, aucun QTL n'a été identifié pour certain métabolite.

Au total, 41 QTL significatifs, 30 mâles et 11 femelles, ont été identifiés avec un LOD supérieur à 3,0. Parmi ces QTL, quatre peuvent être considérés comme identiques en 2010 et 2011. D'autres QTL ont été identifiés sur un même groupe de liaison en 2010 et en 2011 mais à des positions différentes.

Colocalisations

Deux types de colocalisation ont été observés: (1) une localisation identique entre plusieurs caractères et (2) une position identique d'un QTL pour un caractère entre les deux années de l'étude.

Sur la carte mâle, trois métabolites secondaires appartenant à la famille des phénylpropanoïdes (cn, cinnamoyl glucose ; cm, coumaroyl glucose et cf, caffeoyl glucose) ont colocalisé à la même position sur les deux années d'étude. Puisque ces trois métabolites se situent dans la partie de la voie métabolique des polyphénols (phénylpropanoïdes), l'hypothèse d'un seul QTL contrôlant la synthèse de ces trois métabolites est a été émise.

Trois autres colocalisations ont pu être détectés entre les deux années pour un même caractère: la pelargonidine glucoside malonate (pgm), l'acide ellagique (el) et l'HDMF glucoside (hd).

Concernant les résultats sur la carte femelle, on n'observe qu'un seul métabolite secondaire dont le QTL colocalise entre 2010 et 2011, le Kaempferol-glucuronique.

Certains QTL comme celui détecté pour l'acide ellagique sur le M7c ont présenté des intervalles très larges. Les valeurs des pics des LOD pour ces QTL sont très différentes entre les deux années. Cependant, compte tenu de l'imprécision des QTL identifiés, il pourrait s'agir d'un même QTL entre les deux années d'étude.

3) Détection de eQTLs et colocalisations : Exemple des phénylpropanoïdes

Analyse sur le groupe de liaison X

L'analyse QTL des données d'expression de gènes, eQTL, a été effectuée sur le QTL majeur localisé sur le groupe de liaison LGX (non précisé dans ce document) et qui contrôlerait la synthèse des trois métabolites cinnamoyl-glucose, coumaroyl-glucose et caffeoyl-glucose. En se basant sur la position des marqueurs encadrant la zone du QTL localisé sur LGX, nous avons pu identifier une région de 2,4 million de paires de base (pb) sur le groupe de liaison FvbX du

génomique de référence. Les données d'annotation de cette zone ont permis d'identifier 330 gènes. Grâce aux microarrays, les données d'expressions de ces 330 gènes étaient disponibles pour 21 des descendants.

Afin d'effectuer une analyse eQTL, nous avons tout d'abord réalisé un choix des gènes potentiellement candidats parmi les 330 gènes. Ce choix était basé sur l'implication possible du gène candidat dans la voie des phénylpropanoïdes. Au total, 17 gènes candidats ont été retenus pour l'analyse.

Les résultats de l'analyse en SIM ont permis d'identifier un eQTL significatif au seuil de 5% pour trois des 17 gènes candidats. Deux des eQTL sont localisés sur le LGX : un lié à une esterase et l'autre à une ADN cytosine méthyltransferase (méthylation de l'ADN). Compte tenu de sa position dans la voie métabolique des polyphénols, l'« esterase » est le candidat le plus probable sans qu'on puisse exclure les autres.

Un gène candidat codant pour une esterase

Ainsi, le gène codant pour l'esterase montre un eQTL au niveau du LGX à 10 cM avant le pic du QTL observé pour les trois phénylpropanoïdes (Figure 17). Ce eQTL présente un LOD score de 7,67 et explique 79% de la variation de l'expression de cette enzyme.

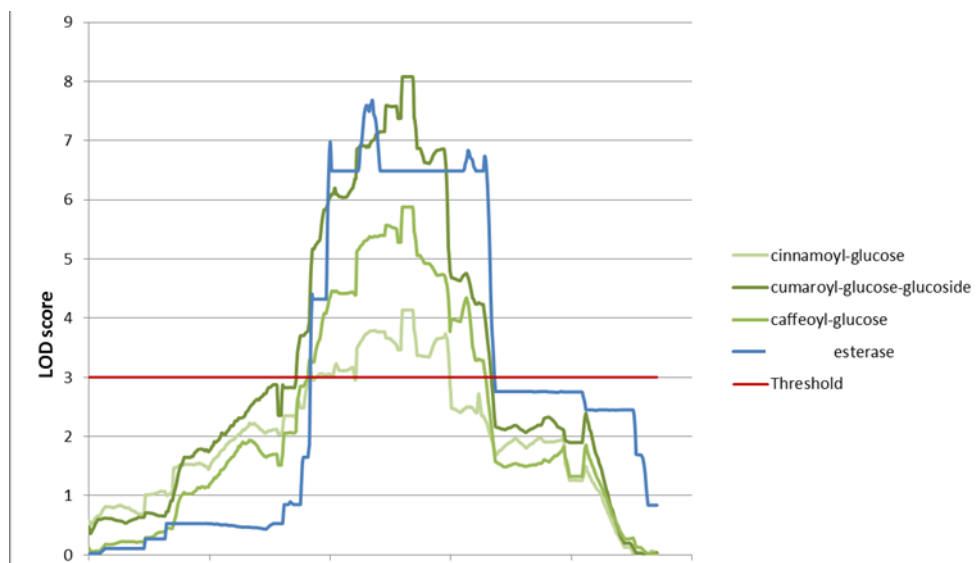


Figure 17. Graphique représentant les résultats de l'analyse en SIM pour les trois phénylpropanoïdes en 2011 (cinnamoyl-glucose, cumaroyl-glucose-glucoside et caffeoyl-glucose) (courbes vertes) et les résultats de l'analyse en SIM pour les données d'expression (eQTL) de l'esterase (courbe bleue) ; Le LOD_{seuil} est fixé à 3 (ligne rouge).

Conclusion

La méthode de cartographie appliquée dans cette étude a permis de construire les premières cartes génétiques femelles et mâles comprenant un nombre important de marqueurs SNP (5216 et 4879 respectivement).

Les QTL observés sur les deux cartes montrent qu'il existe un certain potentiel des polyphénols dans la sélection de nouvelles variétés plus riche en ce type de métabolites secondaires.

Par analyse des données d'expression des gènes cartographiés sur la carte de référence Fvb_a2.0, nous avons pu mettre en avant un gène candidat codant pour une esterase dont la diminution d'expression entraîne une forte augmentation des quantités en phénylpropanoïdes dans le fruit.

Existe-t-il un polymorphisme entre les deux parents, Capitola et CF1116, au niveau du gène de l'esterase (exons et promoteur) ? Pour répondre à cette question, une analyse bioinformatique sera réalisée sur les données de séquençage des deux parents de la population (séquençage Illumina 30X disponible en août 2016).

Y a-t-il une association entre le polymorphisme du marqueur situé au plus proche du gène candidat et la teneur en phénylpropanoïdes (cinamoyl-glucose, coumaroyl-glucose et caffeoyl-glucose) ? Cette étude sera conduite en collaboration avec l'IRTA-CRAAG de Barcelone qui faisait également partie du projet FraGenomics. L'étude de la génétique d'association à partir d'une collection de 40 génotypes analysés en 2010 et 2011 pour leur teneur en polyphénols (Fragenomics) et analysés avec la puce Affymetrix permettra de répondre à cette question.

La validation fonctionnelle du gène sera effectuée par transformation du fraisier. Une transformation transitoire par inoculation d'*Agrobacterium* portant une construction de type 35S::ARNi_esterase sera réalisée. Cette construction permettra d'éteindre le gène candidat, ce qui devrait conduire à une augmentation des teneurs en phénylpropanoïdes.

Les cartes génétiques réalisées dans cette étude constituent une base solide à partir de laquelle il sera possible d'incrémenter les marqueurs hétérozygotes présents en une seule dose chez les deux parents (non utilisés dans cette étude). De plus, les marqueurs qui ne répondaient pas aux paramètres très stringents mis en place pour la construction des cartes pourront également être ajoutés.

Compte tenu du rôle des polyphénols et de la vitamine C dans les résistances aux maladies, des marqueurs pourront être développés pour la sélection assistée par marqueurs (SAM) lors du processus de création de nouvelles variétés et un test de résistance à l'oïdium pourra être réalisé.

Bibliographie

- Agreste, Juillet 2014. Agreste Conjoncture – Légumes – Infos rapides, 4 pages
- Bassil, Nahla V., Thomas M. Davis, Hailong Zhang, Stephen Ficklin, Mike Mittmann, Teresa Webster, Lise Mahoney, 2015. « Development and Preliminary Evaluation of a 90 K Axiom® SNP Array for the Allo-Octoploid Cultivated Strawberry *Fragaria × Ananassa* ». *BMC Genomics* 16: 155. doi:10.1186/s12864-015-1310-1.
- Bernard Cécile, 1992. La production de fraise à Huelva : les avantages de la filière espagnole sur le marché européen. In: *Économie rurale*. N°212, 1992. pp. 12-18.
- Broman, Karl W., Hao Wu, Saunak Sen, et Gary A. Churchill. 2003. « R/Qtl: QTL Mapping in Experimental Crosses ». *Bioinformatics (Oxford, England)* 19 (7): 889-90.
- Darrow, George M., and Henry A. Wallace. 1966. *The Strawberry: History, Breeding, and Physiology*. 1st edition. Holt, Rinehart and Winston.
- Lerceteau-Köhler, E., G. Guérin, F. Laigret, et B. Denoyes-Rothan. 2003. « Characterization of Mixed Disomic and Polysomic Inheritance in the Octoploid Strawberry (*Fragaria X Ananassa*) Using AFLP Mapping ». *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 107 (4): 619-28. doi:10.1007/s00122-003-1300-6.
- Lerceteau-Köhler, E., A. Moing, G. Guérin, C. Renaud, A. Petit, C. Rothan, et Béatrice Denoyes. 2012. « Genetic Dissection of Fruit Quality Traits in the Octoploid Cultivated Strawberry Highlights the Role of Homoeo-QTL in Their Control ». *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 124 (6): 1059-77. doi:10.1007/s00122-011-1769-3.
- Rousseau-Gueutin, Mathieu, Estelle Lerceteau-Köhler, Laure Barrot, Daniel James Sargent, Amparo Monfort, David Simpson, Pere Arús, Guy Guérin, et Béatrice Denoyes-Rothan. 2008. « Comparative Genetic Mapping between Octoploid and Diploid *Fragaria* Species Reveals a High Level of Colinearity between Their Genomes and the Essentially Disomic Behavior of the Cultivated Octoploid Strawberry ». *Genetics* 179 (4): 2045-60. doi:10.1534/genetics.107.083840.
- Rousseau-Gueutin, M., A. Gaston, A. Aïnouche, M. L. Aïnouche, K. Olbricht, G. Staudt, L. Richard, and B. Denoyes-Rothan. 2009. 'Tracking the Evolutionary History of Polyploidy in *Fragaria L.* (Strawberry): New Insights from Phylogenetic Analyses of Low-Copy Nuclear Genes'. *Molecular Phylogenetics and Evolution* 51 (3): 515–30. doi:10.1016/j.ympev.2008.12.024.
- Van Ooijen, J.W., 2006. JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.